

Destek Vektör Regresyon ile Hava Kirliliđi Tahmini

Mahmut Esat Ekinci

YÜKSEK LİSANS TEZİ

Endüstri Mühendisliđi Anabilim Dalı

Nisan 2017

Air Pollution Prediction Using Support Vector Regression

Mahmut Esat Ekinci

MASTER OF SCIENCE THESIS

Department of Industrial Engineering

April 2017

Destek Vektör Regresyon ile Hava Kirliliđi Tahmini

Mahmut Esat Ekinci

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliđi Uyarınca
Endüstri Mühendisliđi Anabilim Dalı
Endüstri Mühendisliđi Bilim Dalında
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır

Danışman: Prof. Dr. A. Sermet Anagün

Nisan 2017

ONAY

Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans öğrencisi Mahmut Esat Ekinci'nin YÜKSEK LİSANS tezi olarak hazırladığı "Destek Vektör Regresyon ile Hava Kirliliği Tahmini" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oybirliği ile kabul edilmiştir.

Danışman : Prof. Dr. A. Sermet Anagün

İkinci Danışman : -

Yüksek Lisans Tez Savunma Jürisi:

Üye : Prof. Dr. A. Sermet Anagün

Üye : Doç. Dr. Gürkan Öztürk

Üye : Doç. Dr. Ezgi Aktar Demirtaş

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve
..... sayılı kararıyla onaylanmıştır.

Prof. Dr. Hürriyet ERŞAHAN
Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. A. Sermet Anagün danışmanlığında hazırlamış olduğum “Destek Vektör Regresyon ile Hava Kirliliği Tahmini” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 12/04/2017

Mahmut Esat Ekinci

İmza

ÖZET

Hava kirliliğinin tahminlenmesi, insan ve çevre sağlığına önemli zararlar veren hava kirliliğine karşı erken önlem alınması açısından oldukça önemlidir. Bu çalışmada günlük ortalama hava kirliliği miktarının, önemli bir hava kirletici olan SO₂ (kükürtdioksit) konsantrasyonu üzerinden tahminlenmesi amaçlanmıştır. Tahminleme modeli, Denizli il merkezinde 2005-2015 yılları arasında ölçülen SO₂, ve PM₁₀ (partikül madde) konsantrasyonu ile sıcaklık, basınç, çığ noktası, rüzgar, yağmur gibi meteorolojik verilerle kurulmuştur. Kirletici miktarı verileri Ulusal Hava Kalitesi İzleme Ağı'nın, meteorolojik veriler ise ABD Ulusal Okyanus ve Atmosfer Dairesi (NOAA)'nin veri tabanından çekilmiştir. Tahminleme çalışmasında, öncelikle radyal tabanlı, polinom, sigmoid ve Pearson VII gibi çeşitli çekirdek fonksiyonları ile çalıştırılan destek vektör regresyon yöntemi kullanılmıştır. Bununla beraber elde edilen sonuçları karşılaştırmak amacıyla, literatürde yaygın olarak kullanılan çok katmanlı sinir ağı ve çoklu doğrusal regresyon yöntemleri de kullanılmıştır.

Anahtar Kelimeler: Hava Kirliliği, Tahminleme, Destek Vektör Makinesi, Yapay Sinir Ağı, Çoklu Doğrusal Regresyon, Veri Madenciliği

SUMMARY

Prediction of air pollution is considerably important in terms of precautions to be taken against air pollution which cause significant harm to human and environmental health. In this study, it was aimed to predict the daily average concentration of SO₂ (sulfur dioxide), which is an important air pollutant. The prediction model was built using historical meteorological data such as temperature, pressure, dew point, wind, precipitation and concentration of SO₂ and PM₁₀ (particulate matter) measured in Denizli city center between the years 2005-2015. Air pollutant concentration data were retrieved from the database of the TR National Air Quality Monitoring System, while meteorological data were retrieved from the US National Oceanic and Atmospheric Administration (NOAA) database. In this prediction study, support vector regression method which was first run various kernel functions such as radial basis, polynomial, sigmoid and Pearson VII was used. In addition to these, multilayer neural network and multiple linear regression method, which are widely used in the literature, also was used in order to compare the results.

Key Words: Air Pollution, Prediction, Support Vector Machine, Artificial Neural Network, Multiple Linear Regression, Data Mining

TEŐEKKÜR

Bu alıőma kapsamında baőta deęerli danıőmanım Prof. Dr. A. Sermet Anagün'e, bütün alıőma süresince yanımda olan ailem ve arkadaşlarıma desteklerinden dolayı sonsuz teőekkürlerimi sunarım.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vi
SUMMARY	vii
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
ŞEKİLLER DİZİNİ	xii
ÇİZELGELER DİZİNİ	xiv
SİMGELER VE KISALTMALAR DİZİNİ	xv
1. GİRİŞ VE AMAÇ	1
2. LİTERATÜR ARAŞTIRMASI	4
3. HAVA KİRLİLİĞİ VE KİRLETİCİLER	11
3.1. Partikül Maddeler (PM)	11
3.2. Karbonmonoksit (CO)	13
3.3. Kükürtdioksit (SO ₂)	14
3.4. Azotdioksit (NO ₂)	15
3.5. Ozon (O ₃).....	16
4. DOĞRUSAL REGRESYON	17
4.1. Basit Doğrusal Regresyon Modeli	17
4.2. Çoklu Doğrusal Regresyon Modeli	20
4.3. Varyans Tahmini.....	22
4.4. Belirlilik Katsayısı	23
5. YAPAY SİNİR AĞLARI	25
5.1. Sinir Hücrelerinin Yapısı	25
5.1.1. Biyolojik sinir hücreleri	25
5.1.2. Yapay sinir hücreleri	26
5.2. Yapay Sinir Ağlarının Yapısı	28
5.3. Yapay Sinir Ağlarında Öğrenme Stratejileri	30
5.4. Yapay Sinir Ağlarında Öğrenme Kuralları	31
5.4.1. Hebb öğrenme kuralı.....	31

İÇİNDEKİLER (devam)

5.4.2. Perceptron öğrenme kuralı	32
5.4.3. Delta öğrenme kuralı.....	32
5.4.4. Kohonen öğrenme kuralı.....	33
5.5. Yapay Sinir Ağı Modelleri	34
5.5.1. Adaptif rezonans teorisi	34
5.5.2. Hopfield ağı.....	35
5.5.3. Kohonen ağı (SOM)	35
5.5.4. Sayaçlı yayılım ağı.....	36
5.5.5. Çok katmanlı algılayıcı	37
5.5.5.1. Genelleştirilmiş delta kuralının türetilmesi	37
5.5.5.2. Öğrenme katsayısı	40
5.5.5.3. Momentum terimi.....	40
6. DESTEK VEKTÖR MAKİNELERİ.....	42
6.1. İstatistiksel Öğrenme Teorisi (Vapnik-Chervonenkis Teorisi).....	42
6.1.1. VC boyutu	44
6.1.2. Yapısal risk minimizasyonu.....	45
6.2. Doğrusal Destek Vektör Sınıflandırıcısı.....	47
6.2.1. Doğrusal olarak ayrılabilme durumu (hard margin)	48
6.2.1.1. Primal (özgün) model.....	51
6.2.1.2. Dual (ikil) model	51
6.2.2. Bir miktar hata ile doğrusal ayrılabilme durumu (soft margin).....	54
6.2.3. Doğrusal olmayan destek vektör sınıflandırıcısı	59
6.3. Destek Vektör Regresyon	64
6.3.1. Doğrusal destek vektör regresyon.....	66
6.3.2. Doğrusal olmayan destek vektör regresyon	70
7. MATERYAL VE YÖNTEM	72
7.1. Veri Seti ve Veri Önışleme	72
7.1.1. Veri bütünleştirme.....	73
7.1.2. Veri temizleme	73
7.1.3. Veri dönüştürme.....	75

İÇİNDEKİLER (devam)

7.1.3.1. Min-max normalleştirilmesi	76
7.1.3.2. Standart skor (Z-score) normalleştirilmesi	76
7.2. WEKA	77
7.2.1. Explorer	79
7.2.2. Experimenter	81
7.2.3. Knowledge flow	82
7.2.4. Simple CLI	83
7.3. Model ve Değerlendirme	84
7.3.1. Genel tahminleme modeli	84
7.3.2. Eğitim ve test verisinin belirlenmesi	85
7.3.2.1. Bagging (bootstrap aggregating)	86
7.3.2.2. k-kat çapraz geçişleme	87
7.3.3. Performans göstergeleri	87
7.3.4. Parametre seçimi	89
7.3.4.1. Destek vektör makinesi için parametre seçimi	90
7.3.4.2. Çok katmanlı algılayıcı için parametre seçimi	92
8. BULGULAR VE TARTIŞMA	96
9. SONUÇ VE ÖNERİLER	102
KAYNAKLAR DİZİNİ	104
EK AÇIKLAMALAR	117
Ek Açıklama-A: Geri yayılım Algoritması	118
Ek Açıklama-B: Destek Vektör Sınıflandırıcı İçin Nümerik Bir Örnek	120
Ek Açıklama-C: Destek Vektör Regresyon İçin Açıklayıcı Bir Örnek	123
Ek Açıklama-D: Örnek Veri Seti	124
Ek Açıklama-E: Örnek Normalleştirilmiş Veri Seti	125
Ek Açıklama-F: Çok Katmanlı Algılayıcı Mimarisi	126

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
3.1. Partikül madde boyutları	12
3.2. PM emisyonunun kaynaklarına göre dağılımı.....	13
3.3. CO emisyonunun kaynaklarına göre dağılımı	13
3.4. SO ₂ emisyonunun kaynaklarına göre dağılımı	14
3.5. NO _x emisyonunun kaynaklarına göre dağılımı	15
4.1. Basit doğrusal regresyon modeli	18
4.2. Çoklu doğrusal regresyon modeli.....	20
5.1. Biyolojik sinir hücresi	26
5.3. Aktivasyon fonksiyonları	28
5.4. Çok katmanlı bir sinir ağı yapısı	29
5.5. ANN’de öğrenme stratejileri: a) gözetimli öğrenme, b) gözetimsiz öğrenme	30
5.6. ART ağlarının genel yapısı.....	34
5.7. Hopfield ağının yapısı	35
5.8. Kohonen ağının (SOM) yapısı.....	36
5.9. Geri yayılım ağının yapısı	39
5.10. Hata fonksiyonunda yerel ve bütünsel minimum	41
6.1. İkili sınıflandırma probleminin üç farklı modelle çözümü.....	43
6.2. \mathbb{R}^2 ’de 3 noktanın ayrımı.....	45
6.3. \mathbb{R}^2 ’de 4 noktanın ayrımı.....	45
6.4. Yapısal risk minimizasyonu	47
6.5. Doğrusal olarak ayrılabilen ikili sınıflandırma probleminde ayırıcı doğrular.....	48
6.6. Destek vektör sınıflandırıcı	49
6.7. Soft marjlinli destek vektör sınıflandırıcı	55
6.8. C parametresinin modele olan etkisi.....	56
6.9. Doğrusal ayırlamama durumu	59
6.10. Verinin üst uzayda doğrusal ayrılabilir duruma dönüştürülmesi.....	60
6.11. Girdi uzayının ϕx dönüşümü ile üst uzayda eşlenmesi.....	61
6.12. Kayıp fonksiyonları	65

ŞEKİLLER DİZİNİ (devam)

6.13. Doğrusal destek vektör regresyon	66
6.14. Doğrusal olmayan destek vektör regresyon.....	70
7.1. k-NN algoritması ile belirlenen aykırı değerler.....	75
7.2. WEKA’da izlenen adımlar	77
7.3. WEKA açılış ekranı.....	78
7.4. Explorer arayüzü.....	79
7.5. Explorer penceresinde classify sekmesi	80
7.6. Explorer penceresinde visualize sekmesi	81
7.7. Experimenter arayüzü.....	82
7.8. Knowledge Flow arayüzü	83
7.9. Simple CLI komut satırı	84
7.10. Genel tahminleme modeli.....	85
7.11. Bagging (torbalama)	86
7.12. 4-kat çapraz geçişleme	87
7.13. γ parametresinin modele olan etkisi.....	90
7.14. (C, γ) parametre ikilisinin test hatasına etkisi	91
7.15. Öğrenme katsayısı ve momentum teriminin test hatasına etkisi	93
7.16. Gizli katmandaki nöron sayısının test hatasına etkisi.....	94
7.17. Eğitim devir sayısının test hatasına etkisi.....	94
8.1. Tahminleme modellerinin gerçek değerlerle uyumu a)SVR b)MLR c)MLP.....	97
8.2. Ölçülen ve tahminlenen SO ₂ miktarları arasındaki doğrusal ilişki (Model 1).....	98
8.3. SVR ile yapılan zaman sıralı yıllık tahminlerin gerçek değerlerle uyumu.....	99
8.4. Ölçülen ve tahminlenen SO ₂ miktarları arasındaki doğrusal ilişki (Model 2).....	100

ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
3.1. Hava kirleticilerin ana kaynakları ve insan sağlığına olan etkileri.....	16
6.1. Çekirdek fonksiyonları	63
7.1. Veri setine ilişkin tanımlayıcı istatistikler	73
7.2. Korelasyon katsayısının yorumu	89
7.3. SVR için belirlenen parametre değerleri	92
7.4. MLP için belirlenen parametre değerleri.....	95
8.1. Test sonuçları.....	96
8.2. Doğalgaz kullanımıyla başlayan döneme ait test sonuçları.....	100

SİMGELER VE KISALTMALAR DİZİNİ

<u>Simgeler</u>	<u>Açıklama</u>
x	Bağımsız değişken, girdi
y	Bağımlı değişken, çıktı-sınıf
β	Regresyon katsayısı
ε	Hata terimi
N	Normal dağılım
δ	Kısmi türev
X	X matrisi
X^T	X matrisinin transpozu
X^{-1}	X matrisinin tersi
\bar{x}	x'in ortalaması
$\ x\ $	x'in normu
\hat{x}	x'in tahmini
$Var(x)$	x'in varyansı
σ^2	Varyans
σ	Standart hata
R^2	Belirlilik katsayısı
R_{adj}^2	Düzeltilmiş belirlilik katsayısı
r, R	Korelasyon katsayısı
w	Ağırlık
b	Eşik değer, sabit terim
z	Net girdi
f	Fonksiyon
$f(z)$	Aktivasyon fonksiyonu
d	Hedef değer
E	Hata değeri
Δ	Değişim miktarı
ρ	Benzerlik katsayısı
η	Öğrenme katsayısı

SİMGELER VE KISALTMALAR DİZİNİ (devam)

<u>Simgeler</u>	<u>Açıklama</u>
μ	Momentum katsayısı
T	Eğitim veri seti
D	Veri seti
\mathbb{R}^m	m boyutlu uzay
ϑ	VC güvenilirliği
h	VC boyutu
N	Küme büyüklüğü
n	Örnek sayısı
$R[f]$	Beklenen risk
$R_{emp}(f)$	Eğitim hatası (ampirik hata)
P	Olasılık
m	Marjin
\mathcal{H}	Hiper düzlem
d	Uzaklık
L	Lagrange fonksiyonu
sgn	İşaret fonksiyonu
ξ	Aylak değişken
C	Ceza parametresi
ϕ	İç çarpım fonksiyonu
I	Girdi uzayı
F	Öznelik uzayı
K	Çekirdek fonksiyonu
μ_e	Ortalama hata

<u>Kısaltmalar</u>	<u>Açıklama</u>
ANFIS	Nöro-Bulanık
ANN	Yapay Sinir Ağları
ARIMA	Otoregresif Bütünleşik Hareketli Ortalama Modeli

SİMGELER VE KISALTMALAR DİZİNİ (devam)

<u>Kısaltmalar</u>	<u>Açıklama</u>
ART	Adaptif Rezonans Teorisi
CO	Karbonmonoksit
cs-SVM	Maliyet Duyarlı Destek Vektör Makinesi
EPA	ABD Çevre Koruma Ajansı
GABPNN	Genetik Algoritmayla Optimize Edilmiş Geriyayılımlı Sinir Ağı
GRNN	Genelleştirilmiş Regresyon Sinir Ağları
k-NN	k-En Yakın Komşu
KKT	Karush-Kuhn-Tucker
MAE	Ortalama Mutlak Hata
MLP	Çok Katmanlı Algılayıcı
MLR	Çoklu Doğrusal Regresyon
MS _e	Artıkların Kareler Ortalaması
NOAA	ABD Ulusal Okyanus ve Atmosfer Dairesi
NO _x	Azotoksitler
NO ₂	Azotdioksit
O ₃	Ozon
Pb	Kurşun
PCA	Temel Bileşenler Analizi
PM	Partikül Maddeler
RBF	Radyal Tabanlı Fonksiyon
RMSE	Karesel Hata Ortalamalarının Karekökü
SS _e	Artıkların Kareler Toplamı
SS _r	Regresyon Kareler Toplamı
SS _t	Genel Kareler Toplamı
s-SVM	Standart Destek Vektör Makinesi
SO ₂	Kükürtdioksit
SOM	Kohonen Ağı (Self-Organizing Maps)
SVM	Destek Vektör Makinesi
SVR	Destek Vektör Regresyon

SİMGELER VE KISALTMALAR DİZİNİ (devam)

<u>Kısaltmalar</u>	<u>Açıklama</u>
THC	Hidrokarbonlar
VC Boyutu	Vapnik-Chervonenkis Boyutu
VC Teorisi	İstatistiksel Öğrenme Teorisi (Vapnik-Chervonenkis Teorisi)
VOC	Uçucu Organik Bileşikler
WEKA	Waikato Environment for Knowledge Analysis
YRM	Yapısal Risk Minimizasyonu

1. GİRİŞ VE AMAÇ

Hava kirliliği, fosil yakıtların kullanımına yoğun bir şekilde başlanılmasından itibaren çevre ve insanlık için önemli bir problem olmuştur. Sanayi Devrimi ile büyük bir ivme kazanan bu sorun, özellikle endüstrileşen ve yoğun nüfuslu yerleşim bölgelerini tehdit etmektedir. Hava kirliliği; küresel ısınma, asit yağmuru ve ötrofikasyon¹ gibi çevresel sorunların yanı sıra kanser, kalp ve solunum yolu hastalıkları gibi ciddi sağlık sorunlarına, hatta ölümle neticelenen vakalara neden olmaktadır. World Bank ve Institute for Health Metrics and Evaluation'ın (2013) araştırmasına göre 2013 yılında 5,5 milyon insan -iç mekân hava kirliliği de dâhil olmak üzere- hava kirliliğine bağlı nedenlerden yaşamını yitirmiştir. Bu sayı 2013 yılındaki her 10 ölümden 1'ine karşı gelmektedir. Bu nedenlerden dolayı hava kirliliğinin önceden tahminlenmesi önem kazanmaktadır. Böylece özellikle çocuklar, astım hastaları ve yaşlılar gibi hassas gruplar için erken önlemler alınabilecektir.

Hava kirleticileri konsantrasyonunun; emisyon kaynakları, coğrafi yapı ve meteorolojik faktörlerle doğrudan ilişkisi olduğu bilinmektedir. Bu faktörlerin modellenmesiyle hava kirliliğinin tahminlenmesi genel olarak iki temel yaklaşımla yapılmaktadır. Bunlardan ilki deterministik yaklaşımdır. Bu yaklaşımda kirletici konsantrasyonunun yayılımı, fiziksel ve kimyasal süreçlerin modellenmesiyle belirlenmeye çalışılmaktadır. Ancak coğrafi yapı, emisyon kaynaklarının çeşitliliği ile kirleticilerin havadaki taşınımı ve etkileşimin modellenmesi oldukça karmaşık bir süreçtir (Vallero, 2007). Hava kirliliğinin tahminlenmesinde kullanılan diğer bir yaklaşım da son yıllarda kullanımı artan istatistiksel modellerdir. Bu yaklaşımda kirletici konsantrasyonu ve bu konsantrasyonu etkileyen faktörler arasındaki doğrusal olmayan karmaşık ilişkilerin matematiksel ifadesine gerek duyulmadan modellenmesi amaçlanmaktadır (Gardner ve Dorling, 1999).

Bir gün sonrasına ait hava kirliliğini tahminlemeyi amaçlayan bu çalışmada, tahminleme modeli kurulurken kirletici verileri ve meteorolojik faktörler kullanılmıştır. Bhaskar ve Mehta (2010) çalışmalarında kirleticilerin atmosferdeki dağılımını; kirletici

¹Su ortamında azot ve fosfor miktarının gereğinden fazla artmasıyla alg ve planktonların artışı ve buna paralel olarak su kalitesinin ve ekosisteminin bozulması olayıdır.

kaynaklarının yanı sıra rüzgâr, yağmur, sıcaklık ve nem gibi meteorolojik faktörlerin de etkilediğini belirtmiştir. Tahminleme modelinde öncelikle destek vektör regresyon yöntemi kullanılmış olup kıyaslama amacıyla çok katmanlı sinir ağı ve çoklu doğrusal regresyon yöntemleriyle de tahminleme modeli kurulmuştur.

Destek vektör makineleri, örüntü tanımlama problemleri için Vapnik (1995) tarafından geliştirilen bir makine öğrenmesi yöntemidir. ϵ -duyarsız kayıp fonksiyonu ile destek vektör makineleri, regresyon analizi için de kullanılabilir olmuştur. Bu yöntem, çekirdek hilesi yardımıyla doğrusal olmayan problemler için de kullanılabilir (Vapnik, 1995; Vapnik vd., 1997; Müller vd., 1997). Geleneksel makine öğrenmesi yöntemleri, deneysel (ampirik) risk minimizasyonu prensibiyle eğitim verisindeki hatayı küçükmeye çalışırken yapısal risk minimizasyonu temeli üzerine kurulan destek vektör makinesi, beklenen riskin üst sınırını enküçükmeye çalışmaktadır. Bu durum destek vektör makinesinin genelleştirme kabiliyetini olumlu yönde etkilemektedir.

Bu çalışma yukarıda belirtilen kapsam ve amaç doğrultusunda on bölümden oluşmaktadır.

İkinci bölümde, öncelikle hava kirliliğinin meteorolojik değişkenlerle olan ilişkisini istatistiksel yöntemlerle inceleyen çalışmalardan bahsedilmiştir. Daha sonra ise bu ilişkinin kullanılmasıyla makine öğrenmesi yöntemleriyle yapılan tahminleme çalışmalarına yer verilmiştir.

Çalışmanın üçüncü bölümünde temel hava kirleticilerinin özellikleri, kaynakları ile insan ve çevre sağlığına olan etkilerinden söz edilmiştir.

Dördüncü bölümde doğrusal regresyon modelinin kurulması ve bu modelin parametrelerinin en küçük kareler yöntemi ile tahminlenmesi anlatılmıştır. Daha sonra tahminlere ve katsayılara ait standart hatanın ve belirlilik katsayısının hesaplanmasına yer verilmiştir.

Beşinci bölümde öncelikle yapay sinir hücrelerinin ve sinir hücreleri ile oluşturulan ağların yapısı incelenmiş, sonrasında sinir ağlarında kullanılan öğrenme yöntemlerine yer

verilmiştir. Bu bölümde son olarak katman sayısı, sinir hücreleri ve katmanlar arasındaki besleme yönü, öğrenme yöntemleri gibi tercihlerin farklı şekillerde kullanılmasıyla oluşturulan yapay sinir ağı modelleri anlatılmıştır.

Çalışmanın altıncı bölümünde destek vektör makinelerinin teorik temelleri anlatılarak sınıflandırıcı destek vektör makinesi ve destek vektör regresyonun yapısı incelenmiştir. Ayrıca doğrusal olmayan problemlerin çözülebilmesi için kullanılan çekirdek hilesi de bu bölümde incelenmiştir.

Yedinci bölümde ise öncelikle veri setine ait tanımlayıcı istatistikler ve veri setinin uygulamaya uygun hale getirilmesi için yapılan ön işlemler anlatılmıştır. Daha sonra ise dördüncü, beşinci ve altıncı bölümde anlatılan yöntemlerin uygulamaya geçirilebilmesi için kullanılan WEKA yazılımının anlatımına yer verilmiştir. Bu bölümde son olarak genel tahminleme modeline ait girdi ve çıktılar, eğitim ve test veri kümesinin belirlenmesi, tahminleme modellerini değerlendirmek amacıyla kullanılacak olan performans göstergeleri, kullanılan yöntemlere ait parametrelerin tahminleme modellerine olan etkisi ve bu parametrelerin belirlenmesine yer verilmiştir.

Çalışmanın sekizinci bölümünde test sonuçları verilmiş olup bu test sonuçlarına ilişkin bulgular tartışılmıştır.

Son bölümde, uygulamalar ile elde edilen sonuçlara ve bu konu ile yapılabilecek sonraki çalışmalara yönelik önerilere yer verilmiştir.

2. LİTERATÜR ARAŞTIRMASI

Hava kirliliğinin seviyesi; meteorolojik faktörler, kirletici kaynakları, topografya ve atmosferik kimyasal süreçlerden etkilenmektedir (Fahimeh ve Azadeh, 2012). Meteorolojik parametreler; atmosferdeki hava kirleticilerinin doğal temizliği, taşınımı ve yayılımı için önemli bir faktördür (Verma ve Desai, 2008). Dolayısıyla atmosferdeki gaz ve partikül konsantrasyonunun yoğunluğu rüzgar, sıcaklık, basınç, bulutluluk ve bağıl nem gibi meteorolojik faktörlerden etkilenir (Jacobson, 2005). Meteorolojik parametreler ile hava kirliliği arasındaki bu ilişkiyi göstermek adına literatürde istatistiksel yöntemlerle çeşitli çalışmalar yapılmıştır.

Bridgman vd. (2002) çalışmalarında kükürdioksit (SO₂) konsantrasyonunun meteorolojik parametrelerle ilişkisini araştırmış ve sonuç olarak platoda SO₂ konsantrasyonunun düşük hava sıcaklığı, yüksek bağıl nem ve düşük rüzgâr hızıyla kuvvetli ilişkisi olduğu belirtmişlerdir.

Miyazaki ve Yamaoka (1991) Osaka'da ölçülen değerler üzerinden yaptıkları çalışmalarında meteorolojik faktörlerle toz konsantrasyonu arasında iyi bir korelasyon yakalamışlardır.

Turalioğlu vd. (2005) çalışmalarında partikül madde (PM) ve SO₂ konsantrasyonunun düşük sıcaklık ile güçlü, düşük rüzgâr hızı, yüksek basınçla anlamlı, yağış miktarının azlığı ve yüksek bağıl nem ile zayıf bir ilişkisi olduğunu göstermişlerdir.

Çuhadaroğlu ve Demirci (1997) ise çoklu doğrusal regresyon (MLR) yöntemiyle kirletici konsantrasyonu ve meteorolojik faktörler arasındaki ilişkiyi incelemişlerdir. Bu çalışma sonucunda orta ve düşük seviyede ilişki tespit edilmiştir.

Dominic vd. (2012) Pearson korelasyon analizini kullandıkları çalışmalarında hava kirleticileri konsantrasyonu ile meteorolojik faktörler (bağıl nem, sıcaklık ve rüzgar hızı) arasında anlamlı bir ilişki olduğunu belirtmişlerdir.

Zhang vd. (2015) kriter hava kirleticileri konsantrasyonun; rüzgâr hızı, rüzgâr yönü, bağıl nem ve sıcaklıkla olan ilişkisini Çin'in üç büyük şehri için göstermişlerdir.

Bhaskar ve Mehta (2010) tarafından yapılan çalışmada partikül maddelerin yayılımıyla meteorolojik faktörler arasında önemli bir ilişki olduğu belirtilmiştir.

Literatürde, hava kirliliği konsantrasyonunu tahminleme çalışmalarında yapay sinir ağları (ANN) ve MLR'nin kullanımı yoğun bir şekilde görülmektedir.

Boznar vd. (1993) bir termal enerji santralının yakın çevresindeki SO₂ konsantrasyonunun yarım saat sonraki değerini; sıcaklık, rüzgâr, nem, solar radyasyon ve SO₂ konsantrasyonu gibi 37 adet girdi kullanarak ANN ile tahminlemeye çalışmıştır. Çalışma sonucunda karmaşık coğrafi yapılarda deterministik yaklaşımların güvenilir sonuçlar üretmediği, ANN modelinin ise ümit verici sonuçlar ürettiği belirtilmiştir.

Comrie (1997) ABD'nin 8 farklı kenti için kirletici konsantrasyonu ve meteorolojik verileri kullanarak günlük ortalama ozon (O₃) değerini, ANN ve MLR ile tahminlemeye çalışmıştır. Çalışmada, aralarında belirgin bir fark olmamakla beraber ANN modellerinin daha başarılı çıktılar ürettiği sonucuna ulaşılmıştır.

Kaminski vd. (1999) 8 adet meteorolojik faktör ile PM ve SO₂ konsantrasyonlarını kullanarak ANN ile günlük kirlilik değerini tahminlemeye çalışmıştır. Uygulamada öncelikle meteorolojik veri üzerinde temel bileşenler analizi (PCA) yöntemiyle boyut azaltma işlemi gerçekleştirilmiştir. Sonuç olarak ANN modeli, ortalama %4,3 hata değeriyle tahminler yapmıştır.

Gardner ve Dorling (1999) Londra kent merkezindeki azotoksitlerin (NO_x) saatlik konsantrasyonunun tahminlenmesi için bulutluluk, görünürlük, basınç, rüzgâr hızı gibi meteorolojik verileri kullanarak ANN ve MLR ile model oluşturmuştur. Sadece meteorolojik verilerin kullanıldığı modele kıyasla kirletici konsantrasyonunun da eklendiği model, yaklaşık %50 oranında daha başarılı olmuştur. Çalışmanın sonucunda, aralarında belirgin farklar olmamasına rağmen ANN ile MLR'ye göre daha iyi çıktılar elde edildiği belirtilmiştir.

Chelani vd. (2002) veri setinde kirletici deęerleri ve meteorolojik faktörleri kullanarak ANN ile havadaki PM₁₀ ve zehirli metallerin miktarını tahminlemeye çalışmışlardır. Çalışmada, tahminleme modelinin PM₁₀ ve zehirli metallerin deęerini düşük miktartlı hatalarla tahminledięi belirtilmiştir.

Kukkonen vd. (2003) ANN, doğrusal istatistiksel model ve deterministik tahminleme yöntemleriyle Helsinki'deki saatlik azotdioksit (NO₂) ve PM₁₀ konsantrasyonu tahminlemeye çalışmıştır. Çalışmada veri olarak kirletici deęerleri, trafik akışı ve 34 adet meteorolojik etmen kullanılmıştır. Çalışmanın sonucunda, ANN'nin dięer yöntemlere kıyasla daha iyi sonuçlar ürettięi ancak geleceęe dair kirlilięi azaltma senaryolarında kullanılmasının uygun olmadığı belirtilmiştir.

Grivas ve Chaloulakou (2006) meteorolojik faktörler ve kirletici deęerlerini kullanarak Atina'daki 4 farklı nokta için saatlik PM₁₀ konsantrasyonunun tahminlenmesi üzerine çalışma yapmıştır. Çalışmada; ilki tüm verilerin kullanıldığı, ikincisi girdi deęişkenlerinin genetik algoritma ile belirlendięi, sonuncusu ise sadece kirletici deęerlerinin kullanıldığı üç farklı ANN modeli oluşturulmuştur. Sadece kirletici deęerlerinin kullanıldığı model en kötü performansı vermiştir. Deęişkenlerinin genetik algoritma ile belirlendięi model ve tüm verilerin kullanıldığı model birbirlerine çok yakın deęerler üretmiştir. Sonuç olarak girdi sayısının çok fazla olduęu modeller için hesaplama maliyetini azaltmak adına genetik algoritma-ANN melezi yöntem önerilmiştir.

Sousa vd. (2007) bir sonraki günün saatlik O₃ konsantrasyonunu tahminlemek amacıyla çok katmanlı algılayıcı (MLP) yöntemini kullanmıştır. Kirletici konsantrasyonu ve meteorolojik ölçümlerin girdi olarak kullanıldığı çalışmada veri setinin boyutu PCA ile azaltılmıştır. Sonuç olarak sinir aęının, kıyaslama amacıyla kullanılan MLR'den daha iyi sonuçlar ürettięi ve PCA'nın, verinin karmaşıklığını azaltarak her iki yöntem içinde model başarımını arttırdığı belirtilmiştir.

Diaz-Robles vd. (2008) PM₁₀ konsantrasyonunu; kirletici verileri ile meteorolojik faktörleri kullanarak MLR, ARIMAX ve ANN ile modelleyip tahminlemeye çalışmıştır. Bu üç yöntem birbirine yakın sonuçlar üretirken en başarılı sonucu ARIMAX modeli vermiştir.

Son olarak ARIMAX-ANN melezi yöntemi kullanılmış olup ortalama mutlak hata değerinde %66'lık bir iyileştirme sağlanmıştır.

Singh vd. (2012) PM₁₀, SO₂ ve NO₂ konsantrasyonu ile sıcaklık, bağıl nem ve rüzgâr hızını girdi olarak kullandıkları çalışmalarında, hava kirliliğini tahminlemek için doğrusal ve doğrusal olmayan modeller kurmuştur. Bu modellerde; kısmi en küçük kareler regresyon, çoklu polinom regresyon ve farklı ANN modelleri kullanılmıştır. Kullanılan ANN modelleri; MLP, RBF (radyal tabanlı fonksiyon) ağı ve genelleştirilmiş regresyon sinir ağıdır (GRNN). Doğrusal olmayan modellerin kısmi en küçük kareler regresyon modeline göre daha iyi sonuçlar ürettiği ve en başarılı yöntemin yapay sinir ağı olduğu görülmüştür. ANN modelleri içerisinde ise en başarılı çıktılar, GRNN modeli üretmiştir.

Ul-Saufie vd. (2013) bir, iki ve üç günlük ortalama PM₁₀ konsantrasyonunu MLR ve ANN ile tahminlemeye çalışmıştır. Çalışmada girdi olarak PM₁₀, SO₂, NO₂ ve CO (karbondioksit) konsantrasyonu ile rüzgâr hızı, bağıl nem ve sıcaklık değerleri kullanılmıştır. İlk aşamada standart verilerle model kurulurken ikinci aşamada PCA ile boyutu azaltılmış olan veri setiyle tahminleme modeli kurulmuştur. Çalışmada PCA ile boyutu azaltılan veriyle daha iyi sonuçlar elde edilmiştir. Bir günlük tahminler için ANN daha başarılıyken iki ve üç günlük tahminlerde MLR daha iyi çıktılar üretmiştir. Son olarak bu üç güne ait ortalama tahmin başarısı incelendiğinde MLR modelinin en iyi sonuçları ürettiği görülmüştür.

Prasad vd. (2016) çalışmalarında kirletici konsantrasyonu ve meteorolojik faktörlerle günlük hava kirliliğini tahminleme modeli kurmuştur. Çalışmada öncelikle gereksiz verilerin ayıklanması için eşdoğrusallık (collinearity) testi yapılmış ve girdi kümelerini belirlemek adına ileriye doğru seçim yöntemi kullanılmıştır. Daha sonra belirlenen girdi kümeleriyle ANFIS (nöro-bulanık) modeli kurulmuş ve gerçek değerlere yakın tahminlerle başarılı sonuçlar elde edilmiştir.

Son yıllarda hava kirliliğinin tahminlenmesinde destek vektör makinelerinin (SVM) kullanımı önem kazanmıştır.

Lu vd. (2002) kirletici konsantrasyonlarını kullanarak saatlik hava kirliliği tahminleme modeli oluşturmuştur. Modeller destek vektör regresyon (SVR) ve RBF ağı ile kurulmuş olup 24 ve 168 saat için tahminleme yapılmıştır. SVR toplam 24 saat için %12,75, toplam 168 saat için de %11,28 daha az ortalama hatayla tahminler yapmıştır.

Lu ve Wang (2005) Hong-Kong için NO_x , PM, SO_2 , CO gibi kirleticiler ve solar radyasyon, sıcaklık, rüzgâr hızı ve yönü gibi 5 meteorolojik ölçüm değeri kullanarak SVR ile Haziran ve Aralık ayları içerisindeki saatlik kirletici (PM, NO_2 ve NO_x) konsantrasyonu tahminlemesi yapmışlardır. Oluşturulan modelin sonuçları RBF ağından elde edilen sonuçlarla karşılaştırılıp SVR'nin daha başarılı olduğu belirtilmiştir.

Osowski ve Garanty (2007) günlük hava kirliliğini tahminlemeyi amaçlayan çalışmalarında; girdi olarak tahminlenen günden önceki 3 günün kirletici konsantrasyonu, meteorolojik ölçüm değerleri, içinde bulunulan günün yılın kaçınıcı ayı ve haftanın kaçınıcı gününe karşı geldiği bilgisini kullanmışlardır. İlk aşamada tahminleme başarısını arttırmak için zaman sıralı kirletici konsantrasyonları üzerinde dalgacık ayrıştırma (wavelet decomposition) yöntemi uygulanmıştır. Daha sonra ise, SVR ile oluşturulan tahminleme modeli MLP ile kıyaslanmış ve SVR'nin gerçek değerlere daha iyi uyum sağladığı gösterilmiştir.

Lu ve Wang (2008) O_3 kirliliğinin olduğu günleri tahminleme amacıyla standart SVM (s-SVM), SVR ve maliyet duyarlı SVM (cs-SVM) yöntemlerini, dengesiz (imbalanced) bir veri seti üzerinde kullanılarak karşılaştırmışlardır. SVR ile sınıflandırma işlemi, sayısal tahminler yapıldıktan sonra bu tahminlerin O_3 eşik değerine göre sınıflara atanmasıyla yapılmıştır. cs-SVM kirliliği günleri tahminlemede biraz daha iyi sonuçlar verirken, O_3 kirliliğinin olmadığı günlerde s-SVM'ye kıyasla daha başarısız olmuştur.

Wang vd. (2008) Hong Kong'da ölçülen hava kirliliği değerlerine göre SVM tabanlı çevrimiçi tahminleme sistemi geliştirmiştir. Dinamik bir yapıya sahip olan bu modelde, veri setine eklenen yeni girdi destek vektörse model güncellenmekte değilse modelin yapısı değişmemektedir. Çevrimiçi-SVM, standart yapıdaki SVM ile kıyaslandığında daha başarılı sonuçlar üretmiştir.

Chelani (2010) son iki günün O₃ miktarı ve meteorolojik verileri kullanarak bir sonraki günün maksimum O₃ konsantrasyonunu tahminlemeye çalışmıştır. Çalışmada tahminleme modelleri SVM, MLP ve MLR ile kurulmuştur. Sonuç olarak SVM'nin MLR'den %52, MLP'den ise %35 daha az ortalama hatayla tahminleme yaptığı gösterilmiştir.

Ortiz-Garcia vd. (2010) çalışmalarında komşu istasyonlarda ölçülmüş O₃ miktarı, geçmiş O₃ miktarı ve solar radyasyon, sıcaklık gibi meteorolojik faktörler üzerinden SVR yöntemiyle Madrid bölgesindeki saatlik O₃ miktarı tahminlemesi yapmışlardır. Bu modelin saatlik O₃ miktarı tahmininde başarılı olduğu gösterilmiştir. Ayrıca çalışmada MLP ve SVR karşılaştırılmış olup SVR ile daha iyi sonuçlar elde edilmiştir.

Zheng vd. (2010) termal santralden yayılan NO_x konsantrasyonu tahminlemek için en küçük kareler-SVR ve GRNN yöntemlerini kullanmıştır. Çalışma girdi olarak kömürün kalitesi, kömür değirmeninin hızı, püskürtülen havanın hızı ve kazan yükü kullanılmıştır. Sonuç olarak en küçük kareler-SVR % 0,21 GRNN ise % 1,6217 göreceli ortalama hatayla tahminler yapmıştır.

Feng vd.nin (2011) Pekin metropolü için yaptıkları çalışmada O₃ ve meteorolojik koşulların (sıcaklık, nem, rüzgâr hızı ve UV radyasyon) girdi olarak kullanıldığı modelde genetik algoritmayla optimize edilmiş olan geriyayılımlı sinir ağı (GABPNN) ve SVM melezi kullanılarak O₃ tahminlemesi yapılmıştır. Sonuç olarak geriyayılımlı sinir ağı, GABPNN ve SVM-GABPNN kıyaslandığında en başarılı ve kararlı tahminlerin SVM-GABPNN ile yapıldığı gösterilmiştir.

Sánchez vd (2011) ile Nieto vd. (2013) İspanya'nın iki farklı şehri için SVR ve MLP yöntemleriyle bir regresyon modeli kurarak hava kirleticileri (SO₂, NO_x, CO, O₃, PM₁₀) arasındaki bağımlılık ilişkisini araştırmışlardır. Sonuç olarak en yüksek korelasyon katsayısı Pearson VII çekirdek fonksiyonunu kullanan SVR ile elde edilmiştir.

Yang vd. (2011) günlük hava kirliliğinin SVR ile tahminlenmesinde çekirdek fonksiyonlarının etkisini araştırmıştır. Tahminleme modelinde girdi olarak içinde bulunan ve bir önceki günün kirletici konsantrasyonu ile meteorolojik ölçüm değerleri kullanılmıştır.

Sonuç olarak doğrusal ve RBF çekirdek fonksiyonlarının gözlemlenen değerlerle daha iyi bir uyum yakaladığını göstermişlerdir.

Lin vd. (2011) logaritma-SVR ile hava kirliliği tahminleme çalışması yapmışlardır. Çalışmada parametreler yapay bağıklık algoritmasıyla belirlenmiştir. Çalışmanın amacı, logaritma-SVR ile ham veri, ölçeklendirilmiş veri ve normalleştirilmiş veri kullanılan SVR'nin kıyaslanmasıdır. Logaritma-SVR ile en düşük hata oranı elde edilmiştir.

Yeganeh vd. (2012) çalışmalarında SVR ve kısmi en küçük kareler-SVR yöntemleriyle Tahran metropolü için günlük ve saatlik CO konsantrasyonu tahminlemesi yapmıştır. İki yöntemin de başarılı sonuçlar verdiği görülse de kısmi en küçük kareler-SVR yöntemi, SVR'ye kıyasla saatlik tahminlerde ortalama %8,6; günlük tahminlerde ise ortalama %19,6 daha az hatayla sonuçlar üretmiştir. Veri setinde PM₁₀, hidrokarbonlar (THC), NO_x, metan (CH₄), SO₂ ve O₃ olmak üzere 6 kirletici ve ek olarak basınç, sıcaklık, rüzgâr yönü, rüzgâr hızı ve bağıl nem gibi meteorolojik etmenler kullanılmıştır.

Sotomayor-Olmedo vd. (2013) Mexico City'nin O₃, NO₂ ve PM₁₀ konsantrasyonlarını SVR ile tahminleyerek çekirdek fonksiyonlarının performanslarını karşılaştırmıştır. Sonuç olarak RBF çekirdeğinin polinom ve spline çekirdek fonksiyonlarına nazaran daha başarılı olduğu gösterilmiştir.

Hajek ve Olej (2015) kirletici konsantrasyonları ve meteorolojik ölçümleri kullanarak iki farklı istasyon için hava kirletici indeksinin sınıfını belirlemeye çalışmıştır. Çalışmada kullanılan yöntemler SVR, ANFIS, RBF ağı, MLP olup en başarılı sınıflandırmayı SVM ve ANFIS gerçekleştirmiştir.

3. HAVA KİRLİLİĞİ VE KİRLETİCİLER

EPA (ABD Çevre Koruma Ajansı) (1992); hava kirliliğini, kirletici maddelerin insan sağlığı ve refahını tehdit edecek veya diğer zararlı çevresel etkilere neden olacak miktarlarda atmosferde bulunması şeklinde tanımlamıştır. Hava kirleticiler, havanın doğal yapısını değiştiren gaz, sıvı veya katı fazlarda bulunabilen kimyasal maddelerdir.

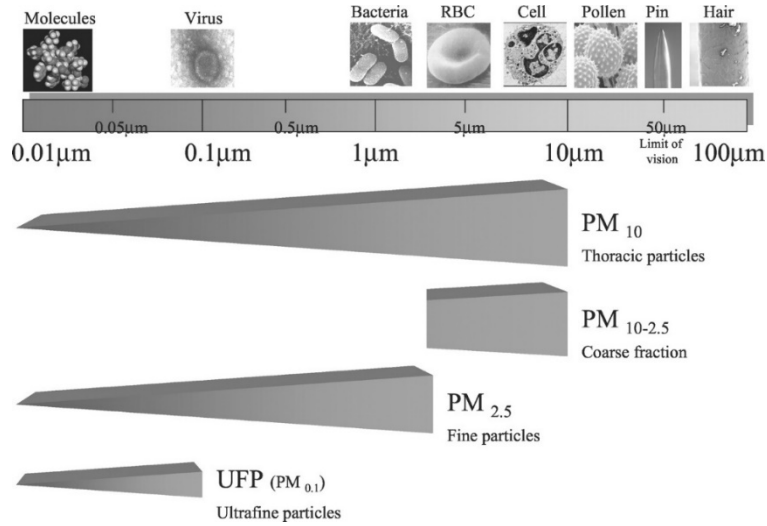
EPA'nın "Clean Air Act of 1970" çalışmaları sonucunda oluşturulan Ulusal Hava Kalitesi Standartları'nda kriter hava kirleticileri olarak adlandırılan altı temel kirleticiden bahsedilmiştir (Vallero, 2007):

1. Partikül Maddeler (PM)
2. Karbon Monoksit (CO)
3. Kükürt Dioksit (SO₂)
4. Azot Dioksit (NO₂)
5. Ozon (O₃)
6. Kurşun (Pb)

EPA'ya (1994) göre Kirlilik Standardı İndeksi (Pollutant Standards Index) hesaplanırken kriter hava kirleticilerinden beş tanesi kullanılmaktadır. Bu beş kirletici; partikül madde (toz, duman, diğer partiküller), kükürtdioksit, karbonmonoksit, azotdioksit ve ozondur. T.C. Çevre ve Şehircilik Bakanlığınca EPA Kirlilik Standardı İndeksi, ulusal mevzuatımız ve sınır değerlerimize uyarlanarak Hava Kalitesi İndeksi'nin (Air Quality Index) hesaplanmasında yine bu beş kirletici kullanılmıştır.

3.1. Partikül Maddeler (PM)

Partikül madde, havada asılı halde bulunan çok küçük tanecikli sıvı ve katı maddelerin heterojen karışımıdır. Bu karışım, asit (nitrat ve sülfatlar), organik kimyasallar, metaller ve toz parçacıkları gibi çeşitli bileşenlerden oluşmakta ve Şekil 3.1'de gösterildiği gibi 0,01 µm ile 100 µm arasında değişen boyutlarda bulunmaktadır. (Brook vd., 2004).

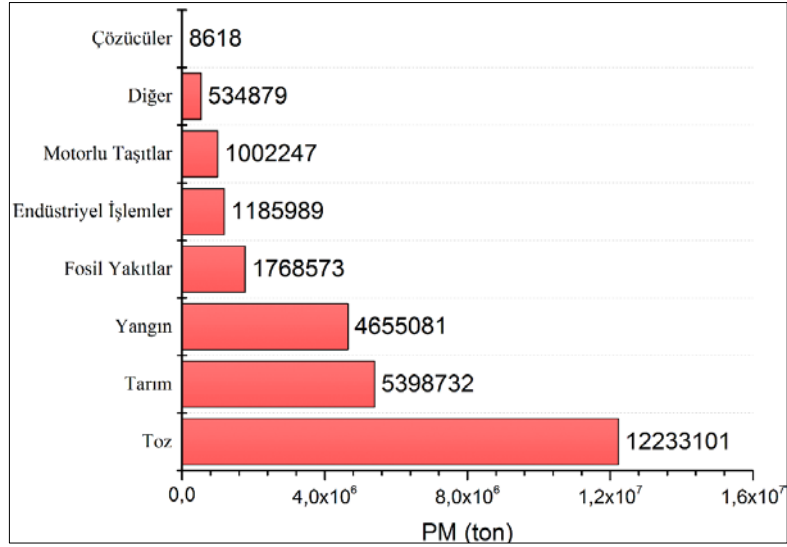


Şekil 3.1. Partikül madde boyutları

Parçacıkların boyutu, neden olabilecekleri potansiyel sağlık problemlerini doğrudan etkilemektedir. Boyutu 10 µm'den küçük olan partikül maddeler ağız ve burun yolunu aşır akciğerlere kadar ilerleyebilmektedir. EPA (2013), sağlık problemlerine neden olabilecek partikül maddeleri boyutlarına göre iki sınıfa ayırmıştır:

- **PM₁₀**: Kaba partikül olarak adlandırılmakta ve boyutu 10 µm ile 2,5 µm arasında olup solunabilmektedirler.
- **PM_{2,5}**: Boyutu 2,5 µm'den küçüktür ve ince partiküller olarak adlandırılmaktadır. Solunduklarında alveollere kadar ulaşabilmektedirler.

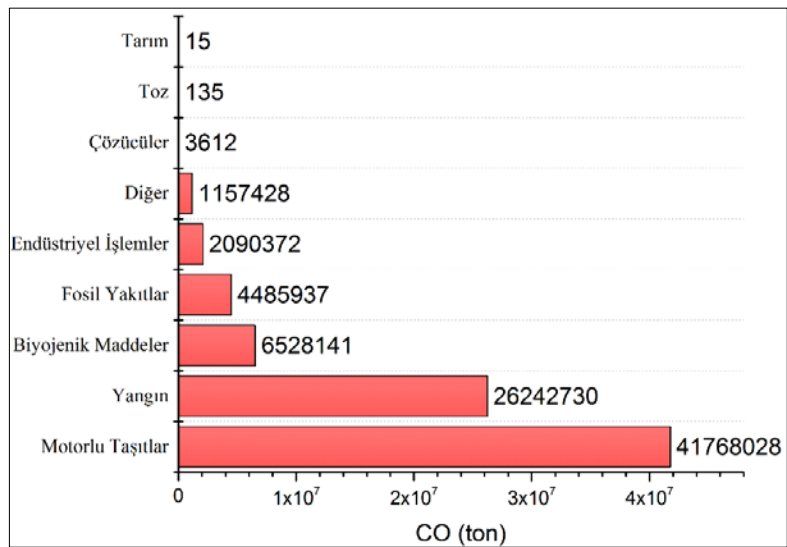
Bu parçacıkların kaynağı büyük oranda tozlar, zirai çalışmalar, orman yangınları, yakıt tüketimi ve endüstriyel çalışmalardır. EPA (2011 a) verilerine göre Amerika Birleşik Devletleri'ndeki toplam PM emisyonunun kaynaklarına göre dağılımı Şekil 3.2'de gösterilmiştir. Solunum yollarına alınan partikül maddelerin boyutu 10µm'den büyük olan kısmı burun ve üst yutakta tutulmakta, 10 µm'den küçük olan kısmı ise bronşlarda birikirken, 1-2 µm çapındakiler alveollerde toplanmakta, 0,5 µm çapındaki partiküller ise alveollerden kılcal damarlar yoluyla diğer organlara yayılmaktadır (Bayram vd., 2006).



Şekil 3.2. PM emisyonunun kaynaklarına göre dağılımı

3.2. Karbonmonoksit (CO)

Karbonmonoksit renksiz, kokusuz ve tatsız bir gaz olup karbon bazlı yakıtların eksik yanması sonucunda açığa çıkar (Atımtay vd, 2010). Özellikle kentsel bölgelerdeki CO emisyonlarının büyük kısmı motorlu taşıtlardan kaynaklanır ve CO konsantrasyonu en üst değerlerine genellikle yoğun araç trafiğinin yaşandığı bölgelerde ve saatlerde ulaşır (Gómez-Perales vd., 2004). EPA (2011 b) verilerine göre Amerika Birleşik Devletleri'ndeki toplam CO emisyonunun kaynaklarına göre dağılımı Şekil 3.3'te gösterilmiştir.

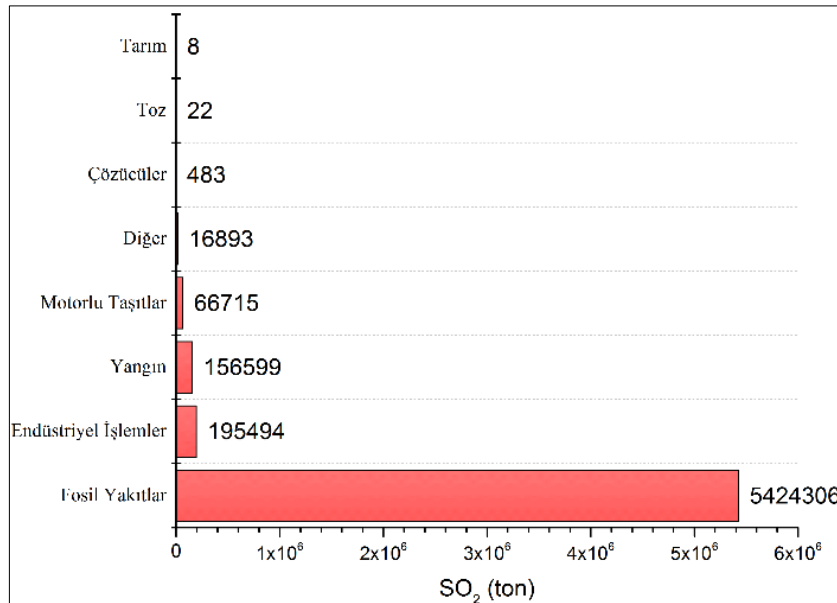


Şekil 3.3. CO emisyonunun kaynaklarına göre dağılımı

CO'in elektron ilgisi O₂'den çok daha fazla olduğu için kandaki hemoglobin ve kaslardaki miyoglobin ile daha kolay ve daha fazla birleşir. Dolayısıyla oksijen taşıma kapasitesinde azalmaya neden olarak ölüme yol açar. Omaye (2002) sanayileşmiş toplumlarda ölümcül zehirlenmelerin %50'sinden fazlasını CO zehirlenmelerinin oluşturduğunu belirtmektedir.

3.3. Kükürtdioksit (SO₂)

Genel olarak katı ve sıvı yakıtların yanması sonucu yakıtta bulunan kükürttten dolayı oluşan SO_x gazlar grubundan olan kükürtdioksit; renksiz, yanmaz, çözünebilir, tahriş edici ve patlamaz bir gazdır (Brook, vd., 2004). Kükürt; ham petrol, kömür gibi fosil yakıt yatakları ve alüminyum, bakır, çinko, kurşun, demir gibi maden cevherlerinde bol miktarlarda bulunmaktadır. Bu nedenle SO₂ gazları petrol, kömür gibi kükürt içeren katı ve sıvı yakıtların yanması, petrolden benzin ekstraksiyonu ve maden cevherlerinden metal zenginleştirilmesi gibi işlemler sonucunda oluşur. Atmosfere salınan SO₂'nin büyük bir kısmı, özellikle kömürün yakıt olarak kullanıldığı termik santraller olmak üzere elektrik üretiminden kaynaklanır (Atımyay vd., 2010). EPA (2011 c) verilerine göre Amerika Birleşik Devletleri'ndeki toplam SO₂ emisyonunun kaynaklarına göre dağılımı Şekil 3.4'te gösterilmiştir.

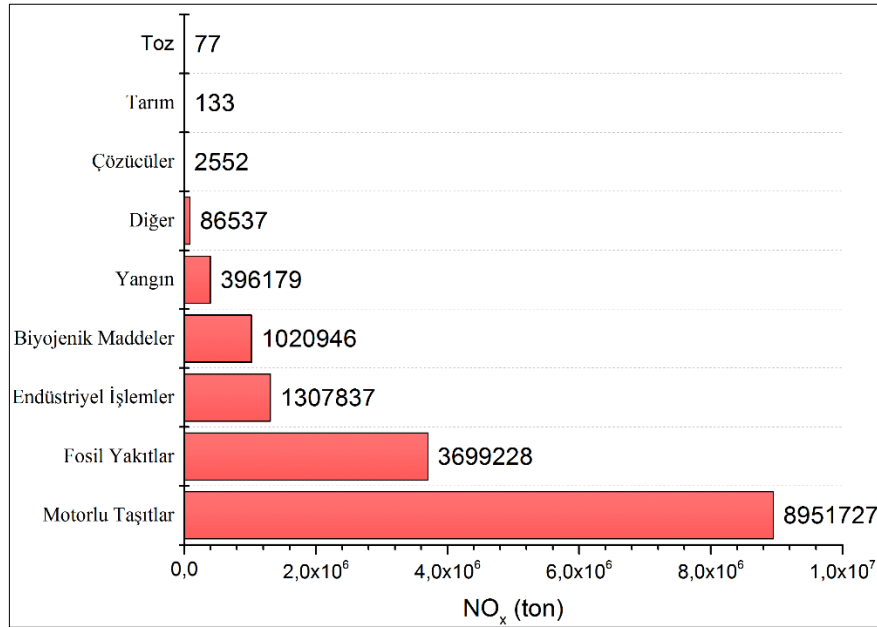


Şekil 3.4. SO₂ emisyonunun kaynaklarına göre dağılımı

SO₂'nin solunum yolu rahatsızlıklarına neden olduğu bilinmekte ve özellikle solunum sistemi hastaları için öldürücü olabildiği düşünülmektedir. Su ile teması halinde kolayca çözünebilmesiyle asit oluşturabilen SO₂, asit yağmurlarının da temel kaynağıdır.

3.4. Azotdioksit (NO₂)

Azotoksitler (NO_x) çok yüksek sıcaklıklardaki yanma tepkimesiyle oluşan oldukça reaktif gazlardır. Bir çok türü renksiz ve kokusuz olan azotoksitler, suda erimediklerinden üst solunum yollarında engellenmeden solunum yollarının en uç noktalarına kadar ulaşabilirler (Atımtay vd., 2010). EPA (2014) standartlarına göre azotoksitlerin göstergesi olarak azotdioksit (NO₂) kullanılır. NO₂ ana kaynakları egzoz gazları, fosil yakıtlar ve organik maddeler olarak sıralanabilir. Özellikle kentsel bölgelerde motorlu taşıt sayısındaki artışa bağlı olarak NO_x konsantrasyonlarında artış gözlenmekte, trafiğin yoğun olduğu bölgelerden uzaklaştıkça azalma görülmektedir (Han ve Neaher, 2006). EPA (2011 d) verilerine göre Amerika Birleşik Devletleri'ndeki toplam NO_x emisyonunun kaynaklarına göre dağılımı Şekil 3.5'te gösterilmiştir.



Şekil 3.5. NO_x emisyonunun kaynaklarına göre dağılımı

3.5. Ozon (O₃)

Ozon, azotoksitler (NO_x) ve uçucu organik bileşiklerin (VOC) ultraviyole ışının katalizörlüğünde reaksiyona girmesiyle oluşan yüksek derecede reaktif, renksiz, keskin kokulu bir gazdır (Rom, 2011; Brook vd., 2004).

Stratosferde bulunan ozon, ultraviyole ışınlarını emerek yeryüzündeki yaşam üzerinde olumlu bir etki yapar. Buna karşın troposferde bulunan ozon fotokimyasal sisin temel bileşenlerinden olup insan sağlığına solunum sistemi problemleri, göz, burun ve boğazda tahriş, astım gibi olumsuz etkilerde bulunur (Beckrich, 2013).

Çizelge 3.1’de beş temel kirleticinin ana emisyon kaynakları ve insan sağlığına olan etkileri özetlenmiştir.

Çizelge 3.1. Hava kirleticilerin ana kaynakları ve insan sağlığına olan etkileri

Kirletici	Ana Emisyon Kaynağı	Etkileri
PM	Tozlar, zirai çalışmalar, yangınlar, yakıt tüketimi, endüstriyel çalışmalar	Kanser, solunum yolu ve kalp hastalıkları, dumanlı sis
SO ₂	Fosil yakıtlar, yangın, metallerin zenginleştirilmesi	Solunum yolu hastalıkları, asit yağmurları, dumanlı sis
CO	Motorlu taşıtlar, yangınlar	O ₂ taşıma kapasitesinde azalma, baş dönmesi, ölüm, sinir sisteminde aksaklıklar
NO ₂	Motorlu taşıtlar, fosil yakıtlar, yüksek sıcaklıklarda yanma tepkimesi	Solunum yolu ve göz hastalıkları, asit yağmurları, su kirliliği, dumanlı sis
O ₃	Trafik kaynaklı azotoksitler ve VOC’lerin güneş ışığıyla değişimi	Solunum yolu hastalıkları, göz, burun ve boğazda tahriş, astım, vücut direncinin düşmesi

4. DOĞRUSAL REGRESYON

Regresyon analizi, iki veya daha fazla değişken arasındaki ilişkiyi araştırmak ve bu ilişkinin kullanılmasıyla kestirim ve tahminleme gibi çıkarımlar yapabilmek için kullanılan istatistiksel bir araçtır (Montgomery vd., 2012). Regresyon analizinde, değişkenler arasındaki ilişkinin açıklanabilmesi için regresyon modeli olarak adlandırılan matematiksel bir model kurulur. Doğrusal regresyon modelleri, basit doğrusal regresyon ve çoklu doğrusal regresyon olmak üzere iki tipten oluşur. Basit doğrusal regresyon modeli, biri bağımlı (açıklanan) değişken diğeri ise bağımsız (açıklayan) değişken olmak üzere iki değişkenle kurulur. Çoklu doğrusal regresyon modelinde ise bağımlı değişken sayısı bir iken bağımsız değişken sayısı birden fazladır.

4.1. Basit Doğrusal Regresyon Modeli

Anakütle verilerinin basit doğrusal regresyon modeli; y bağımlı değişkeni, x bağımsız değişkeni göstermek üzere Denklem (4.1)'deki gibi ifade edilir.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4.1)$$

β_0 , regresyon modelinin sabit parametresi olduğundan $x = 0$ değerini aldığı anda regresyon doğrusunun dikey eksenini kestiği noktayı göstermektedir. β_1 regresyon parametresi, doğrusal denklemin eğimini yani x bağımsız değişkenindeki bir birimlik değişiminin y bağımlı değişkeninde gerçekleştireceği değişim miktarını ifade etmektedir. ε hata terimi ise (4.2) rassal olup normal dağılıma uymaktadır (Chatterjee ve Hadi, 2012).

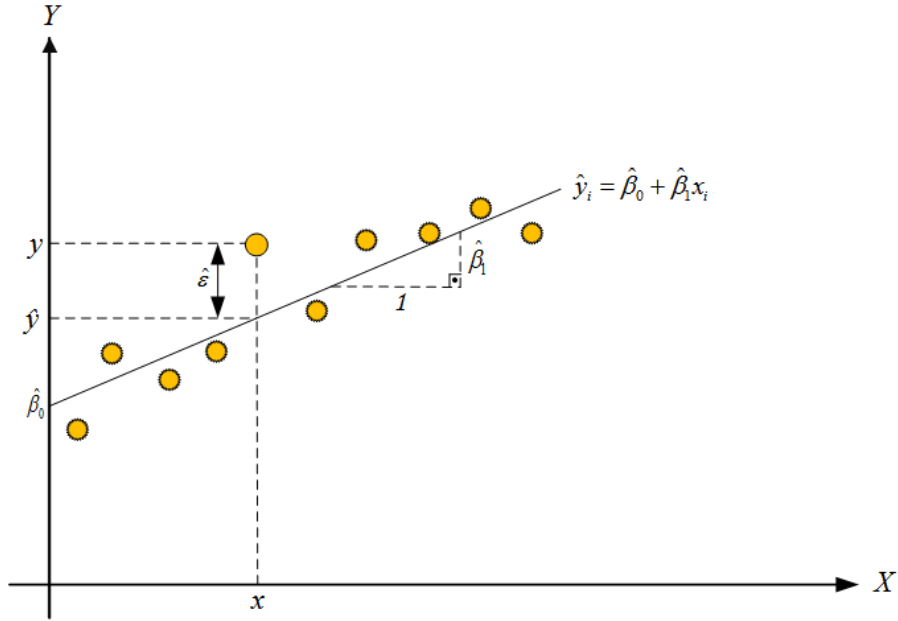
$$\varepsilon \sim N(0, \sigma^2) \quad (4.2)$$

Şekil 4.1'de örneklem kümesine ait gözlemlerin serpmeye diyagramı incelendiğinde (x_i, y_i) noktalarının bir doğru üzerine yerleşme eğilimi içerisinde oldukları ve bu gözlemler için sonsuz sayıda doğrusal fonksiyon çizilebileceği görülmektedir. Bu fonksiyonların en uygunu, y_i gözlem değerlerine diğer bir deyişle Denklem (4.3)'te gösterilen anakütle regresyon modeline en yakın \hat{y}_i tahmin değerlerini veren fonksiyon olacaktır.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (4.3)$$

Gerçekleşen ve tahmin edilen değerler arasındaki fark ile ε_i artık (residual) değerleri oluşacaktır.

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad (4.4)$$



Şekil 4.1. Basit doğrusal regresyon modeli

Bu artıklar, doğrunun üstünde kalan noktalar için pozitif (+), doğrunun altında kalan noktalar için negatif (-) değer alırken doğrunun üzerinde olan noktalar için sıfır (0) değerinde olmaktadır. Bu noktada amaç β_0 ve β_1 'in tahminlenmesiyle, veri noktalarına en iyi uyum gösteren regresyon modelini elde etmektir. Bu amaçla kullanılan en küçük kareler yöntemi, β_0 ve β_1 parametrelerini, Denklem (4.5)'te verilen $\hat{\varepsilon}_i$ artıklarının karelerinin toplamını en küçükleyerek belirler.

$$S(\beta_0, \beta_1) = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.5)$$

Denklem (4.3)'teki \hat{y}_i ifadesinin açık hali denklem (4.5)'te yerine yazılırsa aşağıdaki eşitlik elde edilir.

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (4.6)$$

Denklemi (4.6)'yı enküçükleyecek β_0 ve β_1 değerlerini bulmak için denklemin parametrelere göre kısmi türevleri alınır ve sıfıra eşitlenir.

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4.7)$$

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2x_i \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4.8)$$

Denklem (4.7) ve (4.8)'de gerekli işlemlerin yapılmasıyla aşağıdaki normal denklemler elde edilir.

$$\sum_{i=1}^n y_i = \hat{\beta}_0 \cdot n + \beta_1 \sum_{i=1}^n x_i \quad (4.9)$$

$$\sum_{i=1}^n y_i x_i = \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 \quad (4.10)$$

Normal denklemlerin ortak çözülmesiyle $\hat{\beta}_0$ ve $\hat{\beta}_1$ değerleri bulunur (Birkes ve Dodge, 1993).

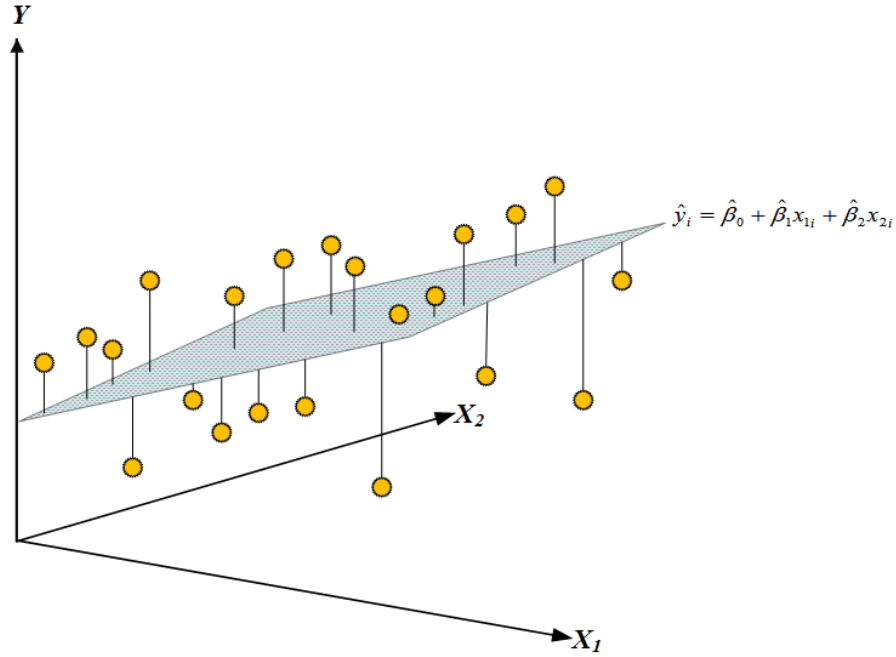
$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (4.11)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (4.12)$$

Denklem (4.11) ve (4.12)'de \bar{x} ve \bar{y} sırasıyla x_i ve y_i değerlerinin ortalamasını ifade etmektedir.

4.2. Çoklu Doğrusal Regresyon Modeli

Birden fazla bağımsız değişkeni olan doğrusal regresyon modeli, çoklu doğrusal regresyon (MLR) modeli olarak adlandırılmaktadır. Şekil 4.2’de iki bağımsız değişkene sahip çoklu regresyon modeli görselleştirilmiştir.



Şekil 4.2. Çoklu doğrusal regresyon modeli

x k adet bağımsız değişkeni, y bağımlı değişkeni ve ε hata terimini gösterirken anakütle regresyon modeli denklem (4.13)’deki gibi ifade edilir.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (4.13)$$

Çoklu regresyon modelinin matris gösterimi kullanılarak ifade edilmesi hem gösterimi hem de yapılacak işlemleri kolaylaştıracaktır (Montgomery vd., 2012). Çoklu regresyon modelinin matris gösterimi aşağıdaki gibidir.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ iken,}$$

$$Y = \beta X + \varepsilon \quad (4.14)$$

Örnekleme regresyon denkleminin matris gösterimi ise Denklem (4.15)'teki gibidir.

$$\hat{Y} = \hat{\beta}X \quad (4.15)$$

Çoklu doğrusal regresyon modelinde de regresyon katsayılarının tahminlenmesi en küçük kareler yöntemiyle yapılmaktadır. Bu yöntemle parametrelerin tahminlenmesi için öncelikle $\hat{\varepsilon}_i$ artıklarının karelerinin toplamının enküçüklenmesi gerekmektedir.

$$S(\beta_0, \beta_k) = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.16)$$

Bu ifadenin matris gösterimi aşağıdaki şekilde yazılır.

$$S(\beta) = \hat{\varepsilon}^T \hat{\varepsilon} = (Y - \hat{Y})^T (Y - \hat{Y}) \quad (4.17)$$

$\hat{\varepsilon}_i$ artıklarının karelerinin toplamı denklem (4.15) kullanılarak daha açık bir biçimde aşağıdaki gibi yazılabilir.

$$\begin{aligned} S(\beta) &= Y^T Y - \hat{\beta} X Y^T - \hat{\beta}^T X^T Y + \hat{\beta}^T X^T \hat{\beta} X \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T \hat{\beta} X \end{aligned} \quad (4.18)$$

Bu denklemi enküçükleyecek β değerlerini bulmak için denklemin parametrelere göre kısmi türevleri alınır ve sıfıra eşitlenir.

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = -2X^T Y + 2\hat{\beta} X^T X = 0 \quad (4.19)$$

Denklem (4.19) ile aşağıdaki normal denklem elde edilir.

$$\hat{\beta} X^T X = X^T Y \quad (4.20)$$

Normal denklemleri çözmek için eşitliğin iki tarafı $(X^T X)^{-1}$ ifadesi ile çarpılır ve $\hat{\beta}$ regresyon katsayısı vektörü Denklem (4.21) ile bulunur.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4.21)$$

4.3. Varyans Tahmini

Regresyon analizinde β katsayılarının tahminlerinin yanı sıra hipotez tezleri ve aralık tahminlerinde varyansın (σ^2) tahminine de gereksinim duyulmaktadır. σ^2 'nin tahmini artıkların kareler toplamı ile elde edilebilmektedir. Artıkların kareler toplamı (SS_e) aşağıdaki gibi ifade edilmektedir (Montgomery vd., 2012).

$$SS_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \hat{\epsilon}_i^2 \quad (4.22)$$

Artıkların kareler toplamının serbestlik derecesine bölünmesiyle elde edilen artıkların kareler ortalaması (MS_e) yani varyansı, σ^2 'nin bir tahminidir.

$$\hat{\sigma}^2 = MS_e = \frac{\sum \hat{\epsilon}_i^2}{n - k - 1} \quad (4.23)$$

MS_e 'in karekökünün alınmasıyla modelin standart hatası elde edilir ve bu standart hata $\hat{\sigma}$ ile gösterilir.

$$\hat{\sigma} = \sqrt{MS_e} \quad (4.24)$$

Tahminlerin standart hatası regresyon denklemi ile yapılan tahminlerin doğruluğunun bir ölçüsüdür. Bu değerın küçülmesi tahminlerin doğruluğunun arttığını göstermektedir.

Parametrelerin standart hatası ise Denklem (4.21)'den yola çıkılarak Denklem (4.25)'teki gibi hesaplanabilir.

$$Var(\hat{\beta}) = Var[(X^T X)^{-1} X^T Y] \quad (4.25)$$

Denklem (4.25)'te gerekli işlemler yapıldığında Denklem (4.26) elde edilir. Bu denklemde $\hat{\sigma}^2$ artıkların kareler ortalamasını temsil etmektedir (Sheather, 2009).

$$\text{Var}(\hat{\beta}) = \hat{\sigma}^2(X^T X)^{-1} \quad (4.26)$$

4.4. Belirlilik Katsayısı

Belirlilik katsayısı (R^2), bağımlı değişkendeki değişimlerin ne kadarının bağımsız değişkenlerdeki değişim miktarıyla açıklanabildiğinin ölçüsüdür. Belirlilik katsayısı, regresyon doğrusunun verileri ne kadar iyi temsil ettiğini gösterdiğinden dolayı uyum iyiliğinin ölçütü olarak da adlandırılabilir.

$$\text{Genel Kareler Toplamı} = SS_t = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4.27)$$

$$\text{Regresyon Kareler Toplamı} = SS_r = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (4.28)$$

olmak üzere belirlilik katsayısı Denklem (4.29)'daki gibi hesaplanır.

$$R^2 = \frac{SS_r}{SS_t} \quad (4.29)$$

$SS_t = SS_r + SS_e$ olduğundan belirlilik katsayısının değeri aşağıdaki ifadeyle de bulunabilir.

$$R^2 = 1 - \frac{SS_e}{SS_t} \quad (4.30)$$

[0,1] aralığında değer alabilen belirlilik katsayısının 1 değerini alması, bağımlı değişkendeki değişimin tamamının bağımsız değişkenlerle açıklanabildiğini, 0 değerini alması ise bağımsız değişkenlerin bağımlı değişkeni hiç bir şekilde açıklayamadığını göstermektedir.

Regresyon modelindeki bağımsız değişken sayısının artması belirlilik katsayısının değerini de arttırmaktadır. Ancak modele giren bağımsız değişken sayısının artması, her zaman tahminlerin standart hatasının azalacağı anlamına gelmemektedir. Özellikle modele giren değişkenler arasında yaşanabilir çoklu bağıntı nedeniyle kötüleşme görülebilmektedir. Bu durumun önüne geçebilmek için düzeltilmiş belirlilik katsayı kullanılmaktadır. Denklem (4.31)'de gösterildiği gibi düzeltilmiş belirlilik katsayı hesaplanırken örneklem büyüklüğü ve bağımsız değişken sayısı da dikkate alınmaktadır (Chatterjee ve Hadi, 2012).

$$R_{adj}^2 = 1 - \frac{SS_e/n - k - 1}{SS_t/n - 1} = 1 - \frac{n - 1}{n - k - 1} (1 - R^2) \quad (4.31)$$

5. YAPAY SİNİR AĞLARI

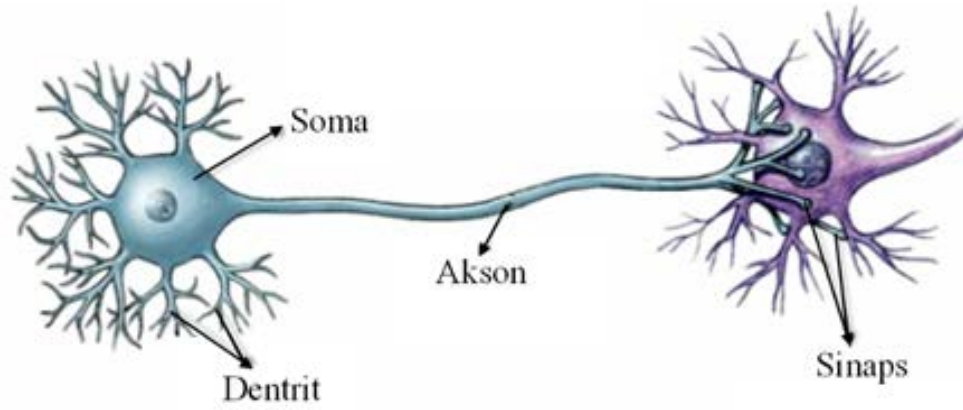
Yapay Sinir Ağları (ANN), biyolojik sinir ağlarının matematiksel modelle ifade edilmesiyle geliştirilmiş bir bilgi işleme sistemidir. Bu sistemdeki bilgi işleme süreci, yapının en küçük bileşeni olan sinir hücreleri (nöron) tarafından gerçekleştirilmektedir (Fausett, 1994). Bu noktada yapay sinir ağları, verilen örneklerle öğrenme işlemini gerçekleştirmekte ve bu işlem sonucunda kazandığı girdi-çıkı ilişkisini kavrama yeteneğiyle daha önce görülmeyen durumlar üzerinde çıkarım yapabilmektedir. Sinir hücrelerinin çeşitli biçimlerle birbirine bağlanmasıyla oluşan bu yapı, yapay sinir ağları olarak adlandırılmaktadır. Bir yapay sinir ağı, nöronlar arasındaki bağlantının örüntüsü (mimarisi), bağlantı ağırlıklarını belirleme yöntemi (öğrenme kuralı) ve aktivasyon fonksiyon ile karakterize edilir (Fausett, 1994).

5.1. Sinir Hücrelerinin Yapısı

Yapay sinir ağlarının başlangıç noktasını, McCulloch ve Pitts'in (1943) biyolojik sinir hücrelerini matematiksel olarak ifade etme çalışmaları oluşturmaktadır. Bu çalışmayla, basit sinir ağlarının mantıksal veya aritmetiksel fonksiyonları öğrenebilmesinin mümkün olduğu gösterilmiştir (Kriesel, 2005).

5.1.1. Biyolojik sinir hücreleri

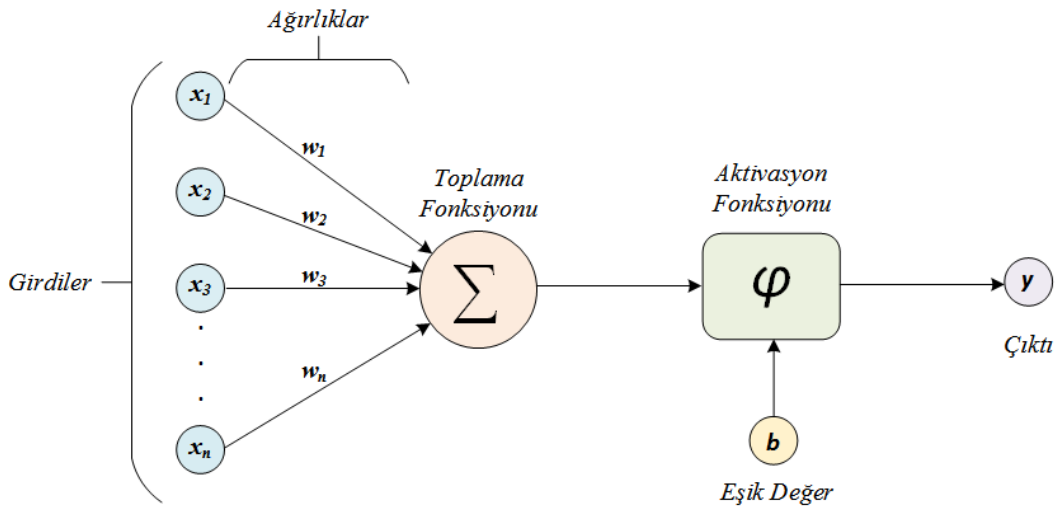
Temel işlevi bilgi aktarımı olan sinir hücreleri, sinir sistemlerinin yapıtaşlarıdır. Biyolojik bir sinir hücrelerinin; dentrit, soma (hücre gövdesi) ve akson olmak üzere sahip olduğu üç ana yapı Şekil 5.1'de gösterilmiştir. Dentritler sinir hücrelerine gelen uyarıları somaya aktarırlar. Gövdeye ulaşan bu uyarılar yeterli seviyede ise akson aracılığıyla diğer sinir hücrelerine iletilirler (Snell, 2010). Aksonlar, bu bilgi iletimini diğer sinir hücrelerinin dentritleri ile kurdukları boşluklu yapı olan sinapslar aracılığıyla gerçekleştirir. Bir sinir hücresi, birçok sinir hücresiyle sinaptik yapı oluşturabilir. Sinaptik yapıda uyarının iletildiği kısım presinaptik, uyarının alındığı kısım ise postsinaptik olarak adlandırılmaktadır (Noback vd., 2005).



Şekil 5.1. Biyolojik sinir hücresi

5.1.2. Yapay sinir hücreleri

McCulloch ve Pitts'in (1943) biyolojik sinir ağını matematiksel olarak ifade etme çalışmalarıyla tanımlanan yapay sinir hücreleri Şekil 5.2'deki gibi bir yapıya sahiptir. Yapay ve biyolojik sinir hücreleri arasında bir analogi kurulacak olursa, ağırlıklar ile sinaptik yapı, toplama fonksiyonu ile dentritler, aktivasyon fonksiyonu ile soma ve son olarak çıktı ile akson arasında eşleştirme yapılabilir.



Şekil 5.2. Yapay sinir hücresi

Şekil 5.2'de $x_i, i = 1, \dots, n$ şeklinde ifade edilen girdiler, işlenmemiş veya başka sinir hücrelerinden iletilen bilgileri temsil etmektedir. Yapay sinir hücresine alınan girdiler,

toplama fonksiyonuna iletilirken üzerinden geçtikleri bağlantının w_i ağırlıklarıyla çarpılırlar. Böylece girdilerin çıktı üzerinde oluşturacağı etki belirlenir.

Toplama fonksiyonları, yapay sinir hücresine iletilmiş olan toplam ağırlıklandırılmış girdiyi hesaplayan yapıdır. Bu fonksiyonları arasında en yaygın kullanıma sahip olan ağırlıklı toplama fonksiyonuyla net girdi aşağıdaki gibi hesaplanmaktadır:

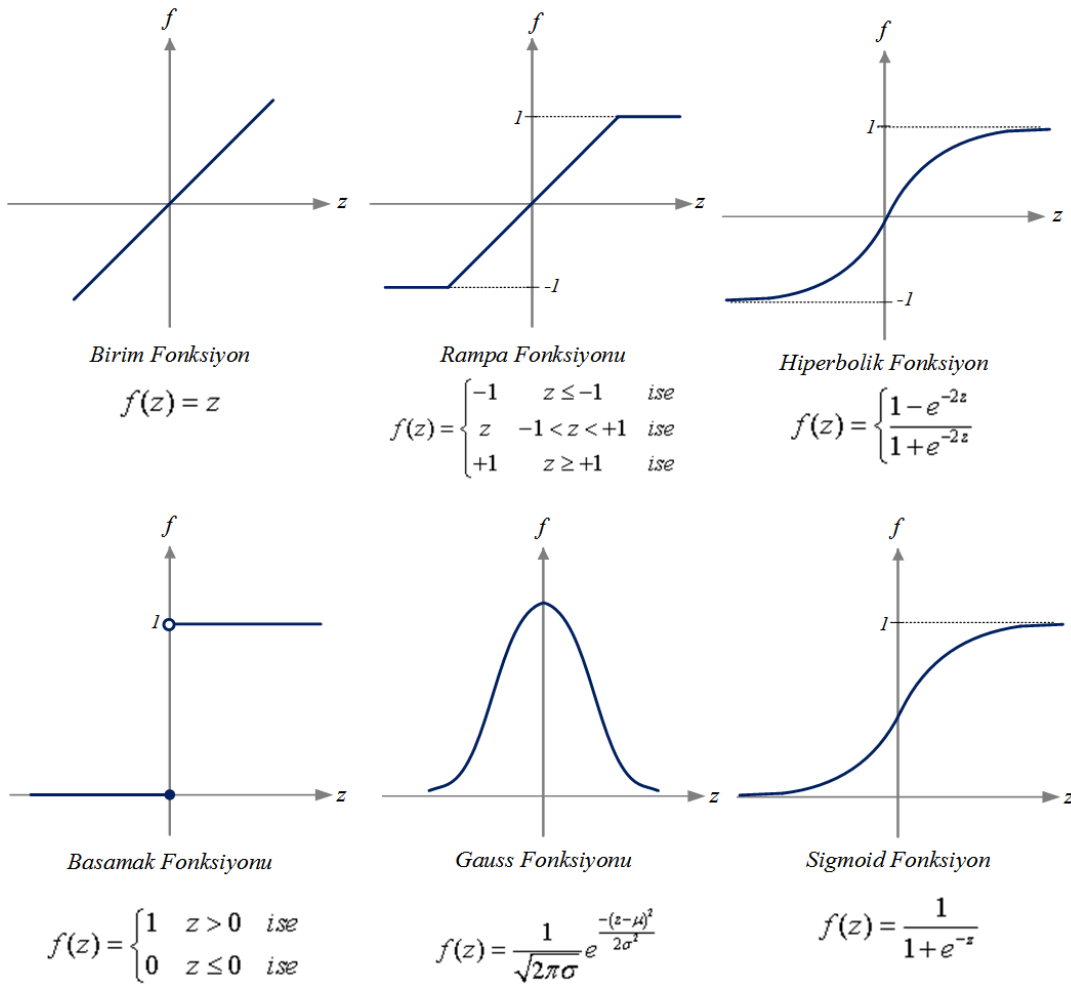
$$z = \sum_{i=1}^n w_i x_i \quad (5.1)$$

Toplama fonksiyonu ile hesaplanan net girdi, b eşik değeri de göz önünde bulundurularak aktivasyon fonksiyonu tarafından işlenip yapay sinir hücresinin çıktısına dönüştürülür. McCulloch-Pitts modelinde net girdiyi çıktıya dönüştürecek araç olarak Denklem (5.2)'de gösterilen basamak (step) fonksiyonu kullanılmıştır (Yegnanarayana, 1999). Bu fonksiyon ağırlıklandırılmış toplam girdiyi 0 ve 1 olmak üzere iki farklı değere dönüştürmektedir.

$$y = f(z) = \begin{cases} 1 & z > b \text{ ise} \\ 0 & z \leq b \text{ ise} \end{cases} \quad (5.2)$$

Bu noktada yapay nöronun işleyişini açıklamak gerekirse; öncelikle hücreye iletilen toplam sinyalin (z) eşik değeri (b) aşp aşamadığı incelenir. Eğer toplam sinyal eşik değeri aşarsa hücre aktif ($y = 1$) olacaktır. Aksi durumda ise düzeneği harekete geçirecek seviyede sinyal hücreye ulaşmamış demektir. Dolayısıyla hücre pasif ($y = 0$) durumdadır (Fausett, 1994).

Bunun yanında, teorik olarak birçok farklı fonksiyon tipi aktivasyon fonksiyonu olarak kullanılabilir. Bu fonksiyonların bazılarını Şekil 5.3'te yer verilmiştir (Kröse ve Smogt, 1996; Jones, 2004; Priddy ve Keller, 2005).

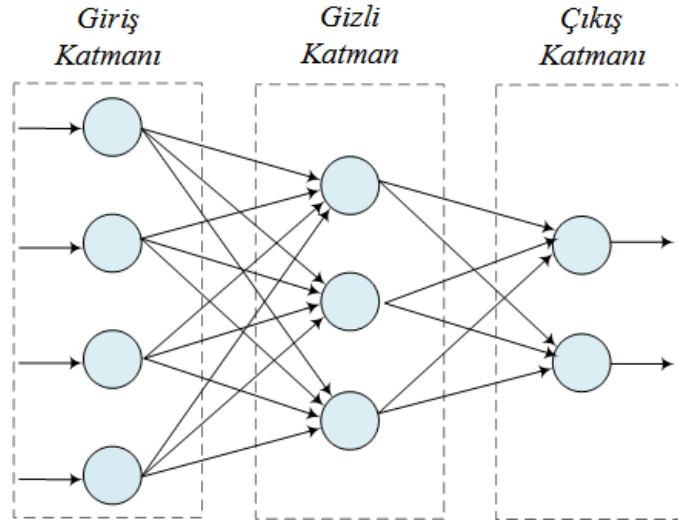


Şekil 5.3. Aktivasyon fonksiyonları

Aktivasyon fonksiyonuyla belirlenen çıktılar, yapay bir sinir hücresinin son süreç elemanlarıdır. Her bir nöron yalnızca bir çıktı değerine sahiptir ve bu çıktılar diğer nöronlar için girdi görevi görebilmektedir (Mehrotra vd., 1996).

5.2. Yapay Sinir Ağlarının Yapısı

Yapay sinir ağları, sinir hücrelerinin çeşitli bağlantılarla bir araya gelip grup halinde çalıştığı yapıdır. Bu yapı temel olarak giriş ve çıkış katmanını üzerinden şekillenmektedir. Sadece giriş ve çıktı katmanından oluşan yapılar genel olarak tek katmanlı algılayıcılar olarak adlandırılmaktadır. Ancak, bazı sinir ağlarında giriş ve çıkış katmanlarının arasında işlem görevini üstlenen gizli katmanlar bulunmaktadır. Gizli katmanın da dâhil olduğu çok katmanlı bir sinir ağı yapısı Şekil 5.4'te gösterilmiştir.



Şekil 5.4. Çok katmanlı bir sinir ağı yapısı

Giriş katmanı yapay sinir ağına gelen bilginin sisteme dâhil edildiği ve bir sonraki katmana yayıldığı kısımdır. Giriş katmanındaki sinir hücresi sayısı, veri kümesinde bulunan öznitelik kadardır. Gizli katman, giriş katmanından alınan bilgilerin işlendiği asıl kısımdır. Çıkış katmanı ise gizli katmandan alınan çıktıların yapay sinir ağının çıktısına dönüştürüldüğü katmandır. Bir sinir ağı modelinde, giriş ve çıkış katmanları bir adet iken gizli katman sayısında bir sınırlama mevcut değildir.

Yapay sinir hücreleri, çıktılarının diğer hücelere aktarım yönlerine göre ileri ya da geri beslemeli ağ yapıları oluştururlar. İleri beslemeli yapay sinir ağları, bir katmanı oluşturan sinir hücrelerinin çıktılarının, bir sonraki katmandaki sinir hücrelerine ağırlıklar üzerinden girdi olarak iletildiği sinir ağı tipidir. Dolayısıyla bilgi akışı tek yönlüdür (Yegnanarayana, 1999). İleri beslemeli yapay sinir ağları geniş anlamda örüntü tanıma ve tahminleme problemlerinin çözümde kullanılırlar (Annema, 2012).

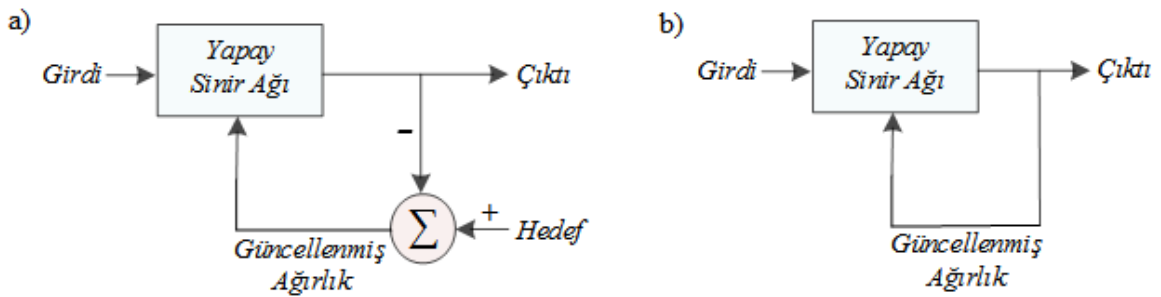
Geri beslemeli yapay sinir ağlarında, çıktısını kendisi dâhil bütün sinir hücrelerine girdi olarak verebilen en az bir sinir hücresi mevcuttur. Standart bir yapısı olmayan geri beslemeli yapay sinir ağlarında, geri besleme hücrenin kendisine ya da bir önceki katmandaki hücelere doğru olabilir (Yegnanarayana, 1999). Geri beslemeli yapay sinir ağları; sinyal işleme, doğrusal olmayan konveks programlama gibi problemlerin çözümünde kullanılırlar (He ve Xu, 2010).

5.3. Yapay Sinir Ağlarında Öğrenme Stratejileri

Genel olarak öğrenme, gözlemlerin yorumlanmasıyla kazanılan tecrübeye ilişkin kalıcı veya uzun süreli davranışsal sonuçlar doğuran bir kavramdır. Öğrenme eylemi canlılar için geçerli olduğu gibi makineler içinde geçerlidir. Bir makine öğrenmesi yaklaşımı olan ANN’de öğrenme süreci, canlılarınkine kıyasla doğrudan gerçekleşmektedir (Zurada, 1992).

Girdi-çıkı ilişkisi üzerinden gerçekleşen öğrenme işlemi, yapay nöronlar arasında kurulan bağlantıların ağırlıkları olarak saklanırlar. Ağırlıklar üzerinde iteratif güncellemeler yapılarak en iyi performansı sağlayacak modeli bulmak amaçlanmaktadır (Jain vd., 1996). En iyi performans, ağırlıklar üzerinde iyi bir genelleme yeteneği kazanmasını temsil etmektedir. İyi bir genelleme yeteneği kazanan sinir ağı, test verisi üzerinde de başarılı sonuçlar vermesi beklenir. Yapay sinir ağları için çoğunlukla kullanılan öğrenme stratejileri; gözetimli, gözetimsiz ve pekiştirmeli öğrenme olarak görülmektedir.

Şematik yapısı Şekil 5.5 (a)’da gösterilen gözetimli öğrenme, etiketlenmiş veri kümesi üzerinden çıkarım yapan bir öğrenme stratejisidir. Gözetimli öğrenme stratejisiyle işleyen bir sinir hücresinde, çıktı değerinin hedeften ne kadar saptığı göz önünde bulundurulurken ağırlıklar güncellenir. Böylece girdi-çıkı ilişkisi, sinir ağı tarafından öğrenilir.



Şekil 5.5. ANN’de öğrenme stratejileri: a) gözetimli öğrenme, b) gözetimsiz öğrenme (Zurada’ dan, 1992)

Şematik yapısı Şekil 5.5 (b)’de gösterilen gözetimsiz öğrenme stratejisi, çıktıların herhangi bir hedef değeriyle karşılaştırılmadığı durumlarda kullanılmaktadır. Bu öğrenme

stratejisinde amaç, veri kümesi içinde var olan örüntülerin açığa çıkarılmasıdır. Bu nedenle daha çok kümeleme problemlerinin çözümünde kullanılmaktadır.

Pekiştirmeli öğrenme stratejisinde, gözetimli öğrenmede olduğu gibi bir hedef değer bulunmamaktadır. Bu stratejide öğrenme işlemi, çıktı değerinin ödüllendirilmesi veya cezalandırılması yöntemiyle gerçekleşir.

5.4. Yapay Sinir Ağlarında Öğrenme Kuralları

Yapay sinir hücrelerinde öğrenme işlemi, temel iki girdisi olan ağırlıklandırılmış girdi ve eşik değer üzerinden gerçekleşmektedir. Ağırlık ve eşik değer uzaylarının araştırılmasıyla en iyi sonuca ulaşmayı amaçlayan çeşitli öğrenme kuralları önerilmiştir.

5.4.1. Hebb öğrenme kuralı

Hebb tarafından 1949 yılında geliştirilen bu gözetimsiz öğrenme kuralı, yapay sinir ağlarında öğrenme süreci adına temel görevi görmüştür. İkili (binary) çıktı değeri üreten sinir ağları düşünülerek tasarlanan Hebb kuralı oldukça basit bir ilkeye dayanmaktadır. Bu kurala göre, aralarında bilgi akışı olan iki nöron aynı anda aktif ya da pasif ise aralarındaki bağ (ağırlıklar) kuvvetlendirilmelidir Aksi durumda ise ağırlık, çıktı değerine göre ya değişmeyecek ya da azalacaktır (Fausett, 1994). Denklem (5.3)'te t . iterasyonda ağırlıkların nasıl güncelleneceği ifade edilmiştir.

$$w_{i(t+1)} = w_{i(t)} + y_i x_i \quad (5.3)$$

Denklem incelenecek olursa; nöronları ikili çıktı değeri $\{(0), (1)\}$ üreten bir sinir ağında, çıktı ve girdinin farklı olması durumunda ağırlıkta bir değişim meydana gelmeyecektir. Nöronları iki kutuplu çıktı değeri $\{(-1), (1)\}$ üreten bir sinir ağında ise çıktı ve girdinin farklı olması durumunda ağırlıkta azalma gerçekleşecektir.

5.4.2. Perceptron öğrenme kuralı

Rosenblatt'ın 1958 yılında ikili sınıflandırma problemleri için geliştirdiği perceptron (algılayıcı) öğrenme kuralı, gözetimli öğrenme kuralları arasındadır. Perceptron öğrenme kuralı, çıktının (y_i), hedef değer (d_i), üzerinden incelenmesiyle tanımlanmaktadır. Eğer çıktı ve hedef değer arasında uyumsuzluk var ise ağırlıkların ve eşik değerinin güncellenmesi gerekmektedir (Alavala, 2012). Bu güncelleme işleminde küçük değerli ve pozitif bir sayı olan öğrenme katsayısı da ($0 \leq \eta \leq 1$) kullanılmaktadır. Öğrenme katsayısı, hedef değere ulaşma aşamasında yapılan güncellemenin adım büyüklüğünü belirler (Mitchell, 1997).

$$w_{i(t+1)} = w_{i(t)} + \eta(d_i - y_i)x_i \quad (5.4)$$

$$b_{i(t+1)} = b_{i(t)} + \eta(d_i - y_i) \quad (5.5)$$

5.4.3. Delta öğrenme kuralı

Widrow ve Hoff tarafından geliştirilen Adaline (adaptive linear element) modeli tek katmandan oluşan bir yapıdır. Birden fazla Adaline ünitesinin bir araya gelerek oluşturduğu iki katmanlı yapı ise Madaline (Multiple Adaline) olarak isimlendirilir (Fausett, 1994). Adaline/Madaline modelinin perceptronlardan farkını, kullanılan öğrenme kuralı belirler. Geliştirilen delta (Widrow ve Hoff) öğrenme kuralında, en küçük ortalama kareler hesaplaması kullanılır. Bu öğrenme kuralındaki amaç çıktı ile hedef değer arasındaki karesel hataların azaltılmasıdır. Delta öğrenme kuralında hata değeri aşağıdaki eşitlikle hesaplanmaktadır (Alavala, 2012):

$$E = \frac{1}{2} \sum_i (d_i - y_i)^2 \quad (5.6)$$

Delta öğrenme kuralında, ağırlıklardaki değişim, hatanın ağırlığa göre olan değişimiyle orantılıdır.

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} \quad (5.7)$$

Hata fonksiyonunu enküçükleyecek olan w_i değerleri, gradyan iniş (gradient descent) metoduyla bulunur (Zurada, 1992).

$$\begin{aligned}
 \frac{\partial E}{\partial w_i} &= \frac{\partial \left[\frac{1}{2} (d_i - y_i)^2 \right]}{\partial w_i} \\
 &= \frac{\partial \left[\frac{1}{2} (d_i - y_i)^2 \right]}{\partial y_i} \cdot \frac{\partial y_i}{\partial w_i} \\
 &= -(d_i - y_i) \frac{\partial y_i}{\partial w_i} \\
 &= -(d_i - y_i) \frac{\partial y_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_i} \\
 &= -(d_i - y_i) x_i
 \end{aligned} \tag{5.8}$$

Böylece ağırlıktaki değişim Denklem (5.9)'daki gibi ifade edilir.

$$\Delta w_i = w_{i(t+1)} - w_{i(t)} = \eta (d_i - y_i) x_i \tag{5.9}$$

Hedef değer ve gerçekleşen çıktı arasındaki fark δ ile ifade edilecek olursa, ağırlıkların ve eşik değerinin güncellenmesi aşağıdaki denklemler ile gerçekleşir.

$$w_{i(t+1)} = w_{i(t)} + \eta \delta_i x_i \tag{5.10}$$

$$b_{i(t+1)} = b_{i(t)} + \eta \delta_i \tag{5.11}$$

5.4.4. Kohonen öğrenme kuralı

Kohonen (1982) tarafından geliştirilen bu kural, gözetimsiz öğrenme yöntemlerinden biridir. Bütün sinir hücrelerinin öğrenmek adına rekabet ettikleri bu öğrenme kuralında, girdi değerine en yakın ağırlıklara sahip olan nöronlar kazanır. Bu öğrenme kuralında kazanan hücrelerin çıktı değeri 1, diğer hücrelerin ise 0 olacaktır. Bu nedenle “kazanan hepsini alır” kuralı olarak da isimlendirilir (Negnevitsky, 2005). Kural Denklem (5.12)'de verilmiştir.

$$w_{i(t+1)} = w_{i(t)} + \eta (x_i - w_{i(t)}) \tag{5.12}$$

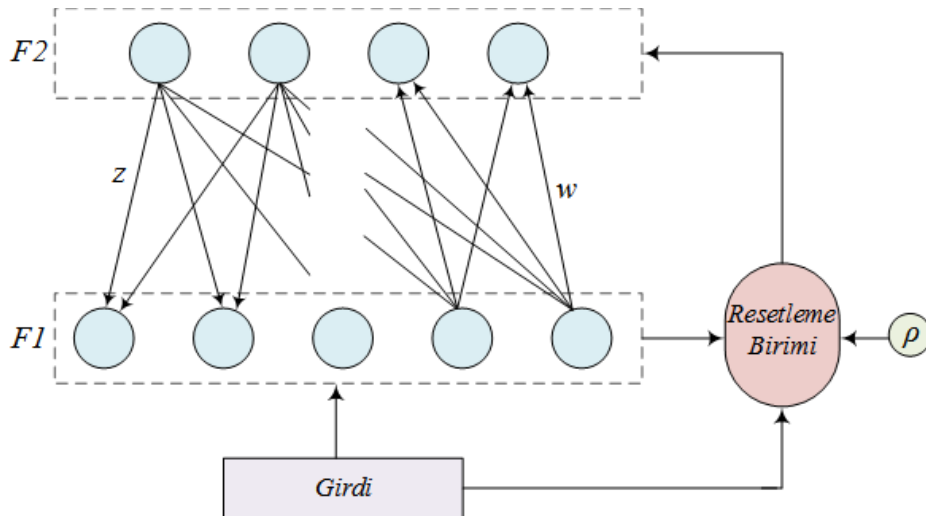
5.5. Yapay Sinir Ağı Modelleri

Katman sayısı, sinir hücreleri ve katmanlar arasındaki besleme yönü, öğrenme yöntemleri gibi tercihlerin farklı şekillerde, farklı ya da benzer problem tiplerine yönelik kullanılmasıyla çeşitli yapay sinir ağı modelleri geliştirilmiştir.

5.5.1. Adaptif rezonans teorisi

Adaptif rezonans teorisi (ART), Carpenter ve Grossberg tarafından 1986 yılında geliştirilmiştir. Gözetimsiz öğrenme stratejisini benimsemiş olan ART, gerçek zamanlı öğrenen bir sinir ağı yapısıdır. Yeni durumlar karşısında adaptasyon yeteneği yüksek olan ART, bu gücünü öğrenme ve unutma işlemlerini bir arada yürütebilmesinden alır (Öztemel, 2012).

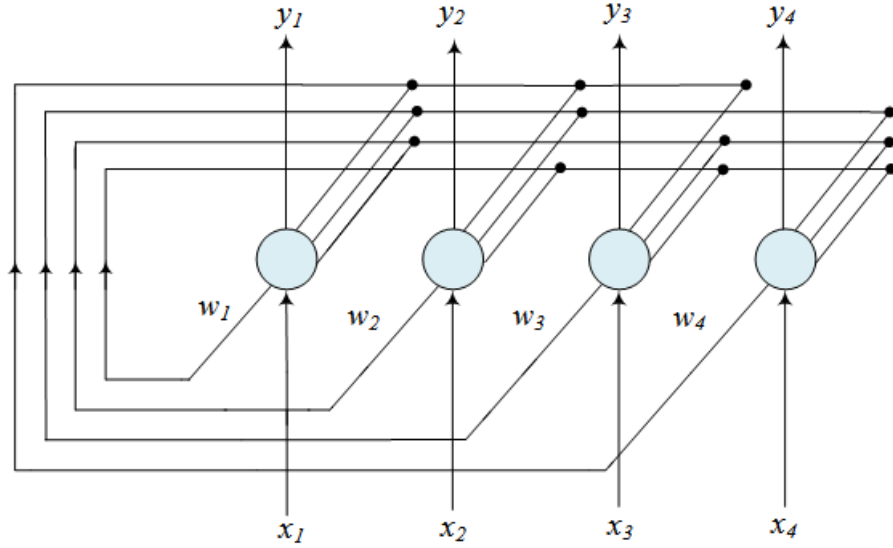
Şekil 5.6'da görüldüğü gibi ART, $F1$ ve $F2$ olmak üzere birbirine iki yönlü bağlı olan iki katmandan oluşmaktadır (Chakraborty, 2010). Dolayısıyla bu ağda nöronlar ileri ve geri beslemeli bir yapı oluştururlar. Burada $F1$ girdinin özelliklerini, $F2$ ise çıktı katmanını temsil etmektedir. $F2$ katmanında rekabetçi bir öğrenme kuralıyla çıktı oluşturulur. Oluşturulan bu çıktının, giriş örüntüsüne olan benzerliği araştırılır. Bu benzerlik katsayısı (ρ) denilen bir parametreyle sınırlanır. Karar ise $F1$ katmanından sinyal alan resetleme ünitesi tarafından verilir (Carpenter ve Grossberg, 1987).



Şekil 5.6. ART ağlarının genel yapısı

5.5.2. Hopfield ağı

Hopfield (1982) tarafından tek katmanlı olarak tasarlanan Hopfield ağı, tekrar beslemeli (recurrent) ve her bir sinir hücresinin birbirine bağlı olduğu simetrik bir yapıya sahiptir (Hajek, 2005). Hopfield ağının yapısı Şekil 5.7’de gösterilmiştir. Hopfield ağları, özelleştirilmiş Hebb öğrenme kuralı ile eğitilirler. Bu yöntemin klasik Hebb öğrenme kuralından tek farkı, öğrenme katsayısının da kullanılmasıdır. Hopfield ağının eğitim aşamasında sisteme verilen desenler hafızada saklanır. Test aşamasında ise gürültülü ya da kısmi ipuçları verilmiş olan desenler, hafızada saklanan desenler ile eşleştirilmeye çalışılır (Sezgin vd.,2012).



Şekil 5.7. Hopfield ağının yapısı

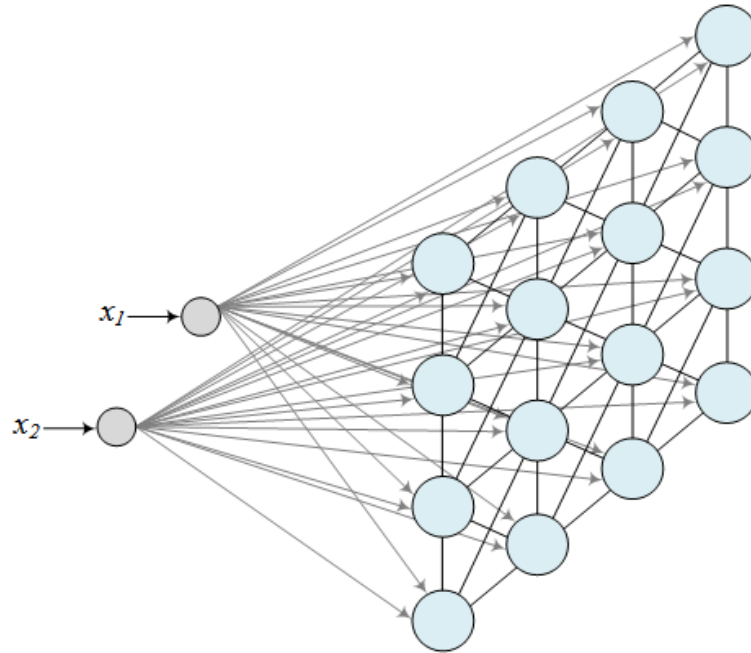
Kesikli Hopfield ağlarında işaret aktivasyon fonksiyonuna sahip McCulloch-Pitts sinir hücresi modelini kullanılır (Negnevitsky, 2005). Sürekli Hopfield ağları ise kesikli ağlarla aynı yapıdadır. Ancak sigmoid fonksiyonu gibi sürekli olan bir aktivasyon fonksiyonuna sahiptir (Rojas, 1996).

5.5.3. Kohonen ağı (SOM)

Kohonen’in (1984) ileri sürdüğü SOM (self-organizing maps), beynin çalışma şekillerinden birini tanımlayan ilişkisel hafıza (associative memory) kavramı üzerine

kurulmuştur (Demartines ve Blayo, 1992). İlişkisel hafıza, duyu organları aracılığıyla beyine iletilen uyarıların, hafızada bulunan örüntülerle ilişkilendirildiği hafıza tipidir (Khan ve Makwana, 2005).

Yapısı Şekil 5.8’de verilen SOM, bir girdi ve iki boyutlu çıktı katmanından oluşmaktadır. Girdi katmanının her bir çıktısı, çıktı katmanındaki tüm sinir hücrelerine iletilir. Çıktı katmanındaki sinir hücreleri rekabetçi öğrenme kuralıyla eğitilirler. Kazanan sinir hücresi ile yakınında bulunan diğer sinir hücreleri arasında komşuluk ilişkisi kurulur. Ağırlık vektörünün güncellenmesi, kazanan sinir hücresi ve komşuları için geçerlidir. Güncelleme sonrası; kazanan sinir hücresi en çok, komşular ise kazanan nörona olan uzaklıklarıyla ters orantılı biçimde girdi vektörüne yakınsarlar (Gutierrez-Osuna, 2002). Öğrenme süreci ilerledikçe komşu sinir hücresi sayısı azalmaktadır ve kazanan nöronlar etrafında gruplaşma meydana gelmektedir. Bu nedenle, Kohonen ağı kümeleme problemlerinin çözümüne oldukça uygun bir yapıya sahiptir (Beale ve Jackson, 1990).



Şekil 5.8. Kohonen ağı (SOM) yapısı

5.5.4. Savaşlı yayılım ağı

Savaşlı yayılım ağı Hecht-Nielsen (1988) tarafından geliştirilmiş olup çok katmanlı bir yapıya sahiptir. Kohonen ağı ve Grossberg outstar öğrenme kuralının kombinasyonu

oluşturulmuştur (Sivanandam vd.,2006). Girdi vektörleri Kohonen katmanına iletilir ve burada kümelenirler. Kümelenmiş veri Grossberg katmanında outstar öğrenme kuralıyla eğitilir ve daha sonra sinir ağının çıktılarını üretir (Taylor, 2006).

5.5.5. Çok katmanlı algılayıcı

Minsky ve Rupert'in (1969) "Perceptrons" adlı kitabında, perceptronların sadece doğrusal ayrılabilen problemleri çözebildiği, fakat XOR gibi doğrusal olamayan problemlere bir çözüm üretmediği gösterilmiştir. Yapay sinir ağları ile ilgili çalışmaları durdurma noktasına getiren bu soruna; Werbos'un (1974) temellerini attığı, Rumelhaart vd.nin (1986) geliştirdiği geriyayılım algoritmasıyla (genelleştirilmiş delta öğrenme kuralı) çözüm bulunmuştur (Smith, 1997). Bu çözümün arkasındaki temel fikir; ileri beslemeli ve çok katmanlı bir sinir ağında hatanın çıktı katmanından itibaren geriye doğru yayılmasıyla ağırlıkların düzenlenmesidir. Bu nedenle bu tip ağlar, geriyayılım ağları olarak adlandırılmıştır (Alavala, 2012). Geriyayılımlı ağlarda öğrenme süreci dört ana aşamada gerçekleşir (Sivanadam vd., 2006):

1. Başlangıç ağırlıklarının atanması
2. Girdilerin ileri yönde akışı
3. Hatanın geriyayılımı
4. Ağırlık ve eşik değerlerinin yeniden düzenlenmesi

5.5.5.1. Genelleştirilmiş delta kuralının türetilmesi

İki kutuplu çıktı değerleri üreten tek bir sinir hücresi için delta kuralıyla hatanın ağırlıklara hangi ölçüde etki edeceği Denklem (5.10)'da gösterilmiştir. Çoğunlukla sigmoid fonksiyon olmak üzere, doğrusal olmayan ve türevlenebilir aktivasyon fonksiyonu kullanılan çok katmanlı algılayıcılar (MLP) için delta kuralının genelleştirilmesi gerekmektedir (Kriesel, 2005). Çıktı katmanından girdi katmanına doğru hatanın geriyayılımını sırasıyla incelenecek olursa;

Şekil 5.9'daki gibi bir yapıya sahip sinir ağının gizli katmanında bulunan j . nöronun, w_{jk} ağırlığı üzerinden girdi alan çıktı katmanının k . nöronunun hata fonksiyonu aşağıdaki gibi yazılır.

$$E = \frac{1}{2} \sum_k (d_k - y_k)^2 \quad (5.13)$$

Hata fonksiyonunu enküçükleyecek olan w_{jk} değerleri gradyan iniş metoduyla araştırılır:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} \quad (5.14)$$

Hatanın ağırlıklara göre değişimi, zincir kuralıyla genişletilecek olursa

$$\begin{aligned} \frac{\partial E}{\partial w_{jk}} &= \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_{jk}} \\ &= \frac{\partial \left[\frac{1}{2} (d_k - y_k)^2 \right]}{\partial y_k} \cdot \frac{\partial y_k}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_{jk}} \\ &= -(d_k - y_k) \frac{\partial y_k}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_{jk}} \\ &= -(d_k - y_k) f'(z_k) x_k \end{aligned} \quad (5.15)$$

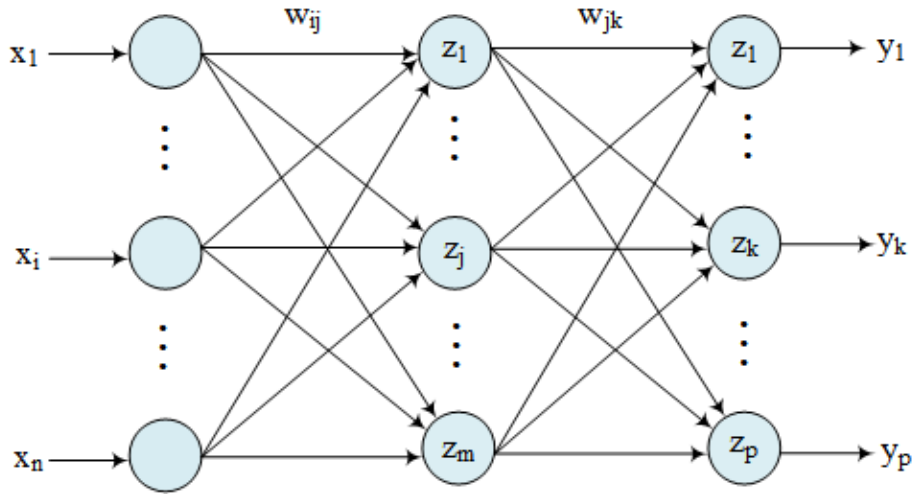
ifadesi elde edilir.

$$\delta_k = -(d_k - y_k) f'(z_k) \quad (5.16)$$

olmak üzere çıktı katmanına bağlanan ağırlıklar Denklem (5.17) ile düzenlenir. Bu denklemdeki η öğrenme katsayısını gösterip ağırlıkların değişim büyüklüğünü belirlemektedir.

$$w_{jk(t+1)} = w_{jk(t)} + \eta \delta_k x_k \quad (5.17)$$

Denklem (5.17)'de tanımlanan işlem, ağırlık değerlerinin güncellenmesi için çıkış katmanına ulaşan bütün bağlantılara uygulanır. Ancak hedef değere ulaşabilmek adına gizli katmana taşınan hatalı ağırlıkların da düzenlenmesi gerekmektedir (Freeman ve Skapura, 1991). Bu bağlamda gizli katman için de hatayı enküçülemek gerekmektedir. Fakat gizli katmanın çıktısını karşılaştırabilecek bir hedef değer bulunmamaktadır. Bu nedenle gizli katmanın ağırlıklarını düzenleyebilmek adına, toplam hata ve gizli katmanla arasında matematiksel bir ilişki kurulmalıdır.



Şekil 5.9. Geriyayılım ağının yapısı

Gizli katmanın j . nöronuna iletilen girdiler, girdi katmanının i . nöronundan w_{ij} ağırlığı ile taşınmaktadır. Hatanın w_{ij} ağırlığına göre değişimi, zincir kuralıyla genişletilerek Denklem (5.18) ile ifade edilebilir.

$$\begin{aligned} \frac{\partial E}{\partial w_{ij}} &= \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial z_k} \cdot \frac{\partial z_k}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}} \\ &= -(d_k - y_k) f'(z_k) w_{jk} f'(z_j) x_j \\ &= \delta_k w_{jk} f'(z_j) x_j \end{aligned} \quad (5.18)$$

Denklem (5.18)'de görüldüğü gibi gizli katmanın hata değeri, çıktı katmanında hesaplanan hatanın (δ_k), ağırlıklar (w_{jk}), üzerinden geriye dağıtılmasıyla hesaplanmaktadır. Buradan yola çıkarak j . gizli katmana iletilen ağırlıkların güncellenmesi Denklem (5.19)'da tanımlı işlemle gerçekleşir.

$$w_{ij(t+1)} = w_{ij(t)} + \eta f'(z_j) x_j \sum_k \delta_k w_{jk} \quad (5.19)$$

Çıktı katmanı ve gizli katmanlar için ağırlıkların yeniden düzenlenmesi sonrasında, ağıya yeni girdiler alınır ve süreç baştan tekrar işler. İstenilen hata değerine ulaşıldığında ise yineleme son bulur. Çok katmanlı bir sinir ağı için geriyayılım algoritmasının adımları Ek Açıklama-A'da verilmiştir (Freeman ve Skapura, 1991; Fausett, 1994).

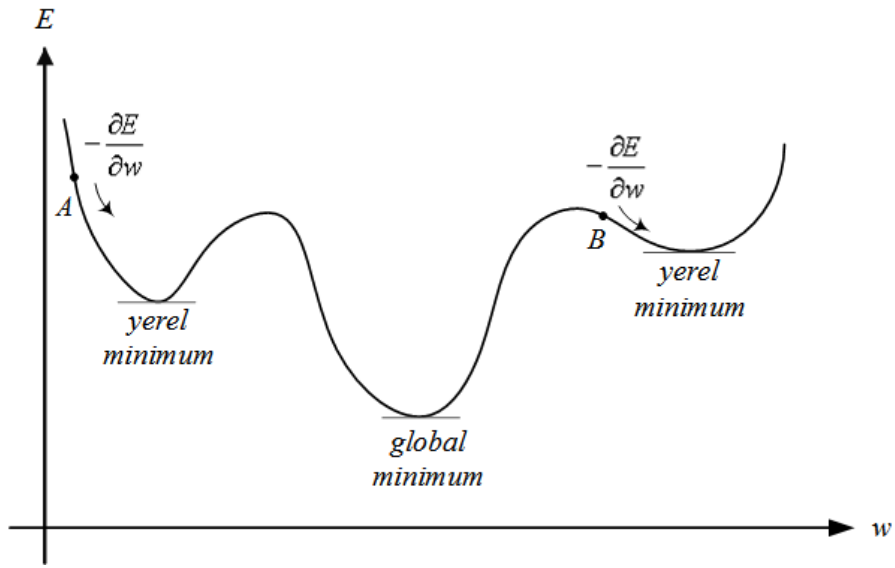
5.5.5.2. Öğrenme katsayısı

Geri yayılım algoritmasında öğrenme katsayısı, ağırlıkların güncellenme büyüklüğünü etkileyen ve $(0,1]$ aralığında değer alan bir parametredir. Öğrenme katsayısının küçük bir değer alması, ağırlıkların güncellenmesi işleminin sık tekrar edilmesine neden olarak modelin eğitim süresini uzatacaktır. Öğrenme katsayısının büyük seçilmesi ise çözümde büyük sıçramalara neden olacağından bütünsel eniyiyi kaçırma riskini ortaya çıkarmaktadır (Haykin, 2009).

5.5.5.3. Momentum terimi

Şekil 5.10'daki gibi birden fazla ekstremum noktası bulunan hata-ağırlık fonksiyonlarında, gradyan iniş A ya da B gibi bir noktadan başladığında çözüm yerel minimuma takılı kalacaktır (Gurney, 1997). Gradyan iniş yöntemi, hatayı azaltma eğiliminde olduğundan yerel minimuma eriştikten sonra hata fonksiyonunun arttığı kısmı tırmanamayacak, dolayısıyla bütünsel minimuma ulaşamayacaktır. Böylece algoritma, tüm eğitim verisine karşılık kısmi bir çözüm üretecektir. Bu sorunun üstesinden gelebilmek adına momentum (μ) adı verilen bir terim geriyayılım algoritmasına eklenmiştir. $0 \leq \mu \leq 1$ aralığında değer alan momentum teriminin eklenmesiyle ağırlıklar Denklem (5.20)'deki eşitlikle güncellenir.

$$\begin{aligned} w_{jk(t+1)} &= w_{jk(t)} + \eta \delta_k x_k + \mu (w_{jk(t)} - w_{jk(t-1)}) \\ w_{jk(t+1)} &= w_{jk(t)} + \eta \delta_k x_k + \mu \Delta w_{jk(t)} \end{aligned} \quad (5.20)$$



Şekil 5.10. Hata fonksiyonunda yerel ve bütünsel minimum

Momentumlu geri yayılım algoritmasında, bir önceki iterasyondaki değişimin mevcut iterasyondaki değişime eklenmesi sağlanır. Böylece yerel minimuma takılan ağırlara, sıçrama ile bütünsel minimuma erişme potansiyeli ve salınımlıların önüne geçilerek sonuca daha hızlı ulaşma yeteneği kazandırılır (Öztemel, 2012). Sonuç olarak Rumelhart vd.nin (1986) belirttiği üzere, momentum katsayısı çözümü hızlandırırken yerel minimumdan kaçınmaya da yardımcı olur.

6. DESTEK VEKTÖR MAKİNELERİ

Destek vektör makineleri (SVM), sınıflandırma ve regresyon analizi için kullanılan gözetimli makine öğrenmesi algoritmasıdır. Vapnik ve Chervonenkis tarafından öne sürülen istatistiksel öğrenme teorisinin üzerine inşa edilen SVM'in teorik temelleri, 1960'lı yıllardaki çalışmalara kadar gitmektedir. Vapnik ve Lerner (1963) ile Vapnik ve Chervonenkis'nin (1964) örüntü tanıma amacıyla geliştirdikleri genelleştirilmiş betimleme metodu (generalized portrait method) SVM'nin temelini oluşturmaktadır. SVM bu algoritmanın doğrusal olmayanına genelleştirilmiş halidir (Smola ve Schölkopf, 2003).

Vapnik ve Chervonenkis'in (1974) istatistiksel öğrenme teorisi üzerine yaptıkları çalışmanın öncüllüğünde Vapnik (1979) SVM'yi geliştirmeye başlamıştır. SVM ilk olarak Boser, Guyon ve Vapnik (1992) tarafından COLT konferansında tanıtılmıştır. Vapnik (1995) ile Cortes ve Vapnik'in (1995) çalışmalarıyla SVM günümüzdeki formuna ulaşmıştır.

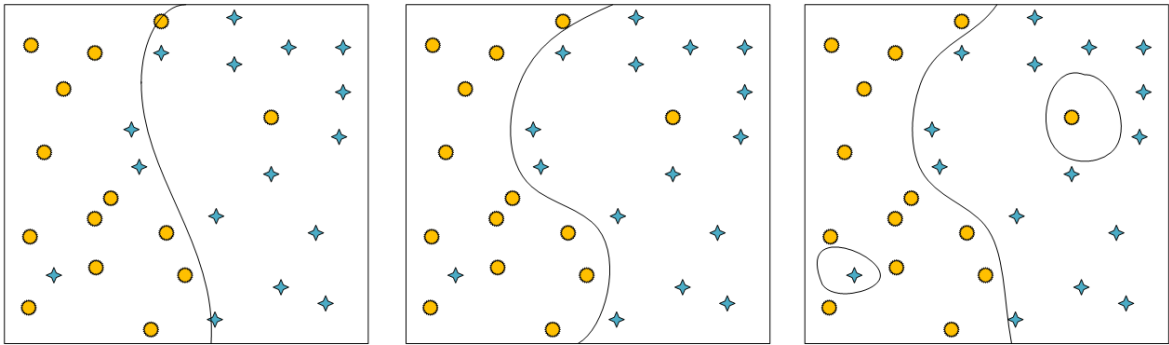
SVM'nin temel fikri, farklı örnek tipleri arasındaki en uygun ayrımı yapacak hiper düzlemin belirlenmesidir (Vapnik, 2000). SVM eğitim veri kümesi üzerinde ikili sınıflandırma yaparken, farklı sınıfların birbirlerine en yakın örneklerini bularak bu örnekler üzerinden destek vektörleri oluşturur. Sınıflandırma sonucu değişmeden birçok vektör çizilebilmektedir. Burada asıl amaç destek vektörlerin ayırıcı hiper düzleme olan uzaklığını enbüyüklemektir.

6.1. İstatistiksel Öğrenme Teorisi (Vapnik-Chervonenkis Teorisi)

İlk çalışmaları 1960'ların sonlarında yapılan istatistiksel öğrenme teorisi (VC teorisi), 1990'lara kadar veriler yardımıyla gerçekleşen makine öğrenmesinin teorik bir analizi olarak kullanılırken; 1990'ların ortalarında bu teoriyi temel alan yeni bir tür öğrenme algoritması (destek vektör makinesi) geliştirilmiştir. SVM'nin geliştirilmesiyle istatistiksel öğrenme teorisi, teorik analizlerde kullanılmamanın dışında gerçek hayat problemlerine de pratik çözümler üretebilecek bir araç halini almıştır (Vapnik, 1999).

Klasik istatistik, verinin bir dağılıma uyduğunu varsayarak dağılımın parametrelerini belirlemeye çalışırken, VC teorisi dağılımdan bağımsız bir şekilde en uygun modeli bulmayı amaçlamaktadır (Cherkassky ve Mulier, 2007).

Şekil 6.1’de görülen üç farklı karar sınırı ile çözülmüş ikili sınıflandırma problemi incelenecek olursa (Schölkopf ve Smola, 2002); en soldaki şekilde yer alan sınıflandırmada neredeyse doğrusal olan karar sınırı, probleme eksik uyum göstererek karar sınırına çok yakın değerleri kaçırmıştır. En sağdaki şekilde sınıflama %100 doğru yapılmasına karşın karar sınırı eğitim seti için fazla özelleştirilmiş yani aşırı uyum göstermiştir. Böyle bir modelde test verisi sınanırken aykırı noktaların yakınında bulunması daha olası olan diğer sınıfa ait verilerin yanlış sınıflandırılması kaçınılmaz olacaktır. Ortadaki şekil ise iki durumun arasını bularak az sayıda sınıflandırma hatası yaparken sınıflar arasında genelleştirilmiş bir karar sınırı oluşturmuştur. VC teorisinin amacı da hatayı sıfırlamak değil en uygun modelle enküçükmektir.



Şekil 6.1. İkili sınıflandırma probleminin üç farklı modelle çözümü

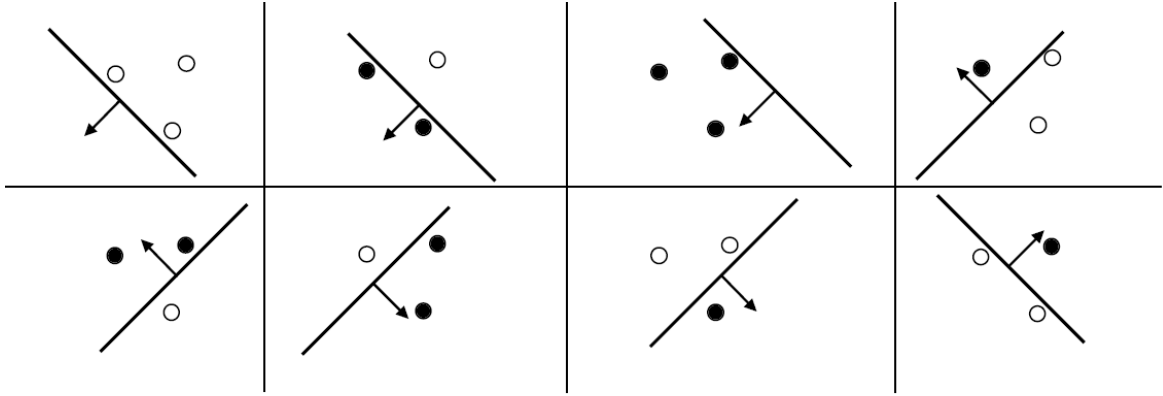
VC teorisi sadece eğitim verisine uyacak bir model inşa etmekten çok, sistem tarafından önceden görülmemiş veriler için de uygun olacak, makul düzeyde genelleştirilmiş bir model oluşturmayı hedefler. Bu durumda genelleştirmenin VC teorisinin özü olduğu söylenebilir (Bar-Joseph vd. 2001). Bu noktada sorulması gereken soru, bir modelin çözüm için uygun olduğuna nasıl karar verildiğidir. VC teorisi, öngörü modelleri kurarken karşılaşılan bu önemli soruyu, VC boyutu ve yapısal risk minimizasyonu prensibini göz önünde bulundurarak cevaplamaktadır.

6.1.1. VC boyutu

Literatürdeki birçok teorik ve deneysel çalışma, öğrenme makinelerinin kapasitesinin genelleme kabiliyeti üzerindeki etkisini ortaya koymuştur. Küçük kapasiteli öğrenme makineleri mümkün olan en iyi çözüme (test verisinde en küçük hata oranı) ulaşabilmek için büyük boyutlu eğitim verisine ihtiyaç duymamaktadır. Yüksek kapasiteli öğrenim makineleri ise daha iyi asimptotik çözümler sağlayabilirken (büyük boyutlu eğitim verisi için daha küçük hata oranı) makul bir test performansı için çok büyük veri kümelerine gereksinim duymaktadır. Böylece kullanılan eğitim verisi için, eğitim hatası ve test hatası arasındaki farklılık yüksek kapasiteli öğrenim makineleri için daha büyük olacaktır. VC boyutunu temel alan istatistiksel öğrenme teorisi, eğitim hatası ile test hatası arasındaki farkın davranışını, eğitim verisi büyüklüğünün bir fonksiyonu olarak tek bir nicelik (VC boyutu) üzerinden karakterize edilebileceğini öngörmektedir (Vapnik vd., 1994). VC boyutu öğrenim makinesinin kapasitesini ölçen skaler bir değerdir (Gunn,1998). Kapasite ise, hipotez uzayının karmaşıklığının bir ölçüsüdür (Sewell, 2008).

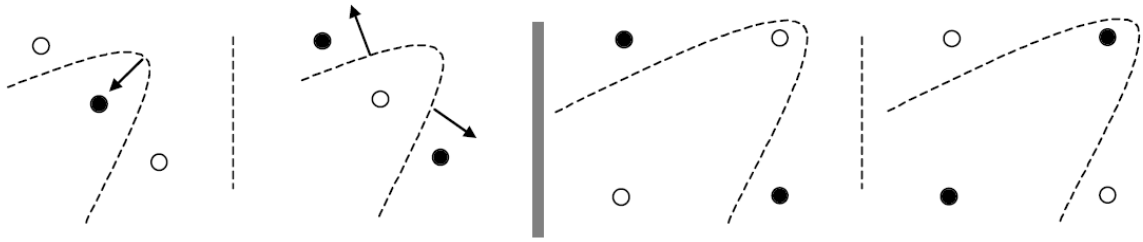
$T = \{(x_i, y_i): i = 1, \dots, n \in \mathbb{Z}\}$ şeklinde tanımlanan bir eğitim veri setindeki tüm vektörler, $y \in \{-1, +1\}$ biçiminde etiketlenilecek olursa, veriler 2^n farklı şekilde parçalanabilecektir. Söz konusu öğrenme problemi için, etiketlemeyi yapabilecek $f \in \mathcal{H}$ şeklinde bir fonksiyon bulunabilir. Böylece, n hacimli veri seti, fonksiyonlar kümesi \mathcal{H} tarafından parçalanabilir. Fonksiyonlar kümesi \mathcal{H} tarafından parçalanabilen maksimum nokta sayısı VC boyutu olarak adlandırılır (Alpaydın, 2010).

Örnek olarak \mathbb{R}^2 'deki 3 nokta $2^3 = 8$ biçimde parçalanabilir. Bu örneğin görsel ifadesi Şekil 6.2'de verilmiştir. Şekilde ok ile gösterilen bölge $y = +1$ yani pozitif sınıfa ait örnekleri, diğer bölge ise $y = -1$ yani negatif sınıfa ait örnekleri tanımlamaktadır.



Şekil 6.2. \mathbb{R}^2 'de 3 noktanın ayrımı (Kecman'dan, 2001)

Ancak dördüncü bir noktanın varlığı durumunda Şekil 6.3'te görüldüğü gibi örnekleri parçalamanın mümkün olmadığı durumlar olacaktır. Dolayısıyla \mathbb{R}^2 'de olası bütün etiketlemenin parçalanabilirliği en fazla 3 nokta için geçerli olduğundan VC boyutu 3'tür.



Şekil 6.3. \mathbb{R}^2 'de 4 noktanın ayrımı (Kecman'dan, 2001)

Genelleştirecek olursak \mathbb{R}^m uzayında doğrusal parçalama için VC boyutu $m + 1$ dir. Doğrusal olarak ayrılabilen kümelerle öğrenme makinelerinin VC boyutu kolaylıkla hesaplanabilirken doğrusal olmayan durumlarda VC boyutunun hesaplanması oldukça zordur (Kecman, 2001). Doğrusal olmayan sınıflandırıcıların VC boyutunu hesaplamak için Vapnik vd. (1994) bazı ampirik yöntemler önermiştir.

6.1.2. Yapısal risk minimizasyonu

Vapnik ve Chervonenkis (1974)'in çalışmasında söz edilen yapısal risk minimizasyonu (structural risk minimization) sonlu veri kümelerinden öğrenme uygulamaları için en uygun modelin seçiminde faydalı olacak tümevarımsal bir prensiptir (Kecman, 2001).

Eđitim hatasının azalması, test hatasının da her zaman azalacađını garanti etmezken, öğrenme makinesinin kapasitesini arttırmak eğitim verisine olan uyumda artışa (aşırı uyum), test hatasında ise performans düşüklüğüne neden olabilir (Hamel, 2009). Bu prensip, modelin karmaşıklığı ve eğitim hatasının uyum niteliđi (ampirik hata) arasında bir ödünleşme sağlar (Sewell, 2008). Böylece kurulan denge ile beklenen riske bir üst sınır oluşturularak optimum düzeyde genelleştirme yapılması amaçlanır.

$$R[f] \leq R_{emp}(f) + \vartheta(h, N, \eta) \quad (6.1)$$

VC güvenilirliđi (VC-confidence) olarak adlandırılan $\vartheta(h, N, \eta)$ terimi;

$$\vartheta(h, N, \eta) = \sqrt{\frac{h \left[\ln \left(\frac{2N}{h} \right) + 1 \right] - \ln \left(\frac{\eta}{4} \right)}{N}} \quad (6.2)$$

şeklinde ifade edilir. Böylece beklenen riskin üst sınırı;

$$R[f] \leq R_{emp}(f) + \sqrt{\frac{h \left[\ln \left(\frac{2N}{h} \right) + 1 \right] - \ln \left(\frac{\eta}{4} \right)}{N}} \quad (6.3)$$

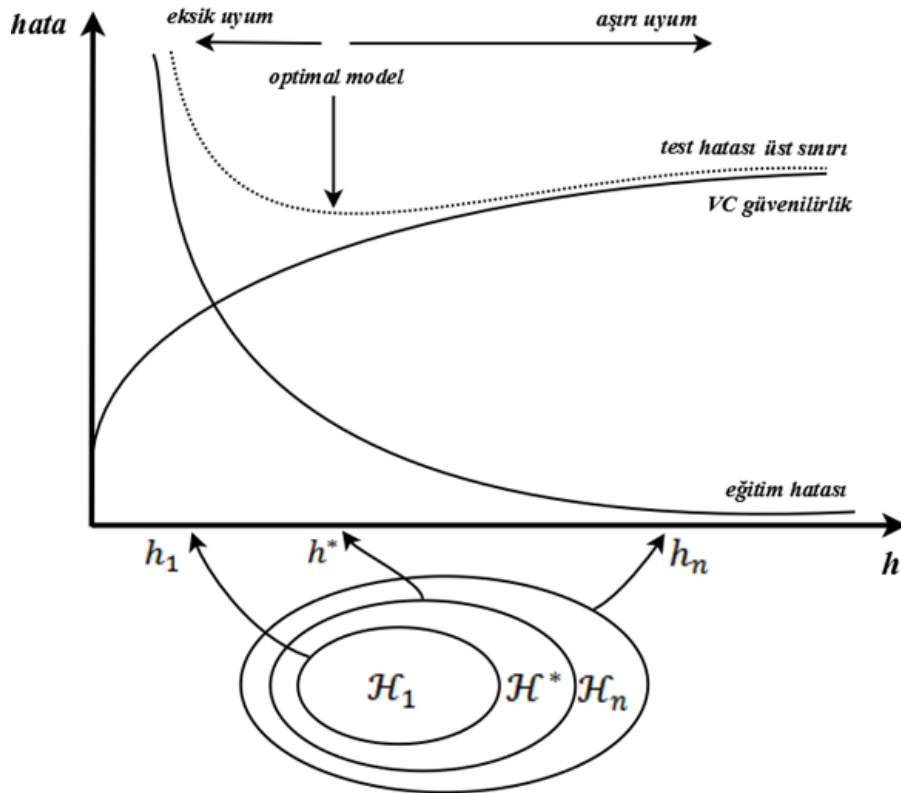
olarak yazılabilir. Burada $R[f]$ beklenen riski, $R_{emp}(f)$ deneysel hatayı (eđitim hatası), $\vartheta(h, N, \eta)$ VC güvenilirlik terimini, h VC boyutunu, N eğitim kümesinin büyüklüğünü, η ise $(0,1)$ aralıđındaki bir deđeri ifade etmektedir. Denklem (6.4)'te verilen sınır $1 - \eta$ olasılıkla oluşturulmuştur.

$$P \left(R[f] \leq R_{emp}(f) + \sqrt{\frac{h \left[\ln \left(\frac{2N}{h} \right) + 1 \right] - \ln \left(\frac{\eta}{4} \right)}{N}} \right) = 1 - \eta \quad (6.4)$$

Yapısal risk minimizasyonunda, artan VC boyutuna (karmaşıklık) karşılık gelen hiyerarşik bir hipotez uzayı söz konusudur. Bu hipotez uzayına ait fonksiyonlar sınıfı da iç içe geçmiş alt kümeler şeklinde bir yapı oluştururlar. Bir başka ifadeyle, fonksiyonlar sınıfı için, $\mathcal{H}_1 \subset \mathcal{H}_2 \subset \dots \subset \mathcal{H}_n$ şeklinde iç içe geçmiş alt kümeler tanımlanırken bu alt kümelere

ait VC boyutları, $h_1 \leq h_2 \leq \dots \leq h_n$ şeklinde sıralanmaktadır (Cherkassky ve Mulier, 2007).

Şekil 6.4'te gösterildiği gibi VC teorisi, öğrenme makinelerinin genelleme yeteneğini Denklem (6.3)'teki beklenen riskin üst sınırını enküçüklemeye çalışarak kontrol eder. Bu işlem hipotez uzayındaki fonksiyon kümesinin karmaşıklığını mümkün olduğu kadar basit tutarken eğitim hatasının da enküçüklenmesiyle gerçekleştirilebilir. Bu dengeyi kurmak VC boyutunu yani modelin karmaşıklığını kontrol etmekten geçer (Cherkassky ve Mulier, 2007).



Şekil 6.4. Yapısal risk minimizasyonu

6.2. Doğrusal Destek Vektör Sınıflandırıcısı

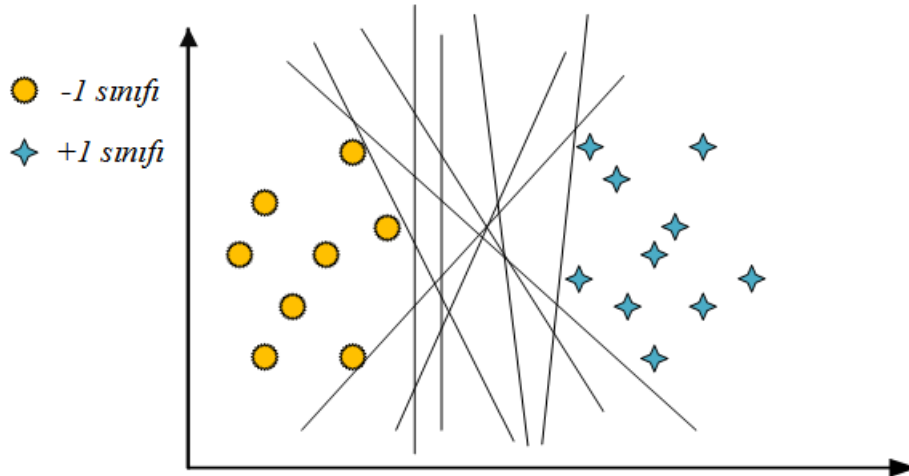
Destek vektör makinelerin temel amacı; ikili sınıflandırma problemi için kullanılacak veri kümesini, olabilecek en uygun şekilde ayıracak en büyük marjli hiper düzlemi bulmaktır. SVM'nin geniş marj sınıflandırıcısı VC teorisine dayanmaktadır. Böylece

mümkün olan en az karmaşıklığa ve yüksek genelleştirme yeteneğine sahip bir ayırıcı fonksiyon araştırılmaktadır (Zhang, 1999; Smola vd., 1999).

Genel olarak SVM’de, verilerin doğrusal olarak ayrılabilirliği ya da doğrusal olarak ayrılamadığı biçiminde iki farklı durum söz konusudur. Verilerin doğrusal olarak ayrılabilirliği durumlarda, iki sınıf birbirinden bir doğru yardımıyla hatasız ya da bir miktar hatayla ayrılabilir. Ancak verinin doğrusal olarak ayrılamadığı durumlar söz konusu olduğunda veri kümesi, bir çekirdek fonksiyonu aracılığıyla doğrusal bir hiper düzlemle ayrılabilirliği üst uzaya taşınarak sınıflandırma işlemi gerçekleştirilir.

6.2.1. Doğrusal olarak ayrılabilirlik durumu (hard margin)

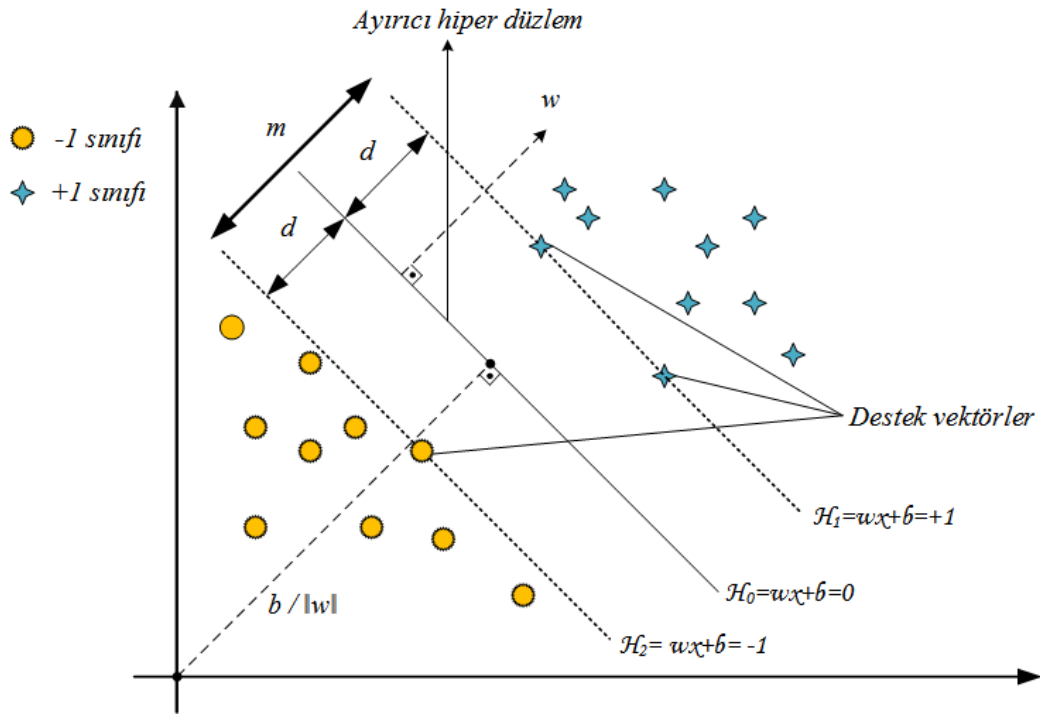
Destek vektör makinelerinin en basit hali, verilerin doğrusal olarak ayrılabilirliği sınıflandırma problemlerinin çözümüyle gösterilir. SVM’nin esasını oluşturan bu çözüm, SVM ile makine öğrenmesi sürecinin nasıl işlediğini göstermesi açısından da anahtar görevi görmektedir. Şekil 6.5’te doğrusal olarak ayrılabilir ikili sınıflandırma probleminde olası ayırıcı doğrular gösterilmiştir.



Şekil 6.5. Doğrusal olarak ayrılabilir ikili sınıflandırma probleminde ayırıcı doğrular

Şekil 6.5’te gösterilen iki sınıflandırma probleminde veriler doğrusal olarak ayrılmakta ve bu ayırma işlemini yapabilecek birçok sayıda hiper düzlem çizilebilmektedir. Bununla beraber bu işlemi yapacak bir tane en büyük marjinli hiper düzlem bulunmaktadır.

(Olson ve Delen, 2008). En uygun \mathcal{H}_0 ayırıcı hiper düzlemi belirlemek için destek vektörlerden faydalanılmaktadır. SVM için marjın (m), destek vektörler yardımıyla belirlenen bu iki doğrunun arasında kalan bölge ile tanımlanır. Şekil 6.6'da destek vektör makinesine ait bu kavramlar ve sınıflandırma yapısı gösterilmiştir.



Şekil 6.6. Destek vektör sınıflandırıcısı

Şekil 6.6'da gösterilen \mathcal{H}_0 ayırıcı hiper düzlemi matematiksel olarak Denklem (6.5) ile ifade edilir.

$$\mathcal{H}_0: W^T \cdot X + b = 0 \quad (6.5)$$

Bu ifadede $W^T \cdot X$ bir iç çarpımı göstermekte olup W ağırlık vektörünü, X hiper düzlem üzerinde bulunan noktaları göstermektedir, b ise sabit bir değerdir. Denklem (6.5) aşağıdaki şekilde yazılabilir.

$$\sum_{i=1}^n w_i x_i + b = 0 \quad (6.6)$$

\mathcal{H}_1 ve \mathcal{H}_2 hiper düzlemleri de benzer şekilde yazılabilir:

$$\mathcal{H}_1: W^T \cdot X + b = 1 \quad (6.7)$$

$$\mathcal{H}_2: W^T \cdot X + b = -1 \quad (6.8)$$

\mathcal{H}_1 'de belirtilen hiper düzlemin üst tarafında kalan noktalar $y = +1$ sınıfına ait olup Denklem (6.9) ile ifade edilir:

$$w^T \cdot x + b \geq 1 \quad (6.9)$$

Benzer şekilde \mathcal{H}_2 'de belirtilen hiper düzlemin alt tarafında kalan noktalar $y = -1$ sınıfına ait olup Denklem (6.10) ile ifade edilir:

$$w^T \cdot x + b \leq -1 \quad (6.10)$$

Bu iki denklemin birleştirilmesiyle Denklem (6.11) ile gösterilen kısıt elde edilir:

$$y_i(w^T \cdot x + b) \geq 1 \quad y_i \in \{-1, +1\} \quad \forall i \quad (6.11)$$

Şekil 6.6'da gösterilen \mathcal{H}_1 ve \mathcal{H}_2 hiper düzlemleri üzerinde bulunan noktaların “destek vektör” olarak adlandırıldığından bahsedilmiştir. Herhangi bir A noktası üzerindeki destek vektörün; optimal hiper düzlem olan \mathcal{H}_0 ile arasındaki d uzaklığı, noktanın doğruya uzaklığı bağıntısından faydalanılarak şöyle ifade edilir:

$$d = \frac{|wx + b|}{\|w\|} \quad (6.12)$$

Bu durumda d uzaklığı \mathcal{H}_1 ya da \mathcal{H}_2 için verilen denklemler dikkate alınarak aşağıdaki gibi hesaplanabilir.

$$d = \frac{|\pm 1|}{\|w\|} = \frac{1}{\|w\|} \quad (6.13)$$

Son olarak marjin Denklem (6.14) ile aşağıdaki gibi ifade edilebilir.

$$m = \frac{2}{\|w\|} \quad (6.14)$$

6.2.1.1. Primal (özgün) model

SVM’de destek vektörler arasındaki uzaklığın enbüyüklenmesi ile optimal ayırıcı hiper düzleminin bulunması amaçlandığından W ’nun normunu ($\|w\|$) enküçüklemek marjini enbüyükleyecektir. Bu durumda Denklem (6.15)’teki ifade enküçüklenmelidir.

$$\|w\|^2 = w^T \cdot w = \sum_{i=1}^n w_i^2 \quad (6.15)$$

Böylece eniyileme problemi Denklem (6.16)’da gösterilen karesel programlama modeli şeklinde ifade edilebilir.

$$\begin{aligned} \text{enk } & \frac{1}{2} \|w\|^2 \\ \text{s. t.} & \\ & y_i(w^T \cdot x + b) \geq 1 \quad \forall i \end{aligned} \quad (6.16)$$

6.2.1.2. Dual (ikil) model

Doğrusal olmayan eniyileme problemlerinde, kısıtlar eşitlik biçiminde olduğunda lagrange çarpanları ile en iyi çözüm aranır. Kısıtların eşitsizlik olduğu durumlarda ise Karush-Kuhn-Tucker (KKT) koşulları sağlanacak şekilde çözüm araştırılır (Arora, 2011).

Lagrange fonksiyonunu elde etmek için kısıt lagrange çarpanıyla çarpılır ve bu çarpım amaç fonksiyonundan çıkarılır:

$$\begin{aligned} & \alpha_i \geq 0 \text{ olmak üzere} \\ L(w, b, \alpha) &= \frac{1}{2} (w^T \cdot w) - \sum_{i=1}^n \alpha_i [y_i(w^T \cdot x_i + b) - 1] \end{aligned} \quad (6.17)$$

Bu eniyileme probleminin çözümü semer noktasının (saddle point) belirlenmesi yoluyla yapılır. Bu noktada fonksiyon w ve b parametrelerine göre minimum değer alırken α langrange çarpanına göre maksimum değer alır (Cortes ve Vapnik, 1995).

$$\frac{\partial L(w, b, a)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (6.18)$$

$$\frac{\partial L(w, b, a)}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (6.19)$$

KKT koşulları ise aşağıdaki gibi elde edilir.

$$\sum_{i=1}^n \alpha_i y_i x_i = w \quad (6.20)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (6.21)$$

$$y_i(w^T \cdot x + b) - 1 \geq 0 \quad \forall i \quad (6.22)$$

$$\alpha_i \geq 0 \quad (6.23)$$

$$\alpha_i [y_i(w^T \cdot x + b) - 1] = 0 \quad \forall i \quad (6.24)$$

Oluşturulan denklemlerle Lagrange fonksiyonu düzenlenerek ikil form aşağıdaki gibi elde edilir:

$$L(w, b, \alpha) = \frac{1}{2}(w^T \cdot w) - \sum_{i=1}^n \alpha_i y_i w_i \cdot x_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \quad (6.25)$$

$$L(w, b, \alpha) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) - \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) + \sum_{i=1}^n \alpha_i \quad (6.26)$$

Son olarak yapılan düzenlemeler ile aşağıdaki ifade elde edilir:

$$L(w, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) \quad (6.27)$$

İkil eniyileme problemi ise Denklem (6.28) ile gösterilir:

$$\begin{aligned}
 \text{enb } L_D(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) \\
 &\quad \text{s. t.} \\
 &\quad \alpha_i \geq 0 \\
 &\quad \sum_{i=1}^n \alpha_i y_i = 0
 \end{aligned} \tag{6.28}$$

Eniyileme problemini ikil forma çevirmenin iki nedeni vardır. Öncelikle ikil form ile çözüme daha kolay ulaşılır. İkinci ve daha önemli neden ise, bu formülasyonla hem eğitim verisi hem de test verisi modelde vektörel iç çarpım şeklinde yer alacaktır. Bu durum doğrusal olmayan SVM modellerin çözümünde çekirdek fonksiyonların kullanılabilmesi adına oldukça önemlidir (Burges, 1998).

KKT koşullarında gösterilen $\alpha_i [y_i(w^T \cdot x + b) - 1] = 0 \quad \forall i$ eşitliği incelenirse iki durum ortaya çıkacaktır (Deng vd., 2013):

- i. Yardımcı hiper düzlemlerin üzerinde bulunmayan x gözlemleri $y_i(w^T \cdot x + b) - 1 > 0$ kısıtını sağlar. Bu durumda $\alpha_i = 0$ olmak zorundadır.
- ii. Yardımcı hiper düzlemlerin üzerinde bulunan x gözlemleri $y_i(w^T \cdot x + b) - 1 = 0$ eşitliğini sağlar. Bu durumda $\alpha_i > 0$ olmak zorundadır.

Sonuç olarak α_i 'nin 0'dan farklı bir değer olabilmesi için $y_i(w^T \cdot x + b)$ ifadesinin 1'e eşit olması gerekmektedir. Bu eşitlik yardımcı hiper düzlemler üzerinde yani destek vektörlerle (Bkz Denklem (6.7) ve (6.8)) sağlanır (Hamel, 2009). $\alpha_i > 0$ olan durumlarda yukarıdaki $y_i(w^T \cdot x + b) = 1$ eşitliğini sağlayan x_{sv} vektörleri destek vektörlerdir ve ait oldukları y_{sv} sınıfıyla belirlenen yardımcı hiper düzlem üzerinde yer alırlar.

\mathcal{H}_0 eniyi hiper düzlemini w ve b belirler (Bkz. Denklem (6.5)). KKT koşullarında belirtilen Denklem (6.20) kullanılarak aşağıdaki eşitlik elde edilir.

$$w^* = \sum_{i=1}^n \alpha_i y_i x_i \quad (6.29)$$

Bu durumda, x_{sv} destek vektörleri, üzerinde buldukları yardımcı hiper düzlemin denkemini sağlayacaktır;

$$y_{sv}(w^* \cdot x_{sv} + b^*) = 1 \quad (6.30)$$

$$b^* = \frac{1}{y_{sv}} - w^* \cdot x_{sv} \quad (6.31)$$

$y_{sv} \in \{-1, +1\}$ olduğundan Denklem (6.31) aşağıdaki şekilde yazılabilir.

$$b^* = y_{sv} - w^* \cdot x_{sv} \quad (6.32)$$

Böylece x test verisinin hangi sınıfa ait olduğu araştırılırken Denklem (6.33)'de verilen karar fonksiyonu kullanılır.

$$f(x) = \text{sgn}(w^* \cdot x + b^*) \quad (6.33)$$

Destek vektör makinesi ile bir sınıflama probleminin matematiksel çözümü Ek Açıklama-B'de verilmiştir.

6.2.2. Bir miktar hata ile doğrusal ayrılabilme durumu (soft margin)

Eğitim veri kümesinin hatasız ayrılabilmediği durumlar, veri kümesinin gürültülü veya aykırı veri içermesi gibi nedenlerden dolayı gerçek hayat problemlerinde pek fazla görülmemektedir (Cristianini ve Shawe-Taylor, 2000). Şekil 6.7'de görüldüğü üzere bu gibi durumlarda verinin hatasız bir biçimde doğrusal olarak ayrılması mümkün olamamaktadır.

Bu problemin üstesinden gelebilmek için Cortes ve Vapnik (1995) tarafından soft marjın eniyilemesi olarak adlandırılan yöntem önerilmiştir. Bu yönteme göre eğitim kümesinde hataya izin verilebilmesi adına daha esnek bir model yapısı gerekmektedir. Bu esnekliği sağlamak için hatalı olarak sınıflandırılan gözlemlerin varlığında, adı geçen gözlemlerin ait oldukları yardımcı hiper düzlemlerle olan ilişkisini gösteren ξ_i aylak

değişkeni modele eklenir. Aylak değişkenlerin modele eklenmesiyle aşağıdaki denklemler yazılabilir.

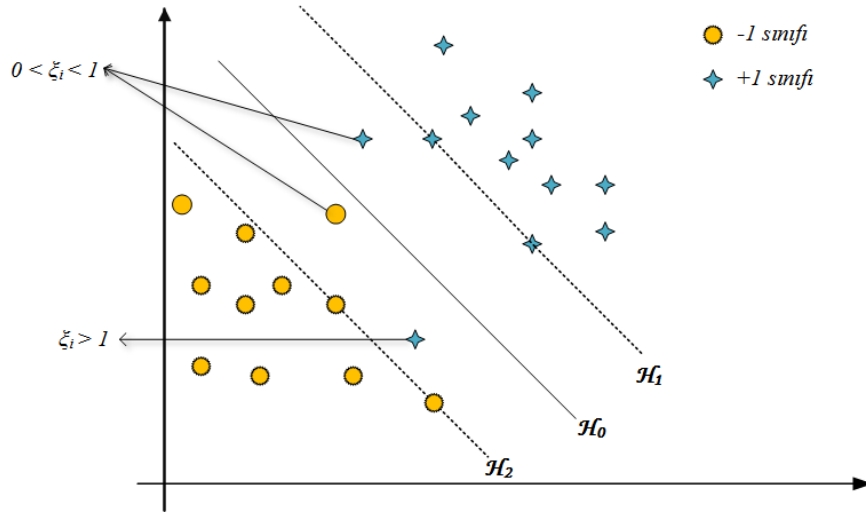
$$w^T \cdot x + b + \xi_i \geq 1 \quad y = +1 \text{ için} \quad (6.34)$$

$$w^T \cdot x + b - \xi_i \leq -1 \quad y = -1 \text{ için} \quad (6.35)$$

$$\xi_i \geq 0 \quad \forall_i \quad (6.36)$$

İki iki kısıtın birleştirilmesiyle Denklem (6.37) elde edilir.

$$y_i(w^T \cdot x + b) \geq 1 - \xi_i \quad y_i \in \{-1, +1\} \quad (6.37)$$



Şekil 6.7. Soft marjlinli destek vektör sınıflandırıcı

Şekil 6.7'de yer alan $\xi_i > 1$ ifadesi, ayırıcı hiper düzlemin yanlış tarafında kalan gözlemleri tanımlarken $0 < \xi_i < 1$ ifadesi ise hiper düzlemin doğru tarafında olan fakat marjin bölgesi içerisinde kalan gözlemleri tanımlamaktadır.

Makinenin öğrenme sürecinde VC teorisinde belirtildiği gibi marjin enbüyüklenmeye çalışılırken eğitim hatası da en küçüklenmelidir. Bu ödünleşmeyi kontrol edebilmek adına C “düzenleme parametresi” amaç fonksiyonuna ilave edilir (Soman vd., 2009).

$$\text{enk } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

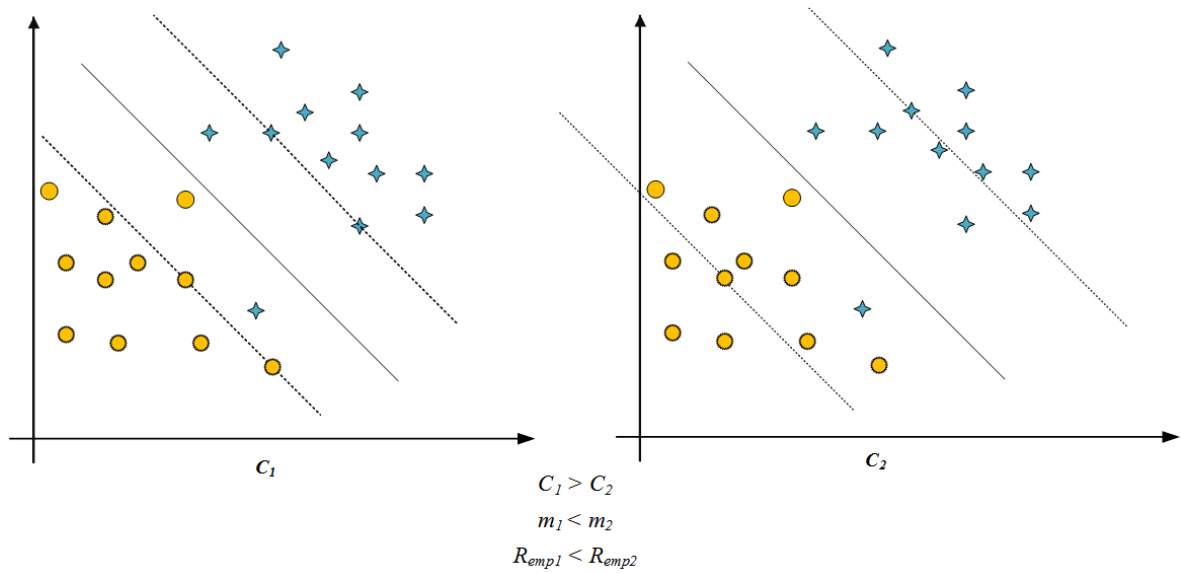
s. t.

$$y_i(w^T \cdot x + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

(6.38)

Burada $C > 0$ olmak üzere kullanıcı tarafından belirlenen bir parametredir. C değerinin azaltılmasıyla marjin genişliği büyüyecek buna karşılık beklenen eğitim hatası artacaktır (Zaki ve Meira, 2014). C parametresine büyük değer verilmesi ise aşırı uyuma neden olabilmektedir. C parametresinin modelde gerçekleştirdiği bu etkiler Şekil 6.8'de gösterilmiştir.



Şekil 6.8. C parametresinin modele olan etkisi

Amaç fonksiyonuna α ve β Lagrange çarpanlarının eklenmesiyle enyileme modelinin Lagrange fonksiyonu elde edilir;

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(w^T \cdot x + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i$$

(6.39)

Bu eniyileme probleminin çözümü de benzer şekilde, semer noktasının belirlenmesi yoluyla yapılır. Bu noktada fonksiyon w , b ve ξ parametrelerine göre minimum değerini alırken α ve β langrange çarpanlarına göre maksimum değerini alır (Kecman, 2001). Çözüm adına yapılanlar Denklem (6.40)-(6.42)'de gösterilmiştir.

$$\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (6.40)$$

$$\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0 \quad (6.41)$$

$$\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial \xi} = C - \alpha_i - \beta_i = 0 \quad (6.42)$$

Böylece eniyileme problemini dual formda çözebilmek için gerekli olan KKT koşulları şöyle yazılabilir:

$$\sum_{i=1}^n \alpha_i y_i x_i = w \quad (6.43)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (6.44)$$

$$C = \alpha_i + \beta_i \quad (6.45)$$

$$y_i(w^T \cdot x + b) - 1 + \xi_i \geq 0 \quad (6.46)$$

$$\xi_i \geq 0 \quad (6.47)$$

$$\alpha_i \geq 0 \quad (6.48)$$

$$\beta_i \geq 0 \quad (6.49)$$

$$\alpha_i [y_i(w^T \cdot x + b) - 1 + \xi_i] = 0 \quad (6.50)$$

$$\beta_i \xi_i = 0 \quad (6.51)$$

Denklemlerin yerine koyulmasıyla Lagrange fonksiyonu şu şekilde de ifade edilebilir:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) \quad (6.52)$$

Dual eniyileme problemi ise Denklem (6.53)'teki gibi olacaktır.

$$\begin{aligned} \text{enb } L_D(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T \cdot x_j) \\ &\text{s. t.} \\ &\sum_{i=1}^n \alpha_i y_i = 0 \\ &0 \leq \alpha_i \leq C \end{aligned} \quad (6.53)$$

Buradaki amaç fonksiyonunun, doğrusal ayrılabilen durum için elde edilen fonksiyonla aynı olduğu görülmektedir. İki durum arasındaki tek fark $C = \alpha_i + \beta_i$ ve $\alpha_i, \beta_i \geq 0$ koşullarıyla sağlanan $0 \leq \alpha_i \leq C$ kısıtının eniyileme modeline eklenmesidir. Böylece α_i Lagrange çarpanına bir üst sınır getirilmiş olur (Burges, 1998).

KKT koşulları incelenecek olursa α_i için üç farklı durum söz konusu olacaktır (Abe, 2010):

- i. $\alpha_i = 0$ olursa, $C = \alpha_i + \beta_i$ eşitliğinden dolayı $\beta_i = C$ olacaktır. Dolayısıyla $\beta_i \xi_i = 0$ koşulunu sağlamak adına $\xi_i = 0$ olmalıdır. Bu durumda x_i doğru sınıflandırılmıştır ve marjinin dışında kalan bölgededir.
- ii. $0 < \alpha_i < C$ olan durumlarda $\alpha_i [y_i (w^T \cdot x + b) - 1 + \xi_i] = 0$ eşitliğini sağlamak adına $y_i (w^T \cdot x + b) = 1$ olduğundan $\xi_i = 0$ olacaktır. Yani x_i bir destek vektörü ifade etmektedir.
- iii. $\alpha_i = C$ olursa, $C = \alpha_i + \beta_i$ eşitliğinden dolayı $\beta_i = 0$ olacaktır. $\beta_i \xi_i = 0$ eşitliğine göre $\xi_i \geq 0$ olacaktır. Böylece,

- a. $0 \leq \xi_i < 1$ durumu x_i gözleminin olası iki şekilde doğru sınıflandırıldığını gösterir.
- $\xi_i = 0$ ise x_i yardımcı hiper düzlem üzerindedir.
 - $0 < \xi_i < 1$ ise marjin bölgesi içerisinde.
- b. $\xi_i \geq 1$ durumunda ise x_i aşağıdaki şekillerde yanlış sınıflandırılmıştır.
- $\xi_i = 1$ ise x_i ayırıcı hiper düzlem üzerindedir.
 - $\xi_i > 1$ ise x_i ayırıcı düzlemin yanlış tarafındadır.

Karar fonksiyonu ile w^* ve b^* değerleri de doğrusal ayrılabilir durumla aynı şekilde ifade edilmektedir:

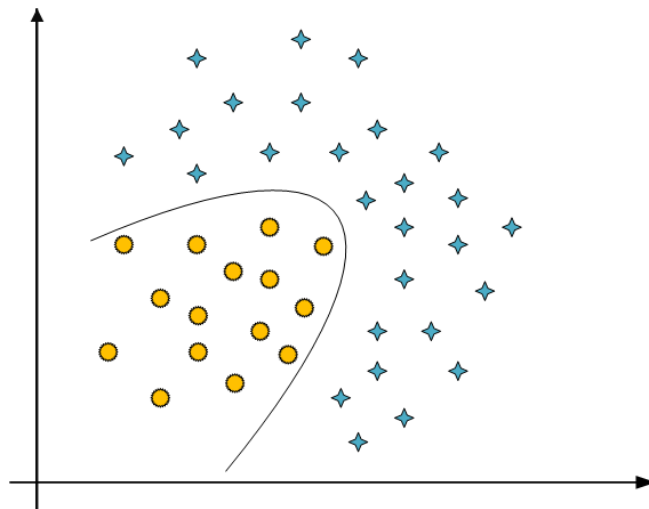
$$w^* = \sum_{i=1}^n \alpha_i y_i x_i \quad (6.54)$$

$$b^* = y_{sv} - w^* \cdot x_{sv} \quad (6.55)$$

$$f(x) = \text{sgn}(w^* \cdot x + b^*) \quad (6.56)$$

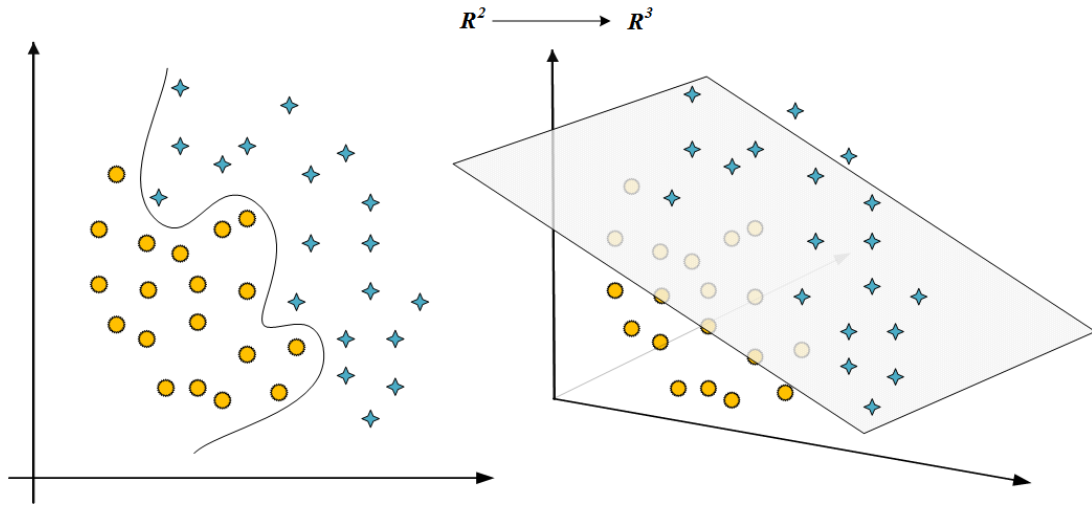
6.2.3. Doğrusal olmayan destek vektör sınıflandırıcı

Doğrusal olmayan SVM, Şekil 6.9'da görüldüğü gibi, veri kümesinin doğrusal bir fonksiyonla hatasız (hard margin) veya belirli bir hata ile (soft margin) ayrılamaması durumunda kullanılan yöntemdir.



Şekil 6.9. Doğrusal ayrılamama durumu

SVM’de ayırma işlemi doğrusal bir karar düzlemi ile gerçekleştirmek üzere tasarlandığından, veri kümesinin bulunduğu uzayda ayırma eğrisini tahmin edebilmesi mümkün olamamaktadır. Bu sorunun üstesinden gelebilmek adına çekirdek fonksiyonları ile girdi uzayındaki iç çarpımlar, Şekil 6.10’da gösterildiği gibi yüksek boyutlu öznitelik uzayına (Hilbert uzayı) eşlenerek doğrusal olarak ayırlamayan durumlar doğrusal ayrılabilir şekle dönüştürülmektedir. (Schölkopf ve Smola, 2002)



Şekil 6.10. Verinin üst uzayda doğrusal ayrılabilir duruma dönüştürülmesi

SVM ilk olarak doğrusal sınıflama problemleri için türetilmişken Denklem (6.52)’de görüldüğü gibi; hem eğitim hem de test aşamalarında sadece veri kümesine ait örnekler arasındaki iç çarpım işlemi gerektiğinden, çeşitli çekirdek (kernel) fonksiyonlarının tanımlanmasıyla doğrusal olmayan durumlara da genişletilebilmektedir (Zhang, 1999). Bu durumda, ilgili denklemdeki $x_i^T \cdot x_j$ iç çarpımının yüksek öznitelik uzayı için yeniden tanımlanması gerekmektedir. Bu nedenle x_i^T ’nin dönüşümü için $\phi(x_i)^T$, benzer şekilde x_j ’nin dönüşümü için de $\phi(x_j)$ fonksiyonu kullanılır (Abe, 2010). Bu dönüşüm; I girdi (input) uzayını, F ise öznitelik (feature) uzayını temsil etmek üzere şöyle ifade edilebilir:

$$\phi: I = \mathbb{R}^2 \rightarrow F = \mathbb{R}^3 \quad (6.57)$$

Örnek olarak (x_1, x_2) girdi uzayı üzerinde yapılan $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ dönüşümünün geometrik ifadesi Şekil 6.11’de gösterilmiştir (Igel, 2014):

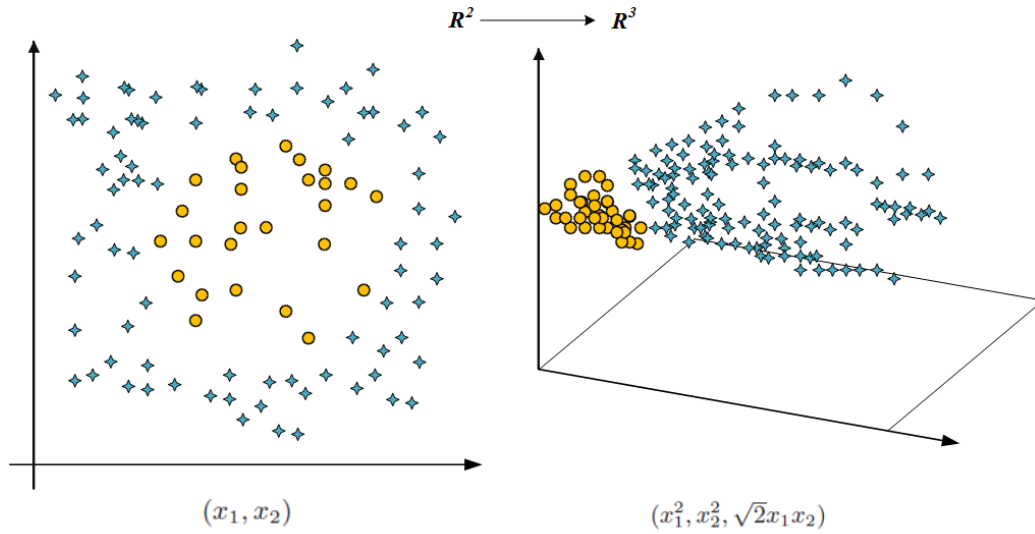
Bu dönüşümlerle öznitelik uzayı için ikil eniyileme modelinin amaç fonksiyonu aşağıdaki şekilde olacaktır.

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j [\phi(x_i)^T \cdot \phi(x_j)] \quad (6.58)$$

Karar fonksiyonu ve w^* ise aşağıdaki gibi ifade edilir.

$$w^* = \sum_{i=1}^n \alpha_i y_i \phi(x_i) \quad (6.59)$$

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i \phi(x_i) \phi(x_j) + b^* \right) \quad (6.60)$$



Şekil 6.11. Girdi uzayının $\phi(x)$ dönüşümü ile üst uzayda eşlenmesi

$\phi(x_i)^T \cdot \phi(x_j)$ çarpımının $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$ şeklinde bir çekirdek fonksiyon olarak tanımlanabilmesi için Mercer teoreminin koşullarını sağlaması gerekmektedir (Cristianini ve Shawe-Taylor, 2000). Bu teoreme göre çekirdek fonksiyonu (Schölkopf, 2000; Genton, 2001; Hofmann, 2006):

- Simetrik fonksiyon olmalıdır.

$$K(x, z) = K(z, x) \quad (6.61)$$

- Cauchy-Schwarz eşitsizliğini sağlamalıdır.

$$\begin{aligned} K(x, z)^2 &= [\phi(x) \cdot \phi(z)]^2 \leq \|\phi(x)\|^2 \|\phi(z)\|^2 \\ &\leq \phi(x)^T \cdot \phi(x) \cdot \phi(z)^T \cdot \phi(z) \\ &\leq K(x, x)K(z, z) \end{aligned} \quad (6.62)$$

- Her bir $K(x, z)$ ile oluşturulan K_{ij} matrisi yarı pozitif tanımlı Gram matrisi olarak gösterilebilmelidir.

$$K_{ij} = \begin{bmatrix} K(x_1 z_1) & \cdots & K(x_n z_1) \\ \vdots & \ddots & \vdots \\ K(x_1 z_n) & \cdots & K(x_n z_n) \end{bmatrix} \quad (6.63)$$

$$\sum_{i,j=1}^n \lambda_i \lambda_j K_{ij} \geq 0 \quad \lambda_i, \lambda_j \in \mathbb{R} \quad (6.64)$$

Sonuç olarak bu koşullar, yüksek boyutlu uzayda çalışılmasını gerektiren durumlarda çekirdek fonksiyonlarının, iki girdi dönüşüm vektörünün iç çarpımı olarak ifade edilmesini sağlamaktadır. Bu yöntem literatürde çekirdek hilesi (kernel trick) olarak adlandırılmaktadır. Çekirdek hilesinin kullanılmasıyla öznitelik uzayında hesaplanması gereken $\phi(x_i)^T \cdot \phi(x_j)$ iç çarpımını hesaplamak yerine; çekirdek fonksiyonu, girdi uzayı vektörlerini kullanarak işlemleri gerçekleştirmektedir (Kecman, 2001). Böylece her bir girdi uzayı, vektörünün öznitelik uzayındaki değerlerinin bilinmesine gerek duyulmadan bu uzaydaki iç çarpımları hesaplanabilecektir. Bu durumda hesaplama maliyetinden kazanç sağlanacaktır (Hoffman vd., 2008). Çekirdek hilesinin sağladığı bu avantajı bir örnek üzerinde göstermek gerekirse,

$x \in \mathbb{R}^2$ girdi vektörünü \mathbb{R}^3 'e taşıyacak olan ϕ dönüşüm fonksiyonu aşağıdaki gibi tanımlansın:

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2) \quad (6.65)$$

Bu durumda $x \cdot z$ iç çarpımı, gerekli dönüşümler yapıldıktan sonra öznitelik uzayında şu şekilde hesaplanır:

$$\begin{aligned}\phi(x)\phi(z) &= \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} (z_1^2, z_2^2, \sqrt{2}z_1z_2) \\ &= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2\end{aligned}\quad (6.66)$$

Aynı hesaplama $K(x, z) = (x^T z)^2$ şeklinde tanımlanan bir çekirdek fonksiyonuyla yapılacak olursa aşağıdaki eşitlik elde edilir.

$$\begin{aligned}K(x, z) &= \left[\begin{pmatrix} x_1 \\ z_1 \end{pmatrix} (z_1, z_2) \right]^2 = (x_1z_1 + x_2z_2)^2 \\ &= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2\end{aligned}\quad (6.67)$$

Örnekte de görüldüğü gibi çekirdek fonksiyonu sayesinde herhangi bir dönüşüm fonksiyonu kullanılmasına gerek kalmadan öznitelik uzayındaki iç çarpım kolaylıkla hesaplanmıştır. Literatürde kullanılan bazı çekirdek fonksiyonları Çizelge 6.1’de verilmiştir.

Çizelge 6.1. Çekirdek fonksiyonları

Çekirdek Fonksiyonu	Matematiksel İfadesi
Polinom	$K(x, x') = (x^T x' + c)^d$
Gauss (Radyal)	$K(x, x') = \exp(-\gamma \ x - x'\ ^2), \quad \gamma = \frac{1}{2\sigma^2}$
Sigmoid	$K(x, x') = \tan(\gamma x^T x' + c)$
Pearson VII	$K(x, x') = \frac{H}{\left[1 + \left(\frac{2(x - x')\sqrt{2^{(1/\omega)} - 1}}{\sigma} \right)^2 \right]^\omega}$

Çekirdek fonksiyonun kullanılmasıyla dual eniyileme modelinin amaç fonksiyonu şu şekilde gösterilir:

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (6.68)$$

Karar fonksiyonu ve b değeri ise aşağıdaki şekilde gösterilir.

$$b^* = y_{sv} - \sum_{i=1}^n \alpha_i y_i K(x_i, x_{sv}) \quad (6.69)$$

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b^* \right) \quad (6.70)$$

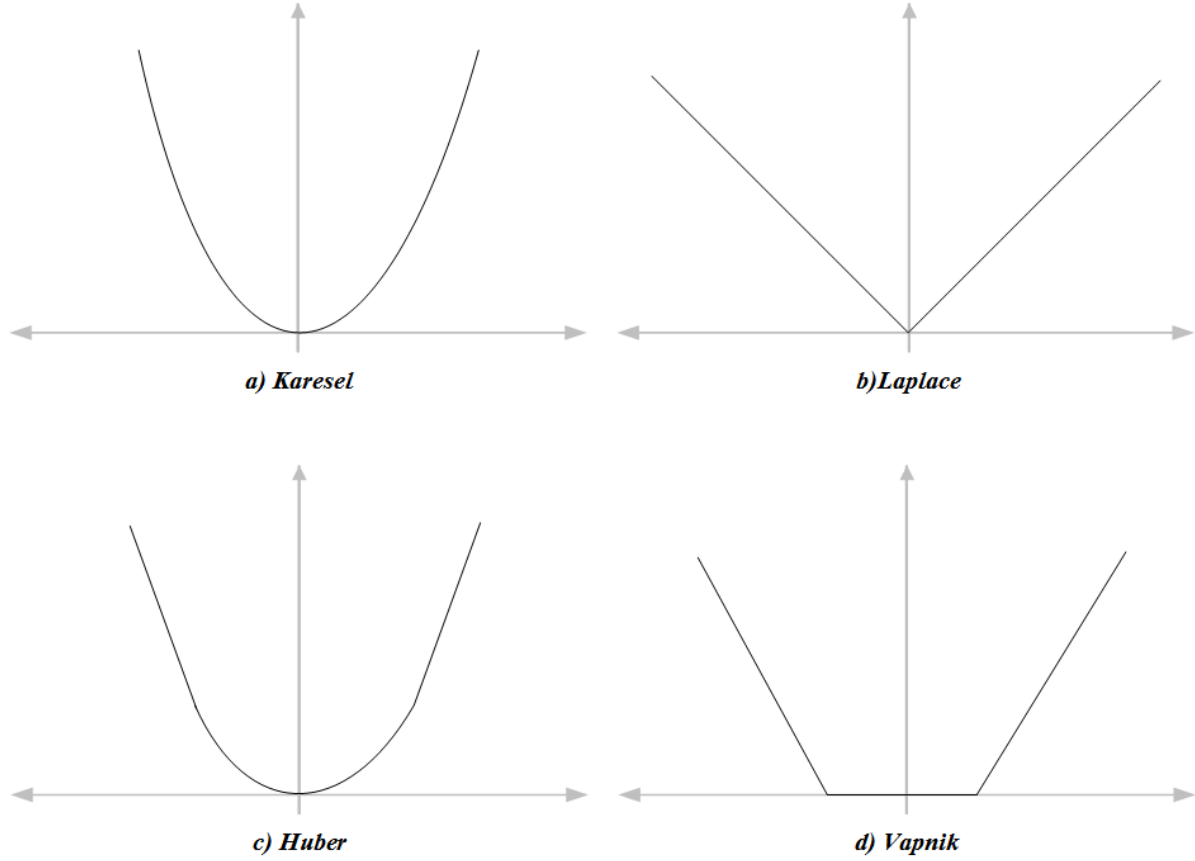
w^* , karar fonksiyonunda yer almadığından değerinin belirlenmesine dolayısıyla $\phi(x_i)$ dönüşümün de yapılmasına gerek duyulmamaktadır.

6.3. Destek Vektör Regresyon

Destek vektör regresyonun (SVR) temel mantığı, VC teorisi kapsamında eğitim hatasını enküçüklemek yerine, beklenen risk hatasını enküçükleyecek regresyon fonksiyonunu belirlemeye çalışmaktır. SVR’de optimal genelleştirme, destek vektörler arasındaki marjinin, belirli bir hata toleransı içerisinde enbüyüklenmesiyle araştırılır. Bu tolerans ise kayıp fonksiyonu aracılığıyla belirlenir. Şekil (6.12)’de bazı kayıp fonksiyonları örnek olarak verilmiştir (Smola, 1996).

Şekil 6.12 (a)’da verilen karesel kayıp fonksiyonu en küçük kareler hata ölçütünü temsil etmektedir. Şekil 6.12 (b)’de ise aykırı gözlemlere karesel kayıp fonksiyonuna kıyasla daha az duyarlılık gösteren Laplace kayıp fonksiyonu gösterilmiştir. Şekil 6.12 (c)’de yer alan Huber kayıp fonksiyonu, veri setinin dağılımının bilinmediği durumlarda optimal özellikli sağlam (robust) bir kayıp fonksiyonu olarak tanımlanmaktadır (Gunn, 1999). Bu üç kayıp fonksiyonun benzer noktası destek vektör sayısını arttırması beklenen seyreklik (sparseness) özelliğine sahip olmamalarıdır (Cherkassky ve Ma, 2002). Bu nedenle SVR’de

marjin oluşumunu sağlamak adına Vapnik (1995) tarafından Şekil 6.12 (d)'deki ε -duyarsız (ε -insensitive) kayıp fonksiyonu önerilmiştir.



Şekil 6.12. Kayıp fonksiyonları

Tanımlanan ε -duyarsız kayıp fonksiyonu, regresyon modelinde $[-\varepsilon, \varepsilon]$ aralığındaki hata değerine sahip verileri tolere ederek marjinin içinde kalmalarına izin vermektedir (Girma, 2009). Böylece $[-\varepsilon, \varepsilon]$ aralığındaki hatalar dikkate alınmamış olacaktır. ε -duyarsız kayıp fonksiyonu matematiksel olarak aşağıdaki gibi ifade edilir (Müller vd., 1997).

$$L_{\varepsilon}(y) = \begin{cases} 0 & |y - f(x)| \leq \varepsilon \text{ ise} \\ |y - f(x)| - \varepsilon & d. d. \end{cases} \quad (6.71)$$

Bu fonksiyonda $f(x)$, $w^T \cdot x_i + b$ düzlemini temsil etmektedir.

6.3.1. Doğrusal destek vektör regresyon

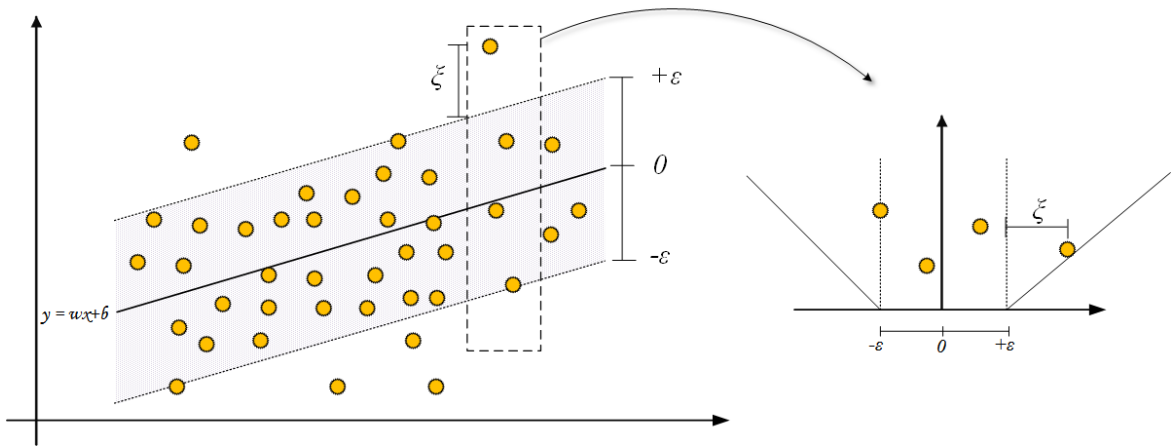
Kayıp fonksiyonunun kullanılmasıyla sınıflama problemlerine benzer şekilde; x_i regresyon modelinin bağımsız değişkenini, y_i ise bağımlı değişkenini temsil etmek üzere tüm verilerin ε tüpü ($[-\varepsilon, \varepsilon]$ hata sınırı) içerisinde kaldığı varsayılırsa Denklem (6.72) yazılabilir.

$$-\varepsilon \leq y_i - w^T \cdot x_i - b \leq \varepsilon \quad (6.72)$$

Ancak Şekil 6.13'te görüldüğü gibi, bu durumun mümkün olmadığı ya da ε tüpünden sapmalara izin verilmek istenildiğinde, soft margin yaklaşımındaki gibi sapmaları temsil edecek aylak değişkenlere gerek duyulur. Böylece denklemdeki her bir kısıt için $\xi_i^+, \xi_i^- \geq 0$ şeklinde aylak değişkenler tanımlanır.

$$y_i - w^T \cdot x_i - b \leq \varepsilon + \xi_i^+ \quad (6.73)$$

$$y_i - w^T \cdot x_i - b \geq -\varepsilon - \xi_i^- \quad (6.74)$$



Şekil 6.13. Doğrusal destek vektör regresyon (Schölkopf ve Smola'dan, 2002)

Sınıflama problemine benzer şekilde genelleştirme yeteneği yüksek bir fonksiyon belirlemek adına marjin $[-\varepsilon, \varepsilon]$ duyarlılığında enbüyüklenmelidir. Bu durumda eniyileme modeli aşağıdaki şekilde gösterilir.

$$\begin{aligned}
&enk \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \\
&\text{s. t.} \\
&y_i - w^T \cdot x_i - b \leq \varepsilon + \xi_i^+ \\
&w^T \cdot x_i + b - y_i \leq \varepsilon + \xi_i^- \\
&\xi_i^+, \xi_i^- \geq 0
\end{aligned} \tag{6.75}$$

Amaç fonksiyonunda kullanılan C parametresi, ε 'dan büyük olmasına karşın kabul edilecek sapmalar ile regresyon fonksiyonunun genelleştirme yeteneği arasındaki ödünleşmeyi temsil etmektedir (Drucker vd., 1997).

Modele $\alpha, \beta, \lambda, \theta$ Lagrange çarpanlarının eklenmesiyle amaç fonksiyonu aşağıdaki şekilde yazılır.

$$\begin{aligned}
L(w, b, \xi^+, \xi^-, \alpha, \beta, \lambda, \theta) &= \frac{1}{2} w^T w + C \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \\
&\quad - \sum_{i=1}^n \lambda_i \xi_i^+ - \sum_{i=1}^n \theta_i \xi_i^- \\
&\quad - \sum_{i=1}^n \alpha_i [-y_i + w^T \cdot x_i + b + \varepsilon + \xi_i^+] \\
&\quad - \sum_{i=1}^n \beta_i [y_i - w^T \cdot x_i - b + \varepsilon + \xi_i^-]
\end{aligned} \tag{6.76}$$

Kısmi türevler alınarak KKT koşulları aşağıdaki gibi elde edilir.

$$\frac{\partial L(w, b, \xi^+, \xi^-, \alpha, \beta, \lambda, \theta)}{\partial w} = w - \sum_{i=1}^n (\alpha_i - \beta_i) x_i = 0 \tag{6.77}$$

$$\frac{\partial L(w, b, \xi^+, \xi^-, \alpha, \beta, \lambda, \theta)}{\partial b} = - \sum_{i=1}^n (\alpha_i - \beta_i) = 0 \tag{6.78}$$

$$\frac{\partial L(w, b, \xi^+, \xi^-, \alpha, \beta, \lambda, \theta)}{\partial \xi^+} = C - \lambda_i - \alpha_i = 0 \tag{6.79}$$

$$\frac{\partial L(w, b, \xi^+, \xi^-, \alpha, \beta, \lambda, \theta)}{\partial \xi^-} = C - \theta_i - \beta_i = 0 \quad (6.80)$$

$$-y_i + w^T \cdot x_i + b + \varepsilon + \xi_i^+ \geq 0 \quad (6.81)$$

$$y_i - w^T \cdot x_i - b + \varepsilon + \xi_i^- \geq 0 \quad (6.82)$$

$$\xi_i^+, \xi_i^- \geq 0 \quad (6.83)$$

$$\alpha_i, \beta_i, \lambda_i, \theta_i \geq 0 \quad (6.84)$$

$$\alpha_i[-y_i + w^T \cdot x_i + b + \varepsilon + \xi_i^+] = 0 \quad (6.85)$$

$$\beta_i[y_i - w^T \cdot x_i - b + \varepsilon + \xi_i^-] = 0 \quad (6.86)$$

$$\lambda_i \xi_i^+ = 0 \quad (6.87)$$

$$\theta_i \xi_i^- = 0 \quad (6.88)$$

Bu koşullar kullanılarak Lagrange fonksiyonu Denklem (6.89)'deki gibi elde edilir:

$$\begin{aligned} L(\alpha, \beta) &= \sum_{i=1}^n y_i(\alpha_i - \beta_i) - \varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) \\ &\quad - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \beta_i)(\alpha_j - \beta_j)(x_i^T \cdot x_j) \end{aligned} \quad (6.89)$$

Dual eniyileme problemi ise Denklem (6.90) ile aşağıdaki gibi gösterilir.

$$\begin{aligned} \text{enb } L_D(\alpha, \beta) &= \sum_{i=1}^n y_i(\alpha_i - \beta_i) - \varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) \\ &\quad - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \beta_i)(\alpha_j - \beta_j)(x_i^T \cdot x_j) \\ &\quad \text{s. t} \\ &\quad \sum_{i=1}^n (\alpha_i - \beta_i) = 0 \\ &\quad 0 \leq \alpha_i \leq C \\ &\quad 0 \leq \beta_i \leq C \end{aligned} \quad (6.90)$$

Lagrange çarpanları adına yorum yapabilmek için KKT koşullarıyla elde edilen aşağıdaki eşitlikler incelenirse,

$$\alpha_i[-y_i + w^T \cdot x_i + b + \varepsilon + \xi_i^+] = 0 \quad (6.91)$$

$$\beta_i[y_i - w^T \cdot x_i - b + \varepsilon + \xi_i^-] = 0 \quad (6.92)$$

$$(C - \alpha_i)\xi_i^+ = 0 \quad (6.93)$$

$$(C - \beta_i)\xi_i^- = 0 \quad (6.94)$$

- i. ε tütünün içinde kalan gözlemler için $|y_i - f(x)| < \varepsilon$ eşitsizliği geçerli olacaktır. Bu durumda KKT koşullarının sağlanması için α_i ve β_i lagrange çarpanlarının ikisinin de sıfıra eşit olması gerekmektedir.
- ii. ε tütünün sınırları üzerinde kalan gözlemler için $|y_i - f(x)| - \varepsilon = 0$ eşitliği geçerli olacaktır. Bu durumda x gözlemi aynı anda iki sınırdaki da yer alamayacağından Lagrange çarpanlarından birisi sıfırdan büyükken diğeri sıfıra eşit olacaktır (Deng vd., 2013).
- iii. ε tütünün sınırları dışında kalan gözlemler için $\xi_i^+ > 0$ ise α_i , $\xi_i^- > 0$ ise β_i C ceza parametresine eşit olacaktır (Campbell ve Ying, 2011).

Yaklaşım hatası ε değerinden büyük ya da eşit olan veri noktaları için $(\alpha_i - \beta_i)$ sıfırdan farklı bir değer alacaktır. Dolayısıyla bu noktalar destek vektörler olacaktır. Genel olarak ε toleransının büyük olması destek vektör sayısını azaltmaktadır. Ancak bununla beraber regresyon fonksiyonun aykırı verilere uyum göstermeye çalışması böylece genelleştirme yeteneğinde düşüş gerçekleşmesi beklenir (Tay ve Cao, 2001).

Son olarak regresyon fonksiyonu ile w^* ve b^* değerleri şu şekilde ifade edilmektedir:

$$w^* = \sum_{i=1}^n (\alpha_i - \beta_i)x_i \quad (6.95)$$

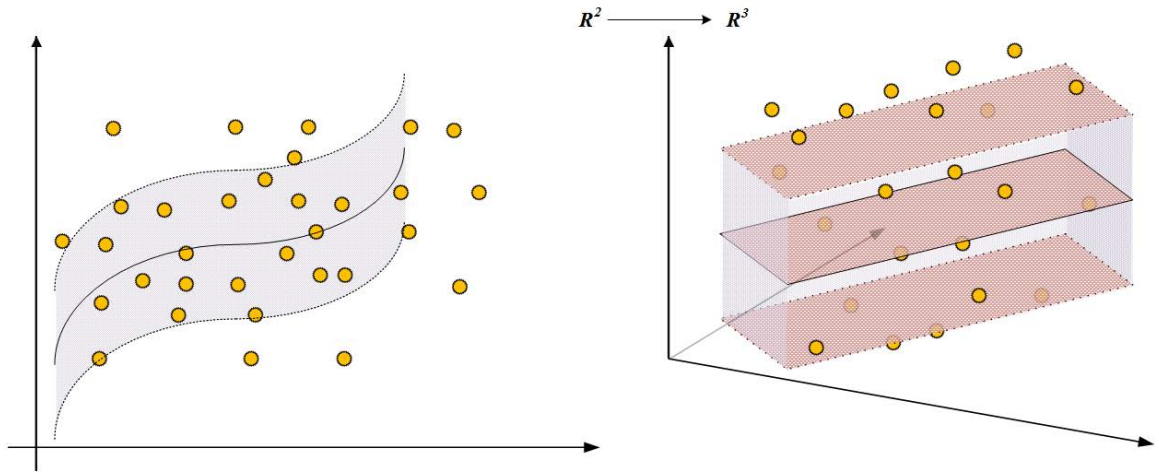
$$b^* = y_{sv\alpha} - \sum_{i=1}^n (\alpha_i)(x_i, x_{sv\alpha}) - \varepsilon \quad \alpha_i > 0 \text{ ise} \quad (6.96)$$

$$b^* = y_{sv\beta} - \sum_{i=1}^n (\beta_i)(x_i, x_{sv\beta}) + \varepsilon \quad \beta_i > 0 \text{ ise} \quad (6.97)$$

$$y_i = \sum_{i=1}^n (\alpha_i - \beta_i)(x \cdot x_i) + b^* \quad (6.98)$$

6.3.2. Doğrusal olmayan destek vektör regresyon

Sınıflandırma algoritmasına benzer olarak destek vektör regresyonda da girdi uzayındaki veriler, dual amaç fonksiyonunda iç çarpım olarak ifade edildiğinden doğrusal olmayan durumlarda çekirdek hilesi kullanılmaktadır. Böylece Şekil 6.14'te görüldüğü gibi girdi uzayında doğrusal olmayan fonksiyonlar, öznitelik uzayına genişletilerek doğrusal öğrenim makineleri ile elde edilebilirler (Christianini ve Shawe-Taylor, 2000).



Şekil 6.14. Doğrusal olmayan destek vektör regresyon

Çekirdek fonksiyonunun kullanılmasıyla dual eniyileme problemi ise aşağıdaki gibi gösterilir:

$$\begin{aligned}
 \text{enb } L_D(\alpha, \beta) &= \sum_{i=1}^n y_i(\alpha_i - \beta_i) - \varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) \\
 &\quad - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \beta_i)(\alpha_j - \beta_j)K(x_i^T, x_j) \\
 &\quad \text{s. t.} \tag{6.99} \\
 &\quad \sum_{i=1}^n (\alpha_i - \beta_i) = 0 \\
 &\quad 0 \leq \alpha_i \leq C \\
 &\quad 0 \leq \beta_i \leq C
 \end{aligned}$$

Regresyon fonksiyonu ve b^* ise aşağıdaki gibi ifade edilir:

$$b^* = y_{sv\alpha} - \sum_{i=1}^n (\alpha_i)K(x_i, x_{sv\alpha}) - \varepsilon \quad \alpha_i > 0 \text{ ise} \quad (6.100)$$

$$b^* = y_{sv\beta} - \sum_{i=1}^n (\beta_i)K(x_i, x_{sv\beta}) + \varepsilon \quad \beta_i > 0 \text{ ise} \quad (6.101)$$

$$y_i = \sum_{i=1}^n (\alpha_i - \beta_i)K(x, x_i) + b^* \quad (6.102)$$

Dört farklı çekirdek fonksiyonu kullanılan destek vektör regresyon için açıklayıcı görsel bir örnek Ek Açıklama-C'de verilmiştir (Gunn,1998).

7. MATERYAL VE YÖNTEM

Bu bölümde, öncelikle çalışmada materyal olarak kullanılan veri setine ilişkin bilgiler ve ham veri setinin işleme süreci anlatılmıştır. Devamında ise MLR, MLP ve SVR yöntemlerinin uygulamasında kullanılan WEKA yazılımının anlatımı üzerinde durulmuştur. Son olarak genel tahminleme modelinin karakutu yapısı, eğitim ve test veri kümeleri oluşturulurken kullanılan yöntemler, tahminleme modellerini değerlendirmek amacıyla kullanılacak olan performans göstergeleri, tahminleme çalışması için kullanılan yöntemlere ilişkin parametrelerin belirlenmesi ve bu parametrelerin tahminleme modellerine olan etkisi yer almaktadır.

7.1. Veri Seti ve Veri Önışleme

Denizli ili için kirletici değerleri T.C. Çevre ve Şehircilik Bakanlığına ait, 29° 05' 13" enlemi ve 37° 46' 53" boylamında yer alan “Denizli-2” isimli istasyon tarafından ölçülmüştür. Şubat 2005-Mayıs 2015 tarihleri arasında ölçülen bu değerler <http://www.havaizleme.gov.tr/> web adresinden çekilmiştir. Meteorolojik veriler ise Denizli iline ait 172370 kodlu meteoroloji istasyonu tarafından ölçülmüş olup ABD Ulusal Okyanus ve Atmosfer Dairesi (NOAA)'nin veri tabanından çekilmiştir. Ham veri setinde kullanılan değişkenler ve değişkenlerin birimi, modelde kullanılan simgeleri, betimsel istatistikleri Çizelge 7.1'de verilmiştir.

Bazı uygulamalarda, üzerinde çalışılacak veri kümesi; eksik, tutarsız, hatalı ve anlamsız veri, aşırı uç değerler gibi olumsuzlukların var olduğu durumları içerebilir. Bu tür veriler çözümü üzerinde olumsuz etkiler yaratacağından, ayrıca farklı algoritmalar nümerik, nominal, ikili gibi farklı veri tiplerine ihtiyaç duyduğundan veri setinin bir önışlemeden geçmesi gerekmektedir (Tapkan vd., 2011). Bu çalışmada veri seti üzerinde yapılan önışlemler; veri bütünleştirme, veri temizleme ve veri dönüştürme alt başlıklarıyla anlatılmıştır.

Çizelge 7.1. Veri setine ilişkin tanımlayıcı istatistikler

<i>Girdiler</i>	<i>Simge</i>	<i>Birim</i>	<i>En küçük</i>	<i>En büyük</i>	<i>Ortalama</i>	<i>Standart Sapma</i>
<i>SO₂</i>	<i>SO₂</i>	µg/m ³	1	200	21,62	26,14
<i>PM₁₀</i>	<i>PM₁₀</i>	µg/m ³	4	417	85,33	62,24
<i>Sıcaklık</i>	<i>T</i>	°C	-6	36	17,92	8,61
<i>En Düşük Sıcaklık</i>	<i>MNT</i>	°C	-11	28	12,17	7,25
<i>En Yüksek Sıcaklık</i>	<i>MXT</i>	°C	-3	44	23,61	9,80
<i>Çiğ Noktası</i>	<i>DP</i>	°C	-14	20	6,50	5,06
<i>Deniz Seviyesi Basıncı</i>	<i>SP</i>	mbar	995,10	1035,40	1013,50	6,28
<i>Hava Basıncı</i>	<i>AP</i>	mbar	946,20	982,80	964,52	5,11
<i>Rüzgâr Hızı</i>	<i>WS</i>	m/s	0,05	4,99	1,16	0,51
<i>Sönümsüz En Büyük Rüzgâr Hızı</i>	<i>MXWS</i>	m/s	0,51	10,8	2,29	1,13
<i>Yağış</i>	<i>R</i>	mm	0	77,72	1,77	5,64
<i>Görünürlük</i>	<i>V</i>	km	4,35	21,40	19,35	1,61

7.1.1. Veri bütünleştirme

Farklı veri kaynaklarından alınan verilerin tek bir veri kümesinde değerlendirmeye alınabilmesi için bu verilerin ortak bir türe dönüştürülmesi yani bütünleştirilmesi söz konusu olacaktır. (Özkan, 2013). Dikkatlice yapılacak bir veri bütünleştirme, veri setindeki tutarsızlık ve uyumsuzlukların önüne geçecektir (Han vd., 2011).

Veri bütünleştirme işleminde gerekli birim dönüşümleri yapılmış, tarihsel uyumluluk gözetilmiş, ondalıklı sayı yazımındaki farklılıklar giderilmiş, nümerik değerlerde bulunan bazı metinsel (string) karakterler silinmiştir.

7.1.2. Veri temizleme

6 Haziran 2008 tarihli ve 26898 sayılı Resmi Gazetede yayımlanan “Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği’nde belirtildiği üzere, günlük verinin geçerli olabilmesi için saatlik olarak geçerliliği kabul edilen ölçümlerin % 75’inin belirlenen

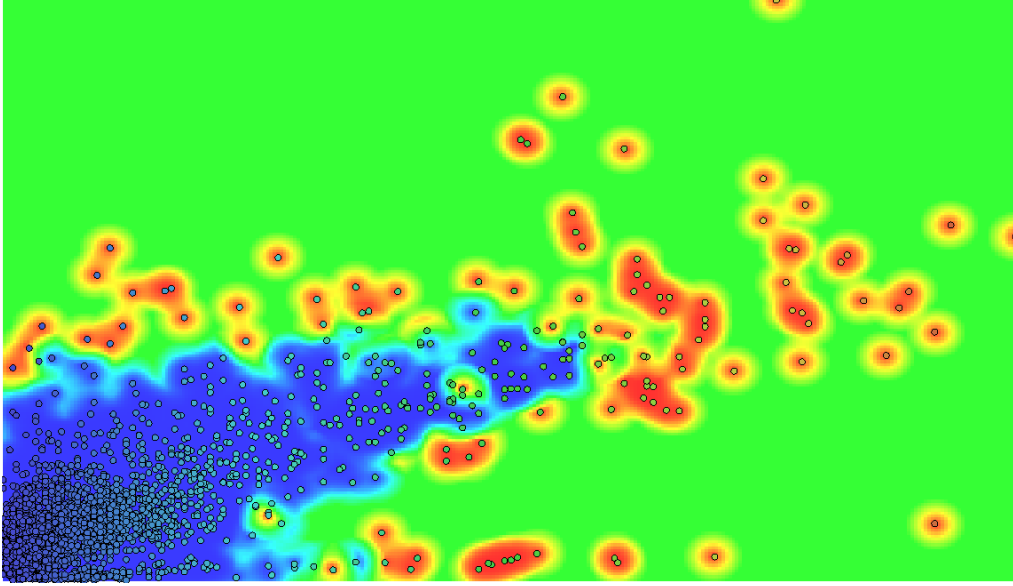
standartlara uygun olarak ölçülmüş olması gerekmektedir. Aksi halde veri ölçülmemiş sayılıp veri tabanında eksik veri olarak görülecektir. Bu durumda eksik veriler silinmeli ya da yerine yeni değerler belirlenmelidir. Eksik verinin var olduğu durumlar için Han vd. (2011) tarafından önerilen aşağıdaki yöntemler kullanılabilir:

1. Eksik değer içeren kayıtlar yok sayılarak veri setinden atılır. Bu yöntemin eksik verilerin veri kümesine oranla az sayıda olduğu veri kümelerinde kullanılması önerilmektedir.
2. Eksik değerlerin tek tek elle girilmesi kullanılabilir diğer bir yöntemdir. Ancak eksik değerlerin fazla olduğu büyük veri kümelerinde bu yöntemin uygun olmayacağı açıktır.
3. Bütün kayıp değerlerin yerine genel bir sabit kullanılabilir. Örneğin “bilinmiyor” değeriyle bu eksik veriler giderilebilir. Basit fakat güvenilir olmayan bir yöntemdir.
4. Eksik değerlerin bulunduğu özelliğin ortalama değeri veya medyanı ile eksik kayıtlar doldurulabilir. Bu yöntemde, verinin normal dağıldığı durumlarda ortalama değerin, asimetrik dağıldığı durumlarda ise medyanın kullanılması önerilmektedir.
5. Sınıf bazında ortalama değer kullanılarak eksik veri doldurulabilir. Örneğin kişinin maaşı bilinmiyor ancak mesleği biliniyor ise veri kümesinde ilgili mesleğe sahip kişilerin maaşlarının ortalaması eksik veri için kullanılabilir.
6. Bayesçi çıkarım, regresyon, karar ağaçları gibi yöntemlerle eksik değerler tahminlenebilir.

Bu çalışmada, eksik değerler k en yakın komşu (k-NN) algoritması kullanılarak tahminlenmiştir. Bu algoritmada, öncelikle kullanıcı tarafından girilen k değeri kadar en yakın gözlem Öklid uzaklığına göre belirlenmektedir. Daha sonra bu gözlem değerlerinin ortalaması alınarak eksik değerler tahminlenmektedir (Batista ve Monard, 2002).

Veri temizleme aşamasında diğer önemli bir nokta ise gürültülü verilerdir. Bu tip veriler hatalı ya da aykırı değerler içerir. Ölçüm yapan cihazın hatalı veri göndermesi veya verileri giren operatörün/elektronik sistemin aksaması sonucu olarak görülebilecek hatalı verilere, havada bulunan kirletici miktarının negatif değer alması örnek verilebilir. Aşırı uç ve aykırı değerlerin varlığının nedeni hatalı ölçümler ve/veya ender gerçekleşen olaylar olabilir. Aşırı uç ve aykırı değerler doğru sonuçlara ulaşmaya engel olabileceği için bu

değerler k-NN algoritması ile kümelendirilerek veri setinden çıkarılmıştır. Belirlenen aykırı değerler Şekil 7.1’de gösterilmiştir. Normalleştirilmiş veri seti üzerinde de k-NN algoritması ile aykırı veriler tespit edilmiş olup bu veri örneklerinin aynı deseni fakat aykırı değerlerin farklı bir desen oluşturduğu görülmüştür. Ancak normalleştirilmemiş veri seti üzerinden aykırı verilerin ayıklanması daha başarılı sonuçlar vermiştir. Ayrıca NOAA veritabanı sistemi, ölçülememiş meteorolojik veriler için sabit bir aşırı uç değer atadığından bu değerlerin bulunduğu kayıtlar ilk aşamada boş hücrelere çevrilmiştir.



Şekil 7.1. k-NN algoritması ile belirlenen aykırı değerler

7.1.3. Veri dönüştürme

Veri setinde yer alan değişkenlerin ortalama ve varyansları birbirlerinden büyük ölçülerde farklı olduklarında, büyük ortalama ve varyansa sahip değişkenler diğer değişkenleri baskılamakta ve bu değişkenlerin rollerini göreceli olarak azaltmaktadır (Özkan, 2013). Değişkenlerin çok büyük ve çok küçük değerleri çözümleme üzerinde olumsuz etkilerde bulunmaktadır. Bu gibi durumlarda verilerin dönüştürülmesi uygun olmaktadır (Özdamar, 2004).

Değişken için kullanılan farklı birimler veri analizi sonucunu farklı şekillerde etkileyebilirler (Han vd., 2011). Örneğin sıcaklık birimi fahrenheit olarak alındığında sıcaklık değerleri daha geniş bir aralığa yayılırken, °C birim olarak alındığında ise aralık

daralmaktadır. Sonuçların birimlerden bağımsız olabilmesi için verilerin normalleştirme ya da standartlaştırma işlemlerinden geçmesi gerekmektedir (Han vd., 2011).

Veri dönüştürmenin sağladığı en önemli avantajı büyük ortalama ve varyansa sahip değişkenlerin diğer değişkenleri baskılamasının önüne geçilmesidir. Diğer bir avantajı ise çekirdek fonksiyonunda yapılan sayısal hesaplamalarda sağladığı kolaylıktır (Hsu vd., 2010). Chang ve Lin (2001) normalleştirilmemiş bir veri kümesiyle destek vektör makinelerinin başarısının düşeceğini belirtirken Shanker vd. (1996) normalleştirilme metotlarının yapay sinir ağlarının başarısına olumlu katkıda bulunduğunu göstermiştir.

7.1.3.1. Min-max normalleştirilmesi

Veri içerisindeki en büyük ve en küçük değerlerin belirlenmesi ve diğer verilerin bu değerlere uygun bir biçimde [0,1] aralığındaki sayısal değerlere dönüştürülmesi işlemidir. Dönüştürme bağıntısı x_i^* normalleştirilmiş değeri, x_i gözlem değerini, x_{enk} en küçük gözlem değerini, x_{enb} en büyük gözlem değerini belirtecek şekilde Denklem (7.1)'deki gibi ifade edilmektedir.

$$x_i^* = \frac{x_i - x_{enk}}{x_{enb} - x_{enk}} \quad (7.1)$$

Ham veri setinden örnekler Ek Açıklama-D'de, normalleştirilmiş veri setinden örnekler ise Ek Açıklama-E'de verilmiştir.

7.1.3.2. Standart skor (Z-score) normalleştirilmesi

Veri setindeki her bir değişken değerinin ortalamadan farkının, standart sapmaya bölünmesidir. Böylece dönüştürülmüş verinin ortalaması 0 ve standart sapması 1 olduğundan değişkenler normal dağılıma uygun hale gelirler.

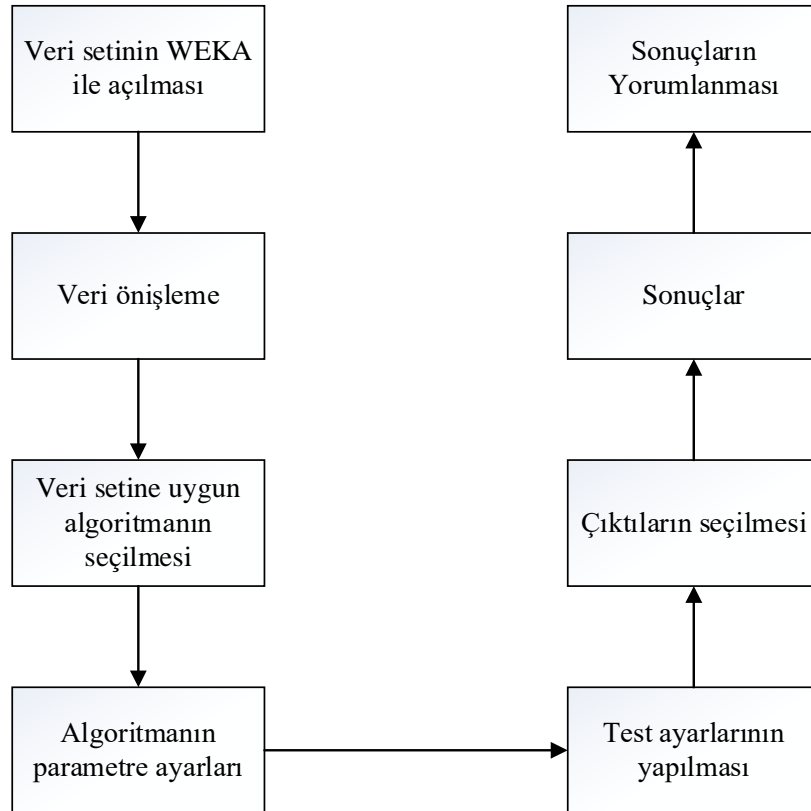
Z-score dönüştürme bağıntısı, z standart değeri, x_i gözlem değerini, \bar{x} gözlem değerlerinin aritmetik ortalamasını ve σ_x gözlem değerlerinin standart sapmasını göstermek üzere Denklem (7.2) ile ifade edilmektedir.

$$z = \frac{x_i - \bar{x}}{\sigma_x} \quad (7.2)$$

7.2. WEKA

WEKA (Waikato Environment for Knowledge Analysis) Waikato Üniversitesi tarafından 1992 yılında Java platformunda geliştirilmeye başlanan, GNU lisanslı, açık kaynak kodlu, özgür bir veri madenciliği (data mining) ve makine öğrenmesi (machine learning) yazılımıdır (Hall vd., 2009).

WEKA'nın; Java Sanal Makinesi ile herhangi bir işletim sisteminden bağımsız çalışabilmesi, birçok algoritma ve gereci içermesi, Java ile entegre olabilen kütüphanelerle birlikte kurulması, kullanıcıların yazdıkları kodların programa eklenebilmesine olanak sağlaması, basit bir kullanıcı arayüzüne sahip olması avantajları olarak sayılabilir. WEKA'da bir uygulama yapılması için izlenmesi gereken temel adımlar Şekil 7.2'de verilmiştir.



Şekil 7.2. WEKA'da izlenen adımlar

WEKA 3.7.12 sürümü itibariyle bünyesinde; 82 adet veri önışleme gereci, 133 adet sınıflama/regresyon algoritması, 12 adet kümeleme algoritması, 32 adet özellik çıkarma algoritması, 7 adet birliktelik kuralı ve 3 adet grafik arayüzü bulundurmaktadır.

WEKA çalıştırıldığında ekrana ilk olarak Şekil 7.3'teki pencere gelmektedir. "Program" sekmesi altından "Log" penceresine ve "Memory usage" (hafıza kullanımı) ayarına erişilebilmektedir. "Visualization" sekmesi adı altında veri setini görselleştirmeye yarayacak gereçler bulunmaktadır. "Tool" sekmesinde ise "Package manager" (paket yöneticisi), "Bayes net editor" ve veri okuyucular yer almaktadır.

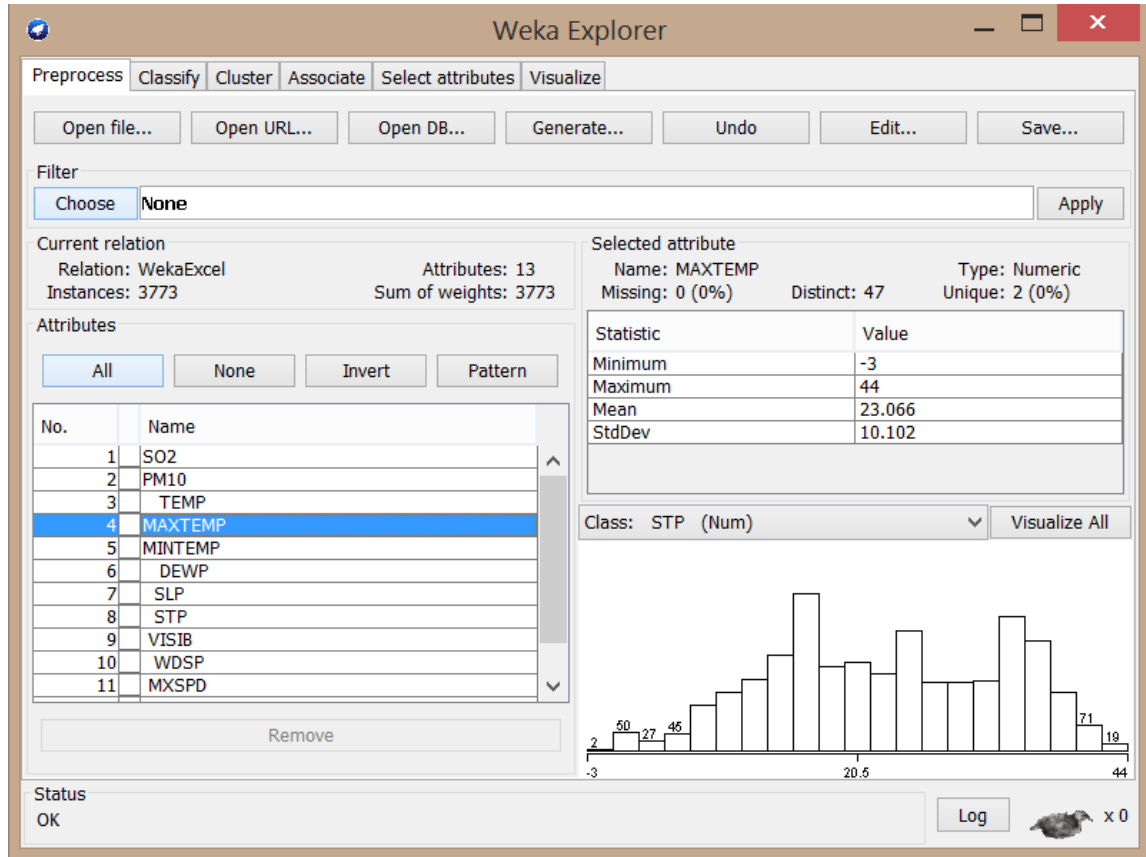


Şekil 7.3. WEKA açılış ekranı

"Applications" kısmındaki ilk seçenek ile WEKA'yı kullanmanın en temel yöntemi olan "Explorer" arayüzüne erişilir. Bu arayüz, veri kümesi üzerinde çeşitli işlemlere olanak sağlayan gereç, filtre ve algoritmalar içermektedir. "Experimenter" seçeneği ile bir ya da birden fazla veri seti için bir ya da daha fazla algoritmanın çalıştırılarak sonuçların kıyaslanabildiği bir kullanıcı arayüzüne erişilir. "KnowledgeFlow" ise yazılımda kullanıcı arayüzüne sahip son seçenek olup "Explorer" seçeneğindeki tüm gereç ve algoritmaların yer aldığı ve bunların akış diyagramı şeklinde tasarlanabildiği alandır. Simple CLI (Command-Line Interface) ise komut ekranı üzerinden işlem yapılmasına olanak sağlar. WEKA'da yer alan arayüzler izleyen alt başlıklarda daha ayrıntılı olarak açıklanmıştır.

7.2.1. Explorer

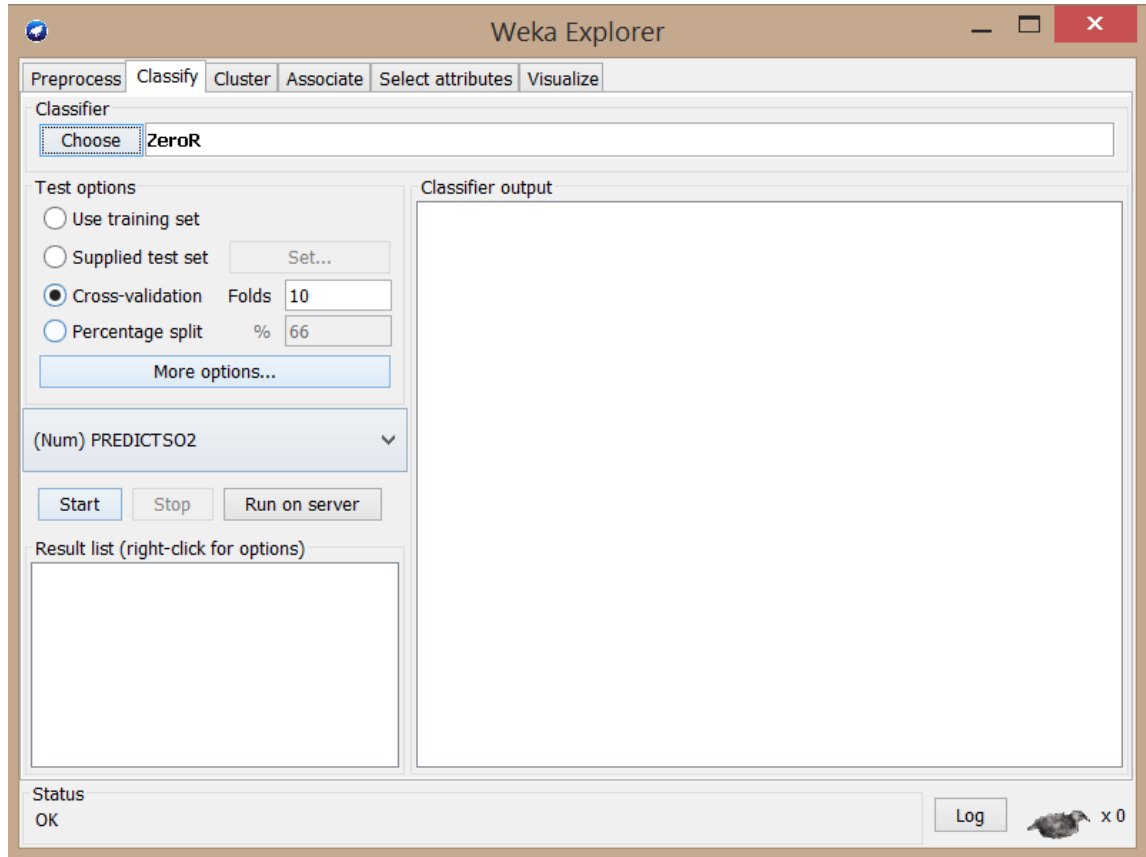
WEKA'nın temel grafik arayüzü olan "Explorer" penceresinde Şekil 7.4'te görüldüğü gibi "preprocess" (önişleme), "classify" (sınıflama), "cluster" (kümeleme), "associate" (birleştirme), "select attributes" (özellik seçimi) ve "visualize" (görselleştirme) olmak üzere altı adet sekme mevcuttur. Herhangi bir veri seti yüklenene kadar "preprocess" sekmesi dışındakiler pasif haldedir.



Şekil 7.4. Explorer arayüzü

Veri setini disk üzerinden yükleyebilmek için "Open file", internet üzerinden erişilebilen bir veri setini kullanabilmek için "Open URL", yerel ağdaki bir veri tabanı sunucusundan veri almak için "Open DB" ve son olarak gerekli parametre ayarlarını yapılmasıyla yapay bir veri kümesi üzerinde çalışabilmek için "Generate" düğmesi kullanılmaktadır. WEKA'nın öntanımlı olarak desteklediği veri uzantıları ARFF, CSV, C4.5, JSON, LibSVM ve XRFF'dir. Bunlara ek olarak paket yöneticisiyle yüklenebilecek eklentilerle Excel ve Matlab dosyaları da WEKA ile açılabilir.

“Preprocess” ekranında veri setine ilişkin basit istatistiksel bilgiler ve grafikler görüntülenebilmektedir. Bu sekmenin yüklendiği asıl görev ise, veri değerlerinin dönüşümü, eksik verilerin doldurulması, aşırı uç değerlerin ayıklanması, normalleştirme, örnekleme gibi işlemlerin yapılmasına olanak sağlamasıdır. Verinin bu filtrelerden geçirilmesi için öncelikle “Choose” düğmesine tıklanarak istenilen filtrenin seçilmesi gerekir. Seçilen filtre hakkında bilgi almak ve gerekli ayarları yapmak içinse “Choose” düğmesinin yanında yer alan filtre ismine tıklanır.



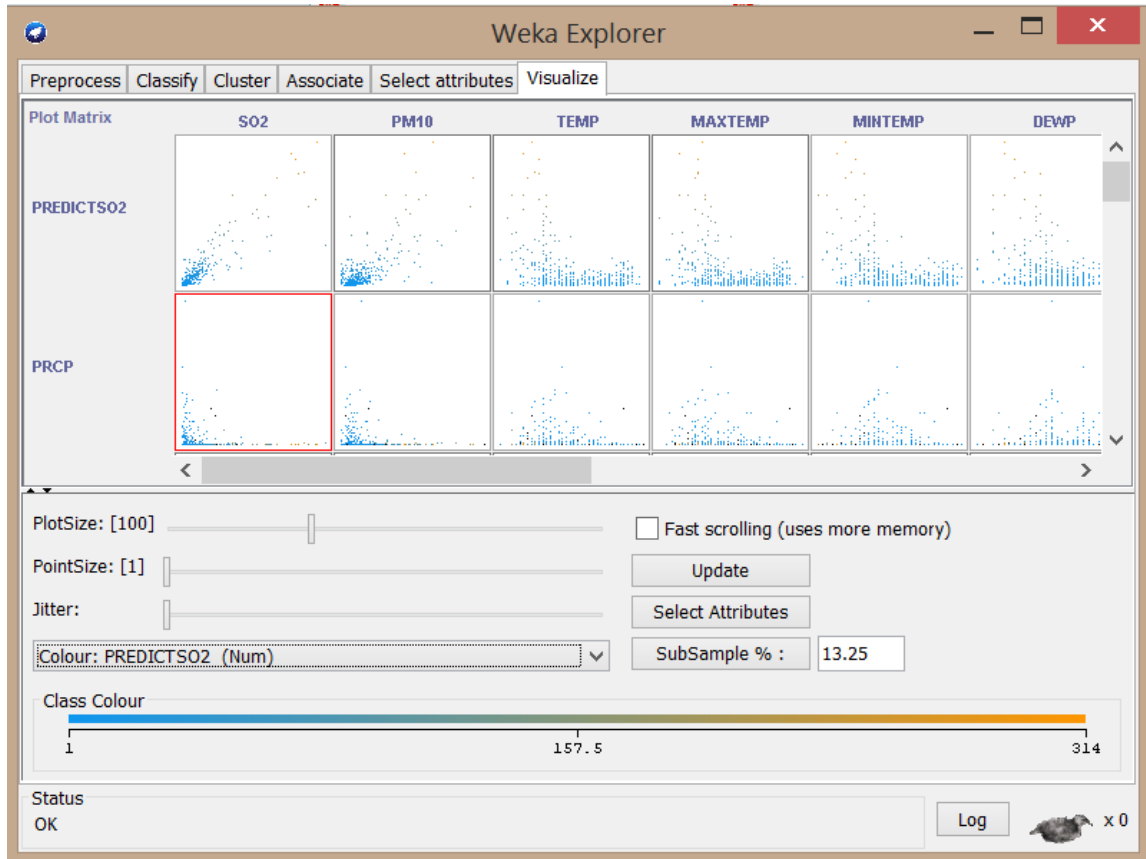
Şekil 7.5. Explorer penceresinde classify sekmesi

“Classify” sekmesinde çeşitli sınıflandırma algoritmalarının kullanılması ve parametrelerinin ayarlanmasıyla ilgili seçenekleri gösteren bir kullanıcı ara yüzü ekrana gelir. Şekil 7.5’teki ekranda “Classifier” başlığı altında bulunan “Choose” düğmesi aracılığıyla istenilen sınıflandırma algoritması seçilir. Yine “Choose” düğmesinin sağında yer alan algoritma ismine tıklanarak gerekli parametre ayarları yapılabilir ve algoritma hakkında detaylı bilgi alınabilir. “Test option” başlıklı option-box ise test verisiyle ilgili ayarların yapıldığı kısımdır. “More options” düğmesine tıklanarak sonuç ekranı için gerekli

ayarlar yapılırken aşağısındaki drop-down list-boxla istenilen hedef sınıf seçilebilir. Son olarak “start” düğmesi aracılığıyla algoritma çalıştırılır.

“Cluster”, “Associate” ve “Select Attributes” sekmelerinde de benzer işlemler yapılmaktadır.

“Visualize” sekmesinde ise Şekil 7.6’da gösterildiği gibi veri setindeki özneliklerin birbiriyle ilişkilerini gösteren iki boyutlu grafikler bulunmaktadır. Çok büyük boyutlu veri seti ile çalışılırken, görsel karmaşıklığı önlemesi açısından istenilen boyuttaki alt örneklem uzayının kullanılması da mümkündür.

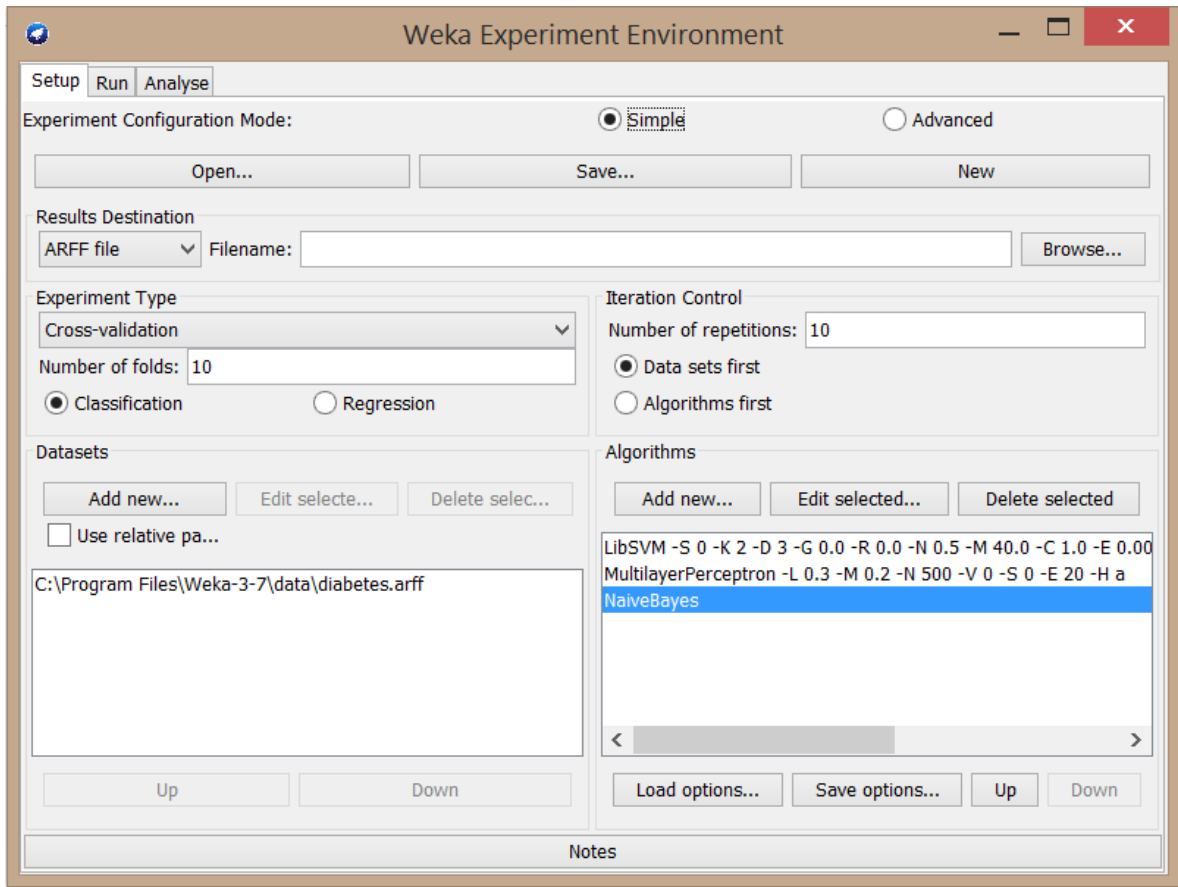


Şekil 7.6. Explorer penceresinde visualize sekmesi

7.2.2. Experimenter

“Experimenter” istenilen veri kümesi veya veri kümelerinin sınıflama veya regresyon çözümlemesinde birkaç algoritmanın ya da bir algoritmanın farklı

parametrelerinin istenilen çıktılar üzerinden kıyaslanabildiği bir kullanıcı arayüzüdür. Şekil 7.7’de gösterilen bu arayüzde yapılan deneyler disk üzerine kaydedilebilmekte böylece bu kayıtlar daha sonra experimenter veya komut satırı üzerinden tekrar çalıştırılabilmektedir. Öncelikle “Setup” sekmesinde istenilen veri/veriler ve algoritma/algoritmalar seçilerek uygun ayarlar yapılır. “Run” sekmesinde deney çalıştırıldıktan sonra “Analyse” sekmesinde sonuçlar arasında kıyaslama yapılabilmektedir. Experimenter arayüzünün önemli bir avantajı da işlem yükünü birden fazla makine üzerine dağıtabilme yeteneğinin olmasıdır (Witten vd, 2011).

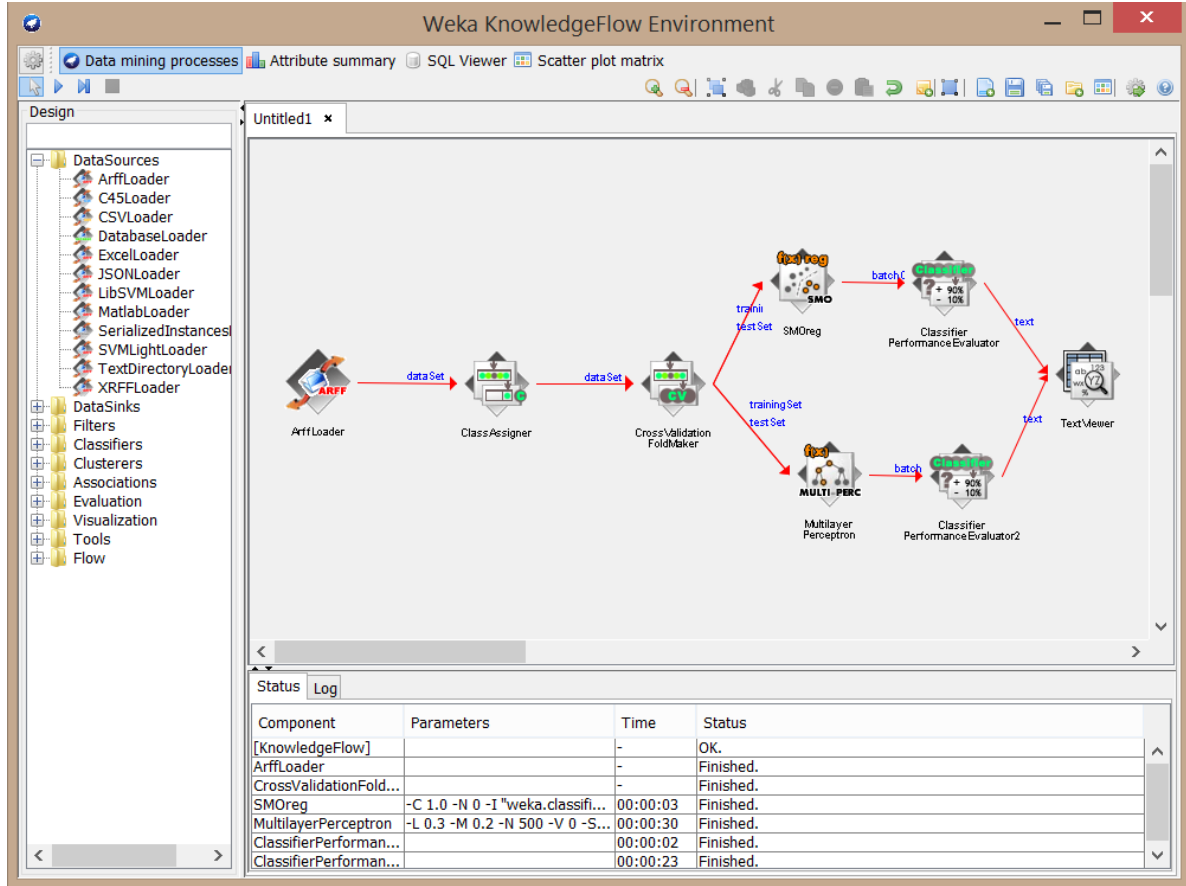


Şekil 7.7. Experimenter arayüzü

7.2.3. Knowledge flow

Tüm özellikleri tek bir ekran üzerinde bulunduğu ve bu işlemlerin akış diyagramı şeklinde tasarlanarak çalıştırıldığı, üçüncü ve son WEKA kullanıcı arayüzüdür. Şekil 7.8’de verilen “Knowledge Flow” arayüzünde yapılacak her bir işlem (veri yükleme, hedef sınıfı seçme, veriyi test ve eğitim seti olarak bölümlenme, sonuçların yazdırılması ve okunması vb.)

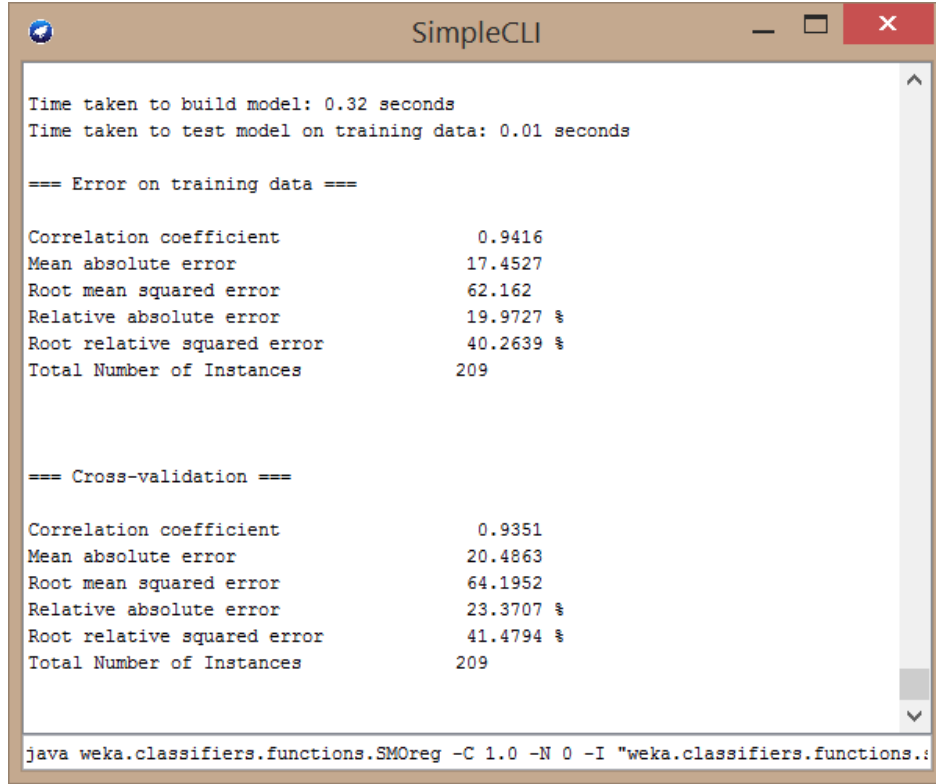
akış diyagramının bir elemanını oluşturmaktadır. Knowledge Flow arayüzünün en önemli avantajlarından ilki birden fazla algoritma veya veri üzerinde aynı anda çalışılabilmesidir. İkinci bir avantajı ise Explorer arayüzünün aksine tüm verinin aynı anda hafızaya alınmaması, verinin kademeli işlenmesiyle büyük boyutlu veriler üzerinde de çalışılabilmesidir (Witten vd, 2011).



Şekil 7.8. Knowledge Flow arayüzü

7.2.4. Simple CLI

WEKA bileşenlerini komut satırı üzerinden kullanmaya olanak sağlar. “Simple CLI” açıldığında WEKA kütüphaneleri yüklenmiş halde gelir. Analizleri komutsal olarak yapmaya olanak sağlayan kısımdır. İlgili uygulama Şekil 7.9’da gösterilmiştir.



```

SimpleCLI
Time taken to build model: 0.32 seconds
Time taken to test model on training data: 0.01 seconds

=== Error on training data ===

Correlation coefficient          0.9416
Mean absolute error             17.4527
Root mean squared error         62.162
Relative absolute error         19.9727 %
Root relative squared error     40.2639 %
Total Number of Instances      209

=== Cross-validation ===

Correlation coefficient          0.9351
Mean absolute error             20.4863
Root mean squared error         64.1952
Relative absolute error         23.3707 %
Root relative squared error     41.4794 %
Total Number of Instances      209

java weka.classifiers.functions.SMOreg -C 1.0 -N 0 -I "weka.classifiers.functions.

```

Şekil 7.9. Simple CLI komut satırı

Şekil 7.9'da aşağıdaki komut satırı örneklenmiştir.

```

java weka.classifiers.functions.SMOreg -C 1.0 -N 0 -I
"weka.classifiers.functions.supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K
"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -t data/cpu.with.vendor.arff

```

7.3. Model ve Değerlendirme

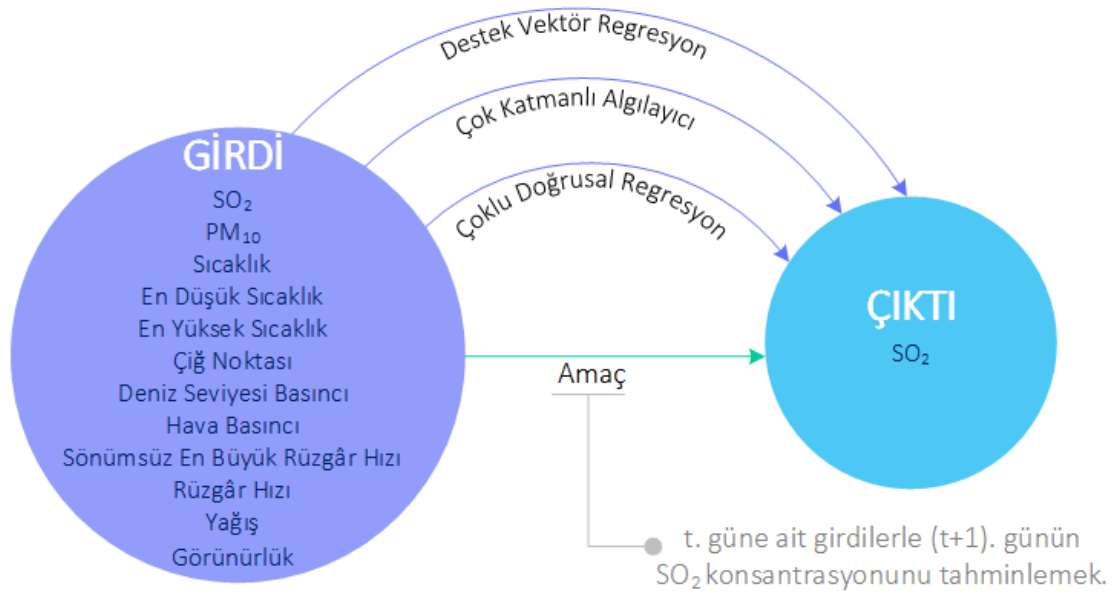
Bu bölümde genel tahminleme modeli ile ilgili bilgiler, eğitim ve test veri kümeleri oluşturulurken kullanılan yöntemler, tahminleme çalışması için kullanılan yöntemlere ilişkin parametrelerin belirlenmesi işlemi yer almaktadır.

7.3.1. Genel tahminleme modeli

t . güne ait hava kirleticileri konsantrasyonu ve meteorolojik verilerin kullanılmasıyla $(t + 1)$. günün hava kirliliğini SO_2 üzerinden tahminlemeyi amaçlayan bu çalışmanın genel tahminleme modeli aşağıdaki gibi ifade edilebilir.

$$SO_{2(t+1)} = f(SO_{2(t)}, PM_{10(t)}, T_t, MNT_t, MXT_t, DP_t, SP_t, AP_t, MXWS_t, WS_t, R_t, V_t)$$

Modele ait girdiler, çıktı, modelin amacı ve bu amaca ulaşmak için girdileri çıktıya dönüştüren yöntemler. Şekil 7.10'da daha açık bir biçimde ifade edilmiştir.



Şekil 7.10. Genel tahminleme modeli

7.3.2. Eğitim ve test verisinin belirlenmesi

Eğitim ve test verisinin belirlenmesi, veri madenciliği uygulamalarının en önemli adımlarından biridir. Eğitim verisi, makinenin ilişkiyi çıkarımlar yapabilmesi için kullanılırken test verisi, makine öğrenmesini sınamak için kullanılmaktadır. Bu iki veri setinin birbirinden bağımsız ve farklı olması modelin güvenilirliği açısından oldukça önemlidir (Witten vd., 2011).

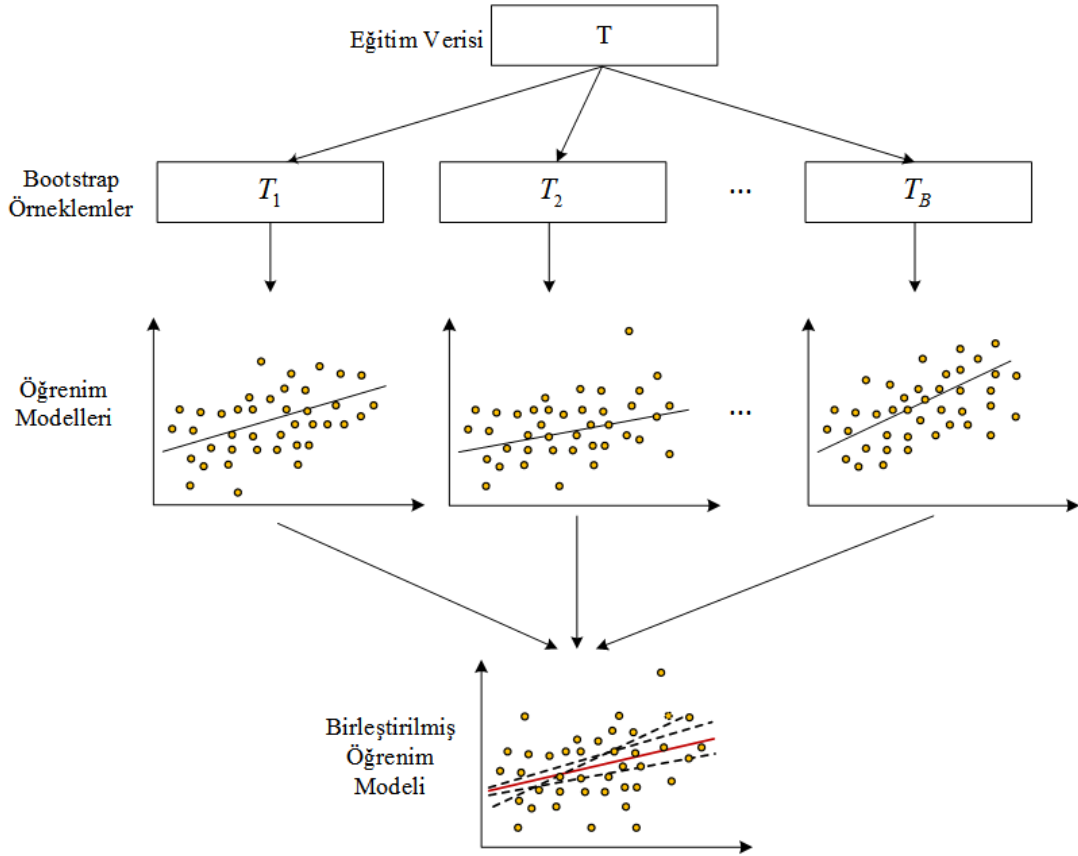
Eğitim ve test verisinin oluşturulması makine öğrenmesi yöntemleriyle elde edilen sonuçların geçerliliği için önemli bir adımdır. Bu adımın temel amacı, aşırı uyumdan uzak ve iyi bir genelleştirme yeteneğine sahip bir model elde etmektir. Genelleştirme yeteneği düşük olan bir model, mevcut veriyle (eğitim verisi) oldukça iyi çalışmasına karşın daha önce karşılaşmadığı verilerle (test verisi) sınıandığında kötü bir performans gösterebilmektedir. Özetle eğitim aşamasındaki düşük hata performansı, test aşamasındaki hatanın da düşük olacağını göstergesi değildir. Bu durumun üstesinden gelmek için

kullanılan çeşitli yöntemler bulunmaktadır. Bu çalışmada modelin genelleştirme yeteneğini arttırmak amacıyla bagging ve çapraz geçirme yöntemleri kullanılmıştır.

7.3.2.1. Bagging (bootstrap aggregating)

Bagging (torbalama), sınıflandırma ve regresyon modellerinin performansını arttırmak amacıyla Breiman (1994) tarafından önerilen birleştirilmeli öğrenim yöntemidir. Yönteminin temel amacı varyansı düşürmek ve aşırı uyumun oluşumunu önlemektir.

Torbalama yönteminde, n boyutundaki orijinal eğitim verisinin, özyükleme (bootstrap) tekniği ile yeniden örneklenmesiyle birbirinden bağımsız B adet n boyutunda yeni eğitim verisinin oluşturulması ilk basamaktır. Bu yöntemde orijinal eğitim verisindeki herhangi bir örneğin, türetilen eğitim setinde birden fazla kez kullanılması olasıdır. Şekil 7.11’de görüldüğü gibi türetilen her bir özyükleme örnekleme için öğrenim modeli kurulur ve bu modellerin birleştirilmesiyle tek bir model elde edilir.



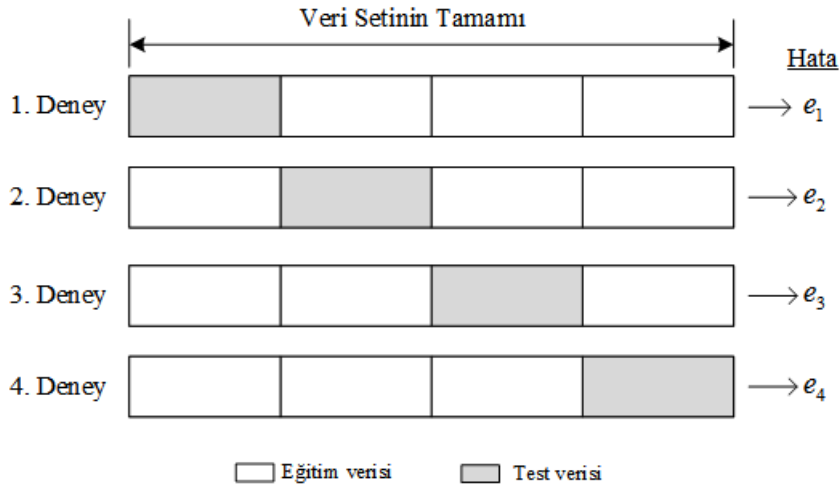
Şekil 7.11. Bagging (torbalama)

7.3.2.2. k-kat çapraz geçерleme

k-kat çapraz geçерleme, aşırı uyum ile yanlılıđı düşürmeyi ve modelin sonuçlarını geçерlemeyi sađlayan bir yöntemdir. k-kat çapraz geçерleme yönteminde öncelikle veri seti eşit büyüklükte ve rassal olarak belirlenen k parçaya ayrılır $D \rightarrow \{D_1, D_2, \dots, D_k\}$. k kez eğitilip test edilecek modelde her bir tekrar için farklı bir parça, test verisi (D_i) olarak ayrılırken geri kalan $k - 1$ parça ($D - D_i$), modeli eğitmek için kullanılır. Her bir tekrar için hesaplanan hataların ortalaması alınarak modelin hatası Denklem (7.3) ile ölçülür (Hastie vd., 2009).

$$e_{cv} = \frac{1}{k} \sum_{i=1}^k e_i \quad (7.3)$$

k parametresi için kabul edilen optimum bir deđer olmamasına karşın literatürde çođunlukla 10-kat çapraz geçерleme kullanılmaktadır. Şekil 7.12'de $k=4$ deđer için çapraz geçерleme yöntemi gösterilmiştir.



Şekil 7.12. 4-kat çapraz geçерleme

7.3.3. Performans göstergeleri

Bir tahminleme modelinin öncelikli amacı, hedef deđerlere göre en isabetli tahminleri yapmaktır. Bu noktada modelin başarısını ölçmek için tahminleme hatası kullanılmaktadır. i . gözlem deđerleri y_i ve bu gözlem deđerleri için yapılan tahmin \hat{y}_i olmak

üzere, bu gözleme ait tahminleme hatası basitçe $e_i = y_i - \hat{y}_i$ şeklinde ifade edilebilir. Tahminlerin doğruluğunu sınamak için e_i hatası üzerine kurulmuş olan ve literatürde sıklıkla kullanılan iki tip performans göstergesi bulunmaktadır (Hyndman ve Athanasopoulos, 2014).

MAE (ortalama mutlak hata), Denklem (7.4)'te gösterildiği gibi, gözlem değerleri ile modelin tahminleri arasındaki mutlak farkın ortalamasını ifade etmektedir. Böylece hesaplanan ortalama hata değeri, i . hatanın yönünden (işaretinden) etkilenmemektedir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| = |\mu_{e_i}| \quad (7.4)$$

RMSE (karesel hata ortalamalarının karekökü), gözlem değerleri ile modelin tahminleri arasındaki farkın karesel ortalamasının karekökünü ifade etmektedir. Böylece $(-1, 1)$ aralığındaki hataların ortalamaya olan etkisi azalırken daha büyük hata değerlerinin ortalamaya olan etkisi artacaktır. Bu durumda RMSE büyük hata değerlerinin istenmediği durumlar için daha kullanışlı olacaktır. RMSE hesabı, Denklem (7.5)'te verilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\mu_{e_i^2}} \quad (7.5)$$

Korelasyon katsayısı (r) ise modelin tahminlerinin doğruluğunu sınamak için kullanılan bir gösterge olmayıp gözlemler ve tahminler arasındaki doğrusal ilişkinin derecesini ve yönünü ifade etmektedir (Murphy, 1995). $[-1, 1]$ aralığında değer alan korelasyon katsayısının, sınır değerlere yaklaşması doğrusal ilişkinin güçlendiğini göstermektedir. Sıfıra yaklaşan korelasyon katsayısı ise doğrusal ilişkinin zayıfladığını ifade etmektedir. Pozitif değerler; gözlemler ve tahminler arasındaki ilişkinin doğru orantılı, negatif değerler ise bu ilişkinin ters orantılı olduğunu göstermektedir.

$S_{y_i \hat{y}_i} = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{n-1}$, $S_{y_i} = \frac{\sum (y_i - \bar{y})^2}{n-1}$, $S_{\hat{y}_i} = \frac{\sum (\hat{y}_i - \bar{\hat{y}})^2}{n-1}$ olmak üzere korelasyon katsayısı Denklem (7.6)'daki gibi ifade edilir (Witten vd., 2011).

$$r = \frac{S_{y_i \hat{y}_i}}{\sqrt{S_{y_i} S_{\hat{y}_i}}} \quad (7.6)$$

Evans (1966) korelasyon katsayısının mutlak değerinin yorumlanması ile ilgili Çizelge 7.2'deki tanımlamayı önermiştir (Beldjazia ve Alatou, 2016)

Çizelge 7.2. Korelasyon katsayısının yorumu

Korelasyon katsayısı	İlişki
0,00 – 0,19	Çok zayıf
0,20 – 0,30	Zayıf
0,40 – 0,59	Orta
0,60 – 0,79	Güçlü
0,80 – 1,00	Çok güçlü

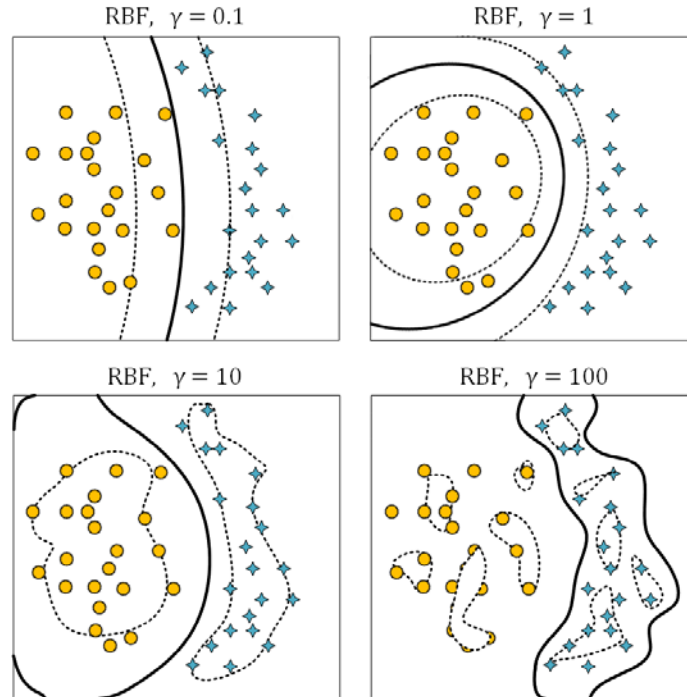
7.3.4. Parametre seçimi

Çoğu makine öğrenmesi algoritmasında kullanıcı tarafından girilmesi gereken parametreler vardır. Model başarımında kilit noktalardan biri de uygun parametrelerin belirlenmesidir. Ancak çoğu zaman parametre eniyilemesi, çok büyük -bazı durumlarda sonsuz- boyutlu çözüm kümesi içinde yapılan ve en iyi çözümü garanti edemeyen yüksek maliyetli bir işlemdir (Camilleri vd., 2014). Parametre eniyilemesi için kullanılan yöntemlerden bazıları; Bayesçi eniyileme, Nelder–Mead metodu, gradyan tabanlı eniyileme, grid search (ızgara araması) ve rassal arama olarak sayılabilir. Bu çalışmada, Bergstra ve Bengio'ya (2012) göre literatürde sıklıkla kullanılan ızgara araması yöntemi ile uygun parametreler araştırılmıştır.

İzgara araması, belirli bir aralık içinde üssel veya adımsal olarak büyüyen parametre kümelerinin kartezyen çarpımıyla oluşturulan ikilileri kullanarak çözüm uzayında arama yapar. Arama aralığı geniş seçilir ve büyüme küçük adımlarla gerçekleşirse ızgara araması metodunun denemesi gereken ikili sayısı, dolayısıyla işlemin maliyeti artacaktır. Dar bir aralığın ve büyük adımlarla büyümenin seçilmesi durumunda ise, uygun çözümden uzaklaşma olasılığı artacaktır. Bu yöntem ile en uygun parametre değerleri çapraz geçirme ile ölçülen hataya göre belirlenmektedir (Hsu vd., 2003). Örnek olarak çok katmanlı algılayıcı ve radyal tabanlı (RBF) çekirdek kullanılan destek vektör makinesine, parametrelerin etkisi ve bu parametrelerin ızgara araması ile eniyilenmesi incelenecektir.

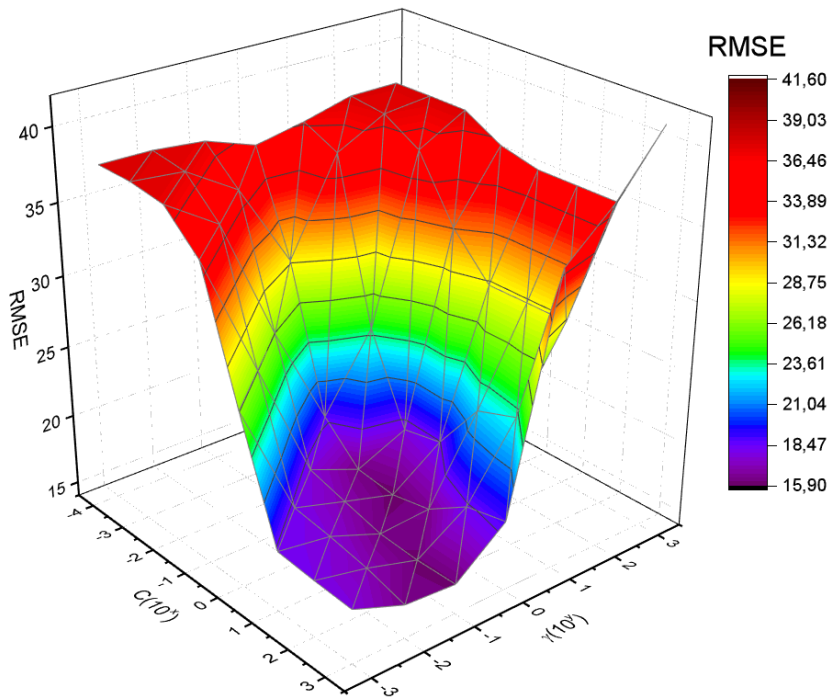
7.3.4.1. Destek vektör makinesi için parametre seçimi

RBF çekirdek kullanılan destek vektör makinelerinde model başarımını etkileyen iki parametre söz konusudur. Bunlar C ceza parametresi ve RBF çekirdek fonksiyonunda bulunan γ 'dır. C değerinin küçük seçilmesi marjini genişletmektedir. Bu durum eksik uyuma neden olarak eğitim hatasını arttırmaktadır. Büyük C değeri ise marjinin daralmasına dolayısıyla aşırı uyum sorununa neden olmaktadır. γ parametresi bir eğitim örneğinin etkisinin ulaşacağı uzaklığı tanımlamaktadır. Küçük γ değerleri için bu uzaklık artacak büyük γ değerleri için ise azalacaktır. Yani büyük γ değerleriyle, model gürültülü verilere karşı daha hassas olacak ve bu durum beraberinde aşırı uyum sorununu getirecektir. Tersisi durumda ise eksik uyum oluşacak, aynı zamanda çekirdek fonksiyonu doğrusal forma yaklaşarak doğrusal olmama özelliğini yitirecektir. γ parametresinin model üzerindeki bu etkisi Şekil 7.13'te gösterilmiştir (Ben-Hur ve Weston, 2010). Şekilde görüldüğü gibi γ değeri büyüdükçe model eğitim verisindeki her bir gözleme uyum sağlamaya çalışarak test verisinde de aynı başarıyı gösterme şansını azaltmıştır. γ değerinin küçülmesi ise ayırıcı hiper düzlemi doğrusallaştırarak modelin doğrusal olmayan veri setindeki başarı şansını azaltmıştır.



Şekil 7.13. γ parametresinin modele olan etkisi

Bu çalışmada, C ve γ parametreleri ızgara araması yöntemine göre $C \in (10^{-4}, 10^{-3}, \dots, 10^3)$, $\gamma \in (10^{-4}, 10^{-3}, \dots, 10^3)$ olmak üzere, her bir (C, γ) ikilisi kullanılarak araştırılmıştır. (C, γ) ikililerinin modele olan etkisi RMSE üzerinden Şekil 7.14'te gösterilmiştir. C parametresi sabit tutulup γ parametresinin değişimi incelendiğinde, γ ($10^{-1}, 10^{-0}$) aralığı civarından sonra küçüldükçe (doğrusallaşma) ve büyüdükçe (aşırı uyum) hata artmaktadır. γ parametresi sabit tutulup C parametresinin değişimi incelendiğinde, genel olarak C parametresi 10^2 değerinden itibaren küçüldükçe (eksik uyum) ve büyüdükçe (aşırı uyum) hata artmaktadır.



Şekil 7.14. (C, γ) parametre ikilisinin test hatasına etkisi

Son olarak ızgara araması ile ilk aşamada belirlenen parametrelerin yakınında daha küçük bir aralıkta ikinci bir arama yapılarak daha iyi parametre değerlerine ulaşılmaya çalışılmıştır. Destek vektör makinesinde beş farklı çekirdek fonksiyonuna göre parametre değerleri araştırılmış ve belirlenen en uygun değerler Çizelge 7.3'te verilmiştir.

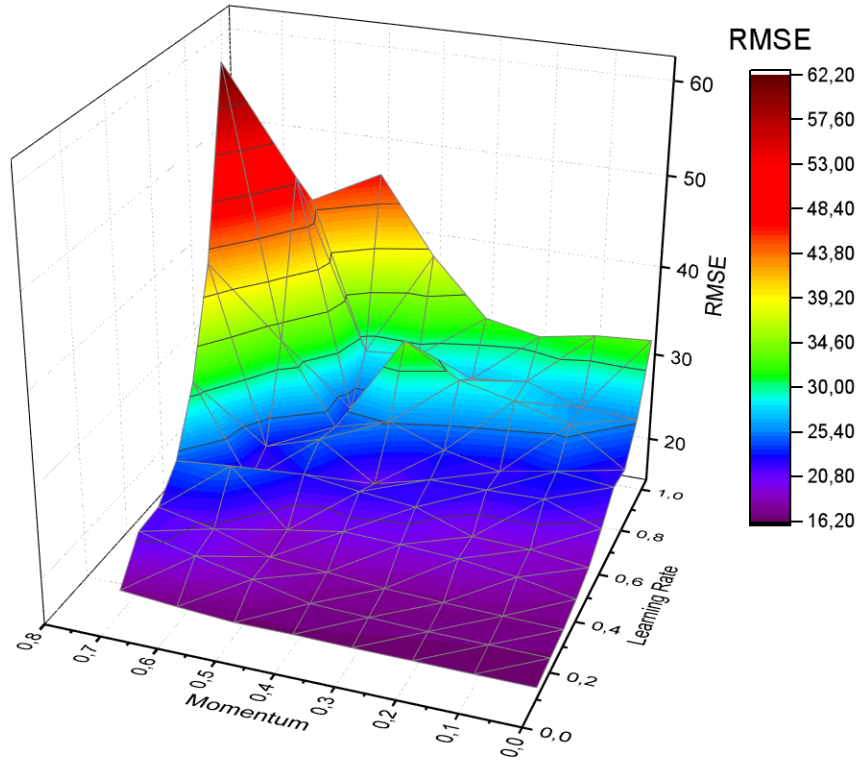
Çizelge 7.3. SVR için belirlenen parametre değerleri

	C	d	γ	ω	σ
<i>Doğrusal</i>	85	-	-	-	-
<i>Polinom</i>	1	2	-	-	-
<i>Sigmoid</i>	1250	-	0,001	-	-
<i>RBF</i>	2,5	-	1	-	-
<i>Pearson VII</i>	0,1	-	-	10	1

7.3.4.2. Çok katmanlı algılayıcı için parametre seçimi

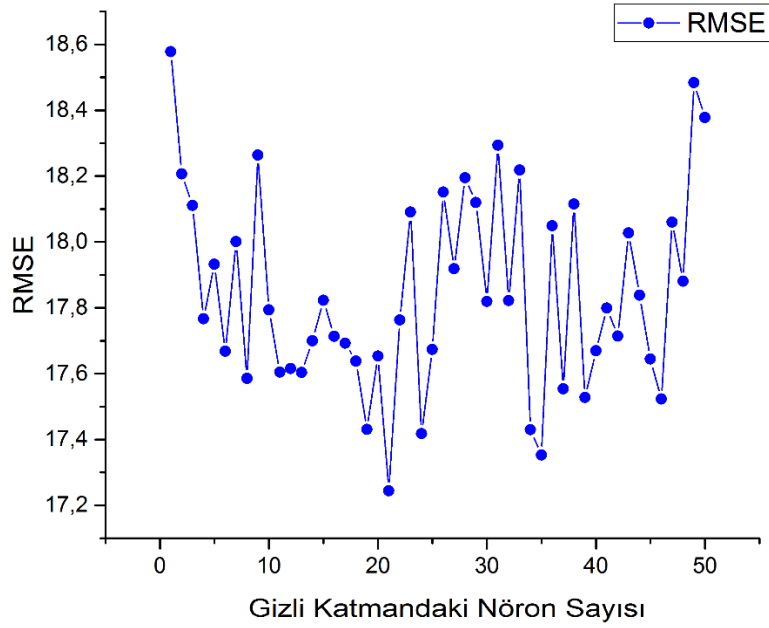
Momentum ve öğrenme katsayısı, gizli katman ve gizli katmandaki nöron sayısı, durdurma kriteri yapay sinir ağlarının performansı üzerinde oldukça etkili olan parametreleridir. Bu parametrelerin belirlenmesiyle ilgili standartlaşmış bir yöntem olmasa da parametrelerin ağırlık öğrenme performansına olan etkileri daha önce yapılmış olan çalışmalarla tecrübe edilmiştir.

Öğrenme katsayısı ağırlıkların değişim miktarıyla ilintilidir. Küçük değerli öğrenme katsayısı, ağırlıkların değişim miktarını azaltacağından çözüme ulaşma süresi uzayacaktır. Diğer taraftan büyük değerli öğrenme katsayısı, ağırlıkların değişim miktarını artırırken salınımlara neden olabilmektedir. Momentum katsayısı, bir önceki iterasyonda yapılan değişim miktarının belirli bir oranda yeni değişime eklenmesini sağlamaktadır. Bu değer küçük olması yerel minimuma takılmaya, büyük olması ise bütünsel minimumu kaçırmaya neden olabilmektedir. Öğrenme katsayısı ve momentum katsayısının modele olan etkisi ızgara araması yöntemi ile araştırılıp en uygun değerler belirlenmeye çalışılmıştır. Öğrenme katsayısı (0,1] ve momentum terimi [0, 1] aralığında, her bir adım için değişim miktarı 0,1 olacak şekilde güncellemiştir. Şekil 7.15'te öğrenme katsayısı ve momentumun modele olan etkisi RMSE üzerinden gösterilmiştir. En düşük hata değerinde momentum katsayısının 0'a oldukça yakın bir değer alması sinir ağının neredeyse standart geriyayılım algoritması ile çalıştığını göstermektedir. Öğrenme katsayısının küçük değerler almasıyla hatanın düştüğü görülmektedir. Bu durum; küçük adımlarla ağırlık güncellenen bu problem için daha iyi sonuçlar verdiğini, aksi durumda algoritmanın bütünsel eniyiyi kaçırdığını göstermektedir.



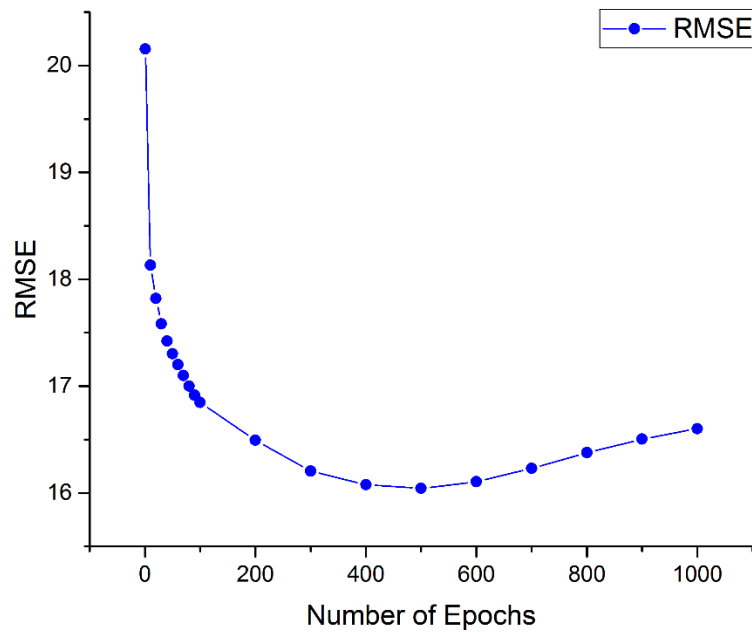
Şekil 7.15. Öğrenme katsayısı ve momentum teriminin test hatasına etkisi

Tek gizli katmanlı ağ yapısının yeterli sayıda nöron içermesi halinde birçok problem için yeterli olduğu bilinmektedir (Cybenko, 1989; Hornik vd., 1989). Bu nedenle bu çalışmada gizli katman sayısı bir olarak seçilmiştir. Gizli katmanın az sayıda nöron içermesi öğrenme sürecini güçleştirebilirken nöron sayısının artırılması hesaplama maliyetini yükseltmektedir. Ayrıca nöron sayısının artışı, bir noktadan sonra sinir ağının eğitim verisine aşırı uyum göstermesine neden olmaktadır. Böylece sinir ağının genelleştirme yeteneği azalacak dolayısıyla test performansında düşüş gerçekleşecektir (Mehrotra vd., 1996). Bu çalışmada gizli katmandaki nöron sayısı [1-50] aralığında her bir adım için değişim miktarı 1 olmak üzere güncellenmiş ve nöron sayısının modele olan etkisi RMSE üzerinden Şekil 7.16'da gösterilmiştir. Şekil incelendiğinde az sayıdaki nöronun eksim uyuma, çok sayıdaki nöronun ise aşırı uyuma neden olmasıyla bu değerlerde hatanın yüksek olduğu görülmektedir.



Şekil 7.16. Gizli katmandaki nöron sayısının test hatasına etkisi

Geri yayılım algoritmasının durdurulması çeşitli kriterlerle sağlanabilir. Bunlardan ilki maksimum yineleme sayısını (number of epoch) belirlemektir. Diğer bir yöntem, çok küçük bir ε hata değerinin belirlenmesiyle iterasyonun, hata $\leq \varepsilon$ oluncaya dek devam etmesidir. Diğer bir yöntem ise, hatadaki değişim oranının yeteri kadar küçülmesidir. Çünkü minimuma yaklaştıkça hatanın değişim oranı da küçülecektir (Gurney, 1997). Belirli aralıklarla değiştirilen yineleme sayısının modele olan etkisi Şekil 7.17’de gösterilmiştir.



Şekil 7.17. Eğitim devir sayısının test hatasına etkisi

Çok katmanlı algılayıcı için belirlenen parametre değerleri Çizelge 7.4'te ve bu değerlerle oluşturulan sinir ağının mimarisi ise Ek Açıklama-F'de gösterilmiştir.

Çizelge 7.4. MLP için belirlenen parametre değerleri

<i>Öğrenme Katsayısı</i>	<i>Momentum Katsayısı</i>	<i>Nöron Sayısı</i>	<i>Yineleme Sayısı</i>
0,1	0,0001	21	500

8. BULGULAR VE TARTIŞMA

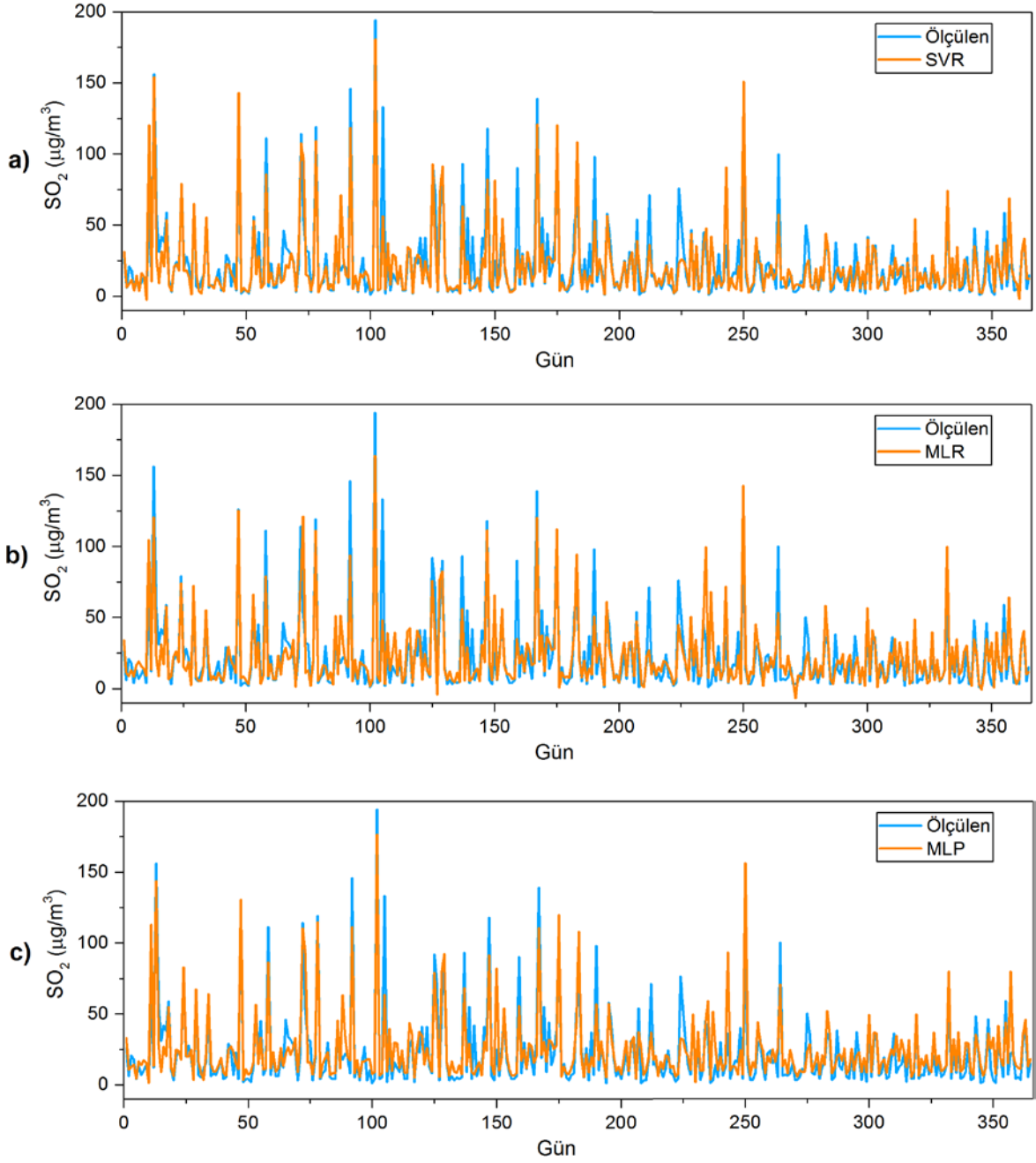
Bu başlıkta, çalışmanın amacına yönelik sonuçlar elde etmek için kullanılan yöntemlerin ürettiği çıktılar yer almaktadır. Çalışmada eksik verilerin tamamlanması (imputation) ve aykırı verilerin belirlenmesinde RapidMiner Studio, tahminleme modellerinin kurulmasında ise WEKA yazılımı kullanılmıştır.

Bu çalışmada, bir gün sonrasına ait hava kirliliğinin tahminlenmesinde doğrusal, polinom, sigmoid, Pearson VII ve RBF çekirdek fonksiyonları ile çalıştırılan destek vektör regresyon yönteminin başarısı araştırılmıştır. Kıyaslama amacıyla kullanılan diğer yöntemler ise geriyayılım algoritması kullanan çok katmanlı algılayıcı ve çoklu doğrusal regresyondur. MLP, MLR ve farklı çekirdek fonksiyonları kullanılan SVR ile elde edilen en iyi 10-kat çapraz geçiş sonuçları Çizelge 8.1’de verilmiştir.

Çizelge 8.1. Test sonuçları

	<i>R</i>	<i>MAE</i>	<i>RMSE</i>
<i>SVR-Doğrusal</i>	0,8577	8,0321	14,4207
<i>SVR-Polinom</i>	0,8749	7,6516	13,5961
<i>SVR-Sigmoid</i>	0,8574	8,0334	14,4221
<i>SVR-Pearson VII</i>	0,8773	7,5626	13,4843
<i>SVR-RBF</i>	0,8818	7,4355	13,2261
<i>MLP</i>	0,8738	8,1603	13,5789
<i>MLR</i>	0,8630	8,5795	14,0891

Öncelikle SVR’de kullanılan çekirdek fonksiyonları birbirleriyle kıyaslanacak olursa; en iyi performansı radyal tabanlı çekirdek, en kötü performansı ise birbirlerine oldukça yakın sonuçlar üreten doğrusal ve sigmoid çekirdek fonksiyonları göstermiştir. Buna karşın bütün çekirdek fonksiyonları tatmin edici sonuçlar üretmiştir. Kıyaslama amacıyla kullanılan yöntemler (MLP, MLR) kabul edilebilir sonuçlar üretmesine rağmen SVR kadar başarılı olamamıştır. Tahminleme çalışmasında başarı sırası SVR, MLP ve MLR şeklindedir. Test sonuçlarına göre SVR; MLP’ye oranla %8,88 MLR’ye oranla ise %13,33 daha az ortalama mutlak hatayla tahminler gerçekleştirmiştir.

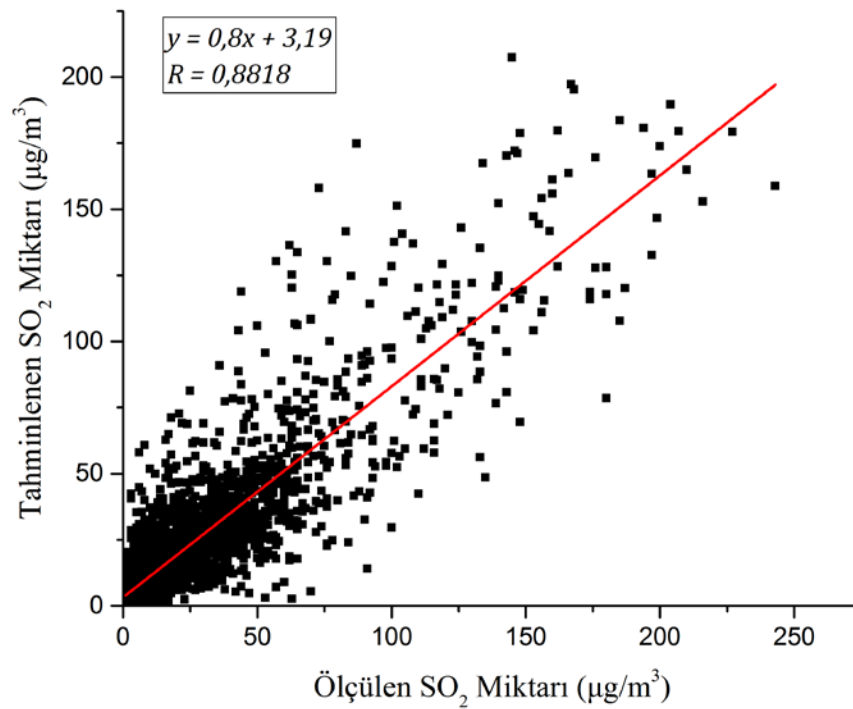


Şekil 8.1. Tahminleme modellerinin gerçek değerlerle uyumu a)SVR b)MLR c)MLP

Şekil 8.1’de zaman sıralı olmayan 365 gün için hava kalitesi izleme istasyonu tarafından ölçülen SO₂ konsantrasyonu ve kurulan modellerle tahminlenen SO₂ konsantrasyonunun kıyaslama grafiği verilmiştir. Şekil incelendiğinde üç yöntemin de seçilen günler için genel olarak iyi sonuçlar ürettiği görülmektedir. Ancak bu üç yöntem için de pik noktalarına ve ani değişimlere karşı yapılan tahminlemelerde, geri kalan tahminlemelere göre biraz daha başarısız olduğu söylenebilir. Bununla beraber kurulan modellerin SO₂ miktarının değişim yönüne uyum sağladıkları da belirtilmelidir. Son olarak

kullanılan yöntemler arasında kıyaslama yapıldığında, destek vektör regresyon yönteminin ürettiği sonuçların orijinal veri ile daha iyi bir uyum yakaladığı gözlemlenmiştir.

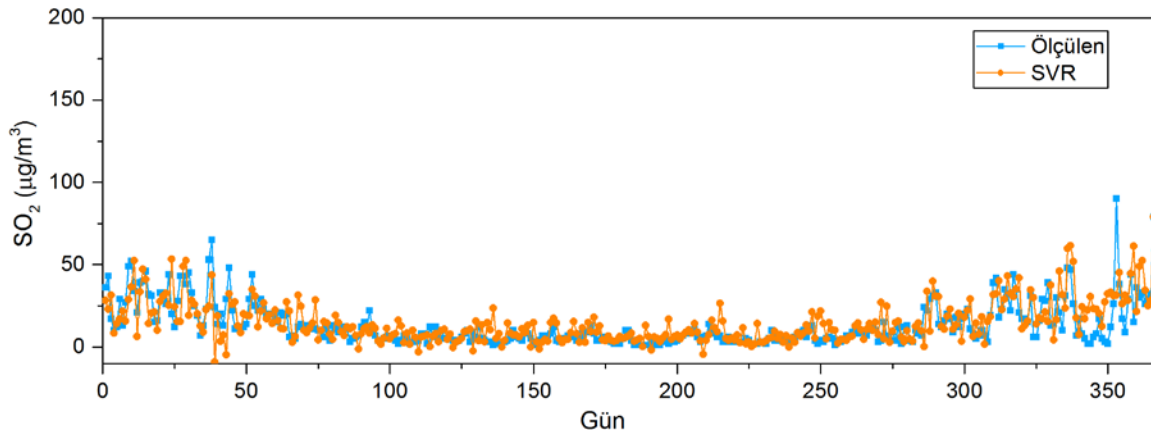
Şekil 8.2’de ölçülen ve SVR-RBF modeli tarafından tahminlenen SO₂ konsantrasyonlarının serpmeye diyagramı verilmiştir. Ölçülen ve tahminlenen SO₂ konsantrasyonları arasında model tarafından çok güçlü bir ilişki kurulduğu 0,8818 korelasyon katsayısıyla gösterilmiştir.



Şekil 8.2. Ölçülen ve tahminlenen SO₂ miktarları arasındaki doğrusal ilişki (Model 1)

Şekil 8.3’te hava kirliliği istasyonu tarafından ölçülen zaman sıralı 365 güne ait SO₂ konsantrasyonu ve SVR modeliyle tahminlenen SO₂ konsantrasyonunun kıyaslama grafiği verilmiştir. SO₂ konsantrasyonunun dar bir aralıkta değer aldığı, yani sıralı günler arasında küçük farklarla değiştiği bahar ve yaz dönemlerinde, tahminlerin gerçek veriyle daha iyi örtüştüğü görülmektedir. Yılın ilk ve son dönemlerinde ise sıcaklık ortalamasının yüksek olduğu aylara kıyasla, SO₂ konsantrasyonunda büyük değişimler gerçekleşmiş ve model bu ani değişikliklere göreceli olarak iyi tepkiler verememiştir. Bu durumun nedeni olarak Denizli’nin ılıman bir iklime sahip olması, dolayısıyla sıcaklık ortalamasının yüksek olduğu

dönemlerin yıl içinde daha fazla olması gösterilebilir (Meteoroloji Genel Müdürlüğü, 2016). Diğer bir neden olarak da bu dönemler içerisinde SO₂ konsantrasyonunun küçük bir aralık içinde değer alması gösterilebilir. Her iki durum da bu dönemlerdeki günlük hava kirliliği örüntüsün çeşitliliğinde azalmaya neden olacak dolayısıyla veri kümesi içinde daha sık tekrar edilmesini sağlayacaktır. Bu nedenle tahminleme modelinin bu örüntülere daha iyi uyum sağlaması, veri kümesinde oransal olarak az sayıda yer alan kış aylarındaki günlük hava kirliliği örüntülerine karşı ise daha duyarsız olması beklenmektedir.



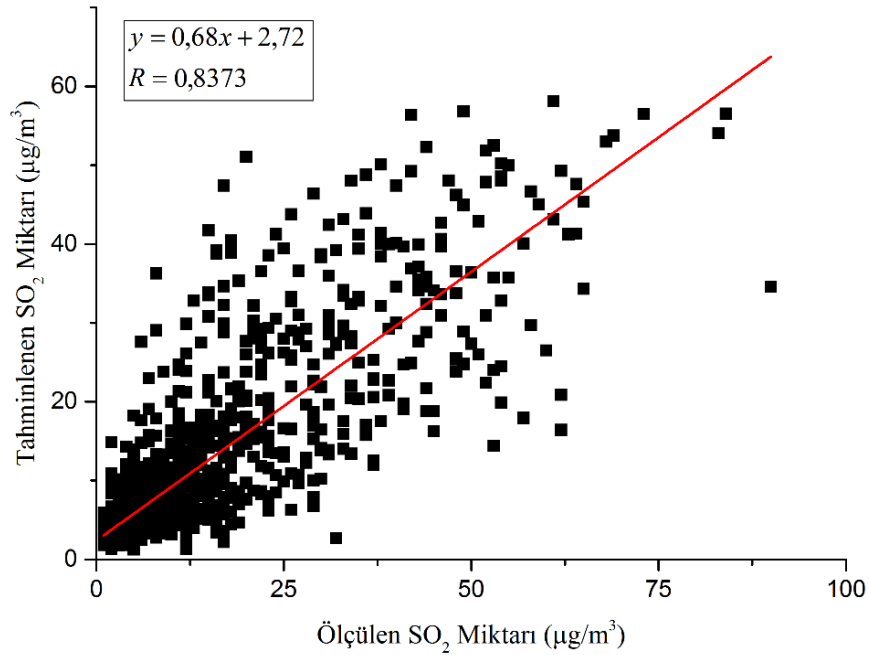
Şekil 8.3. SVR ile yapılan zaman sıralı yıllık tahminlerin gerçek değerlerle uyumu

Denizli’de 2008 yılı itibariyle doğalgaz kullanımına geçiş tam anlamıyla sağlanmış ve buna paralel olarak evsel ısınma ihtiyacını karşılayan kömür ve petrol kökenli yakıtların tüketiminde azalma gerçekleşmiştir. Doğalgazın yanma tepkimesi sonucunda açığa çıkan SO₂ konsantrasyonu, ihmal edilecek derecede düşük seviyededir. Bu nedenle Denizli’de doğalgaz kullanımına geçildikten sonra SO₂ miktarında önemli ölçüde düşüş gerçekleşmiştir (Yazıcı vd., 2010). Doğalgaz kullanımı öncesindeki verilerin yüksek değerli SO₂ konsantrasyonu içermesiyle, bu değerlerin veri setinin geneline göre daha farklı bir davranış göstermesi beklenmektedir. Ayrıca bu dönemde ölçülen SO₂ konsantrasyonların kısa ve orta vadede tekrar ölçülmesinin düşük olasılıklı olması nedeniyle tahminleme modelini olumsuz yönde etkileyeceği düşünülmektedir. Bu durumun tahminleme modeline olan etkisini incelemek adına doğalgaz kullanımı sonrası elde edilen verilerle bir tahminleme çalışması daha yapılmıştır. 10-kat çapraz geçiş sonrası elde edilen test sonuçları Çizelge 8.2’de gösterilmiştir.

Çizelge 8.2. Doğalgaz kullanımıyla başlayan döneme ait test sonuçları

	<i>R</i>	<i>MAE</i>	<i>RMSE</i>
<i>SVR-RBF</i>	0,8373	3,6361	6,7089
<i>MLP</i>	0,8336	4,0168	6,7614
<i>MLR</i>	0,8202	4,1879	6,9450

Test sonuçları incelendiğinde, doğalgaz kullanımına geçilmeden önce elde edilen verilerin modelden çıkarılmasıyla daha isabetli tahminler yapıldığı görülmüştür. İki durum kıyaslandığında doğalgaz kullanımına geçilmeden önce elde edilen verilerin modelden çıkarılmasıyla SVR ile kurulan model için ortalama mutlak hatada %51,1 azalma olmuştur. Bu tahminleme çalışmasında da SVR modelinin, MLP ve MLR modellerine göre daha başarılı sonuçlar ürettiği görülmüştür. Test sonuçlarına göre SVR, MLP'ye oranla %9,48 MLR'ye oranla ise %13,18 daha az ortalama mutlak hatayla tahminler gerçekleştirmiştir.



Şekil 8.4. Ölçülen ve tahminlenen SO₂ miktarları arasındaki doğrusal ilişki (Model 2)

Şekil 8.4'te doğalgaz kullanımından itibaren ölçülen ve SVR-RBF modeli ile tahminlenen SO₂ miktarlarının serpmeye diyagramı verilmiştir. Ölçülen ve tahminlenen SO₂

miktarları arasında model tarafından çok güçlü bir ilişki kurulduğu 0,8373 korelasyon katsayısıyla gösterilmiştir. Ayrıca Şekil 8.2 ile kıyaslama yapıldığında, doğalgaz kullanımına geçilmesiyle yüksek değerli SO₂ konsantrasyonlarının gözlemlenmediği görülmüştür.

9. SONUÇ VE ÖNERİLER

Hava kirliliği çevre ve insan sağlığı için önemli bir sorun olup ekolojik hayatı tehdit etmekte ve ölümlerle dahi sonuçlanan hastalıklara neden olmaktadır. Bu nedenle hava kirliliğinin, özellikle büyük yerleşim yerlerinde analiz edilmesi ve erken uyarı sistemlerinin geliştirilmesi gereklilik oluşturmaktadır. Bu amaçla bu çalışmada, bir sonraki günün hava kirliliği destek vektör regresyon ile tahminlenmeye çalışılmıştır. Kıyaslama yapabilmek adına çok katmanlı algılayıcı ve çoklu doğrusal regresyon yöntemleri ile de tahminleme modelleri kurulmuştur.

İlk tahminleme modeli, elde edilen verilerin tamamıyla kurulmuştur. Ancak doğalgaz kullanımına geçilmesiyle elde edilen verilerle yeni bir tahminleme modeli daha kurma ihtiyacı doğmuştur. Bu ihtiyacın doğmasında, doğalgaz kullanımına geçilmeden önceki ölçümlerin verinin geri kalanına göre oldukça yüksek değerlerde seyretmesi etkili olmuştur. Bu denli yüksek hava kirliliğine kısa ve orta vadede rastlanması beklenilmediğinden tahminleme modeline yapacağı olumsuz etkiler ortadan kaldırılmak istenmiştir. Beklenildiği gibi yeni tahminleme modeliyle daha başarılı sonuçlar elde edilmiştir.

Kurulan iki modelin de test sonuçları incelendiğinde, kullanılan üç yöntemin de kabul edilebilir sonuçlar ürettiği ancak SVR yönteminin daha başarılı olduğu görülmüştür. SVR'nin daha küçük hatalarla tahminler yapmasının sebebi olarak MLP'nin tipik sorunları olan yerel minimuma takılma ve aşırı uyumun üstesinden gelebilmesi gösterilebilir. SVR'nin karesel eniyilemeye dayalı bir yöntem olması, çözüm uzayında yerel minimuma takılma sorununu ortadan kaldırmakta ve çözüm noktasının her zaman bütünsel minimuma ulaşmasını sağlamaktadır. Ayrıca SVR'nin yapısal risk minimizasyonu üzerine kurulmuş olması genelleştirme kabiliyetini arttırmakta, aşırı uyum sorununu aşmasını sağlamaktadır. SVR'nin MLP'ye göre diğer bir avantajı ise, belirlenmesi gereken parametre sayısının daha az olmasıdır. Ayrıca bu çalışmada SVR'nin parametrelerinin çözüme olan etkisinin MLP'ye kıyasla daha küçük olduğu görülmüştür. SVR'nin dezavantajının ise, uygun çekirdek fonksiyonunu belirleme işleminin hesaplama maliyetini artırması olduğu söylenebilir.

Kirletici seviyelerindeki ani deęişimlerin tahminleme başarısını dūşürdüęü Bölüm 8'de tartışılmıştır. Kirletici deęerlerindeki bu dalgalanmaları azaltmak için günlük ortalamalar yerine saatlik veriler kullanılabilir. Böylece birbirini takip eden kirletici deęerleri daha yumuřak geçiřlerle deęiřeceęinden gerçek deęerlere daha yakın tahminler yapılabilir. Ayrıca bu durum veri kümesindeki örnek sayısını dolayısıyla çeřitlilięi arttıracaktır. Modelin daha fazla örnekle eęitilmesiyle genelleřtirme yeteneęinin de artması beklenmektedir.

Kirletici konsantrasyonuyla doğrudan iliřkili olan evsel ısınma, řehirdeki trafik yoğunluęu, endüstriyel faaliyetler gibi emisyon kaynaklarının modele eklenmesi tahminleme başarımını olumlu yönde etkileyebilir. Ayrıca kirleticilerin benzer emisyon kaynaklarından etkilenmesi sebebiyle modele CO, NO₂, gibi kirleticilerin de eklenmesi daha isabetli tahminler yapma konusunda yardımcı olabilir.

Sıcaklık ortalamasının yüksek olduęu dönemlerde kirletici konsantrasyonu duraęan bir seyir izlerken kış mevsiminde kirletici konsantrasyonunda dalgalanmalar görölmektedir. Çalışmanın yapıldıęı ilde sıcaklık ortalamasının yüksek olduęu dönemlerin yıl içinde oransal olarak daha fazla yer alması, kış döneminin baskılanmasına neden olmaktadır. Böylece model daha sık tekrar eden örüntüler içeren duraęan dönemlere daha iyi uyum göstermektedir. Bu sorunun önüne geçmek için kış dönemi ve sıcaklık ortalamasının yüksek olduęu dönemler için iki ayrı model kurulabilir. Böylece kurulan modellerin ilgili dönemlere daha iyi uyum saęlayacaęı söylenebilir.

Ayrıca meteorolojik verilerin ve buna baęlı olarak hava kirlilięinin zaman serisi analizine uygun bir yapıda olması nedeniyle bu doğrultuda tahminleme modeli de kurularak elde edilen sonuçlar dięer yöntemlerle karşılaştırılabilir.

KAYNAKLAR DİZİNİ

- Abe, S., 2010, Support Vector Machines for Pattern Classification, 2nd Edition, Springer, p.30-31.
- Alavala, C.R., 2012, Fuzzy Logic and Neural Networks: Basic Concepts & Applications, New Age International, p.131, 135, 139.
- Alpaydın, E., 2010, Introduction to Machine Learning, 2nd Edition, The MIT Press, p.27
- Annema, J., 2012, Feed-Forward Neural Networks: Vector Decomposition Analysis, Modelling and Analog Implementation, Springer Science & Business Media, p.9.
- Arora, J., 2011, Introduction to Optimum Design, Academic Press, p.139.
- Atımtay, A., Bayram, H., Can, A., Çımrın, A.H., Demiral, B., Elçi, M.A., vd., 2010, Türkiye'nin Hava Kirliliği ve İklim Değişikliği Sorunlarına Sağlık Açısından Yaklaşım, T.C. Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü, s.85.
- Bar-Joseph, Z., El-Yaniv, R., Lischinski, D., Werman, M., 2001, Texture mixing and texture movie synthesis using statistical learning, IEEE Transactions on Visualization and Computer Graphics, 7, 2, 120-135.
- Batista, G.E.A.P.A., Monard, M.C., 2002, A study of k-nearest neighbour as an imputation method, Soft Computing Systems: Design, Management and Applications, IOS Press, 251-260.
- Bayram, H., Dörtbudak, Z., Fişekçi, F.E., Kargın, M., Bülbül, B., 2006, Hava kirliliğinin insan sağlığına etkileri, dünyada, ülkemizde ve bölgemizde hava kirliliği sorunu paneli ardından, Dicle Tıp Dergisi, 33, 2, 105-112.
- Beale, R., Jackson, T., 1990, Neural Computing: An Introduction, Taylor & Francis Group, p.118.
- Beckrich, A., 2013, Good ozone, bad ozone, The Science Teacher, 80, 7, 10.
- Beldjazia, A., Alatou, D., 2016, Precipitation variability on the massif Forest of Mahouna (North Eastern-Algeria) from 1986 to 2010, International Journal of Management Sciences and Business Research, 5, 3, 21-28.
- Ben-Hur, A., Weston, J., 2010, A User's Guide to Support Vector Machines, Data Mining Techniques for the Life Sciences, O. Carugo, F. Eisenhaber (Eds.), Humana Press, p.223-239.
- Bergstra, J., Bengio, Y., 2012, Random search for hyper-parameter optimization, Journal of Machine Learning Research, 13, 281-305.

KAYNAKLAR DİZİNİ (devam)

- Bhaskar, B.V., Mehta, V.M., 2010, Atmospheric particulate pollutants and their relationship with meteorology in Ahmedabad, *Aerosol and Air Quality Research*, 10, 4, 301-315.
- Birkes, D., Dodge, Y., 1993, *Alternative Methods of Regression*, John Wiley & Sons, p.32.
- Boser, B., Guyon, I., Vapnik, V.N., 1992, A training algorithm for optimal margin classifiers, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144-152.
- Boznar, M., Lesjak, M., Mlakar, P., 1993, A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain, *Atmospheric Environment Part B: Urban Atmosphere*, 27, 2, 221-230.
- Breiman, L., 1994, *Bagging predictors*, Technical report, University of California, 20 p.
- Bridgman, H.A., Davies, T.D., Jickells, T., Hunova, I., Tovey, K., vd., 2002, Air pollution in the Krusne Hory region, Czech Republic during the 1990s, *Atmospheric Environment*, 36, 21, 3375–3389.
- Brook R.D., Franklin B., Cascio W., Hong Y., Howard G., vd., 2004, Air pollution and cardiovascular disease: A statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association, *Circulation*, 109, 21, 2655-2671.
- Burges, C.J.C., 1998, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2, 2, 121-167.
- Camilleri, M., Neri, F., Papoutsidakis, M., 2014, An algorithmic approach to parameter selection in machine learning using meta-optimization techniques, *WSEAS Transactions on Systems*, 13, 203-212.
- Campbell, C., Ying, Y., 2011, *Learning With Support Vector Machines*, Morgan & Claypool, p.39.
- Carpenter, G.A., Gjaja, M.N., 1994, Fuzzy ART chooses function, *World Congress on Neural Networks*, 2, 713-722.
- Carpenter, G.A., Grossberg, S., 1987, A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics, and Image Processing*, 37,1, 54-115.

KAYNAKLAR DİZİNİ (devam)

- Chakraborty, C., 2010, Adaptive Resonance Theory: Soft Computing Course Lecture 25-28, http://www.myreaders.info/05-Adaptive_Resonance_Theory.pdf, erişim tarihi: 20.12.2015.
- Chang, C.C., Lin, C.J., 2001, LIBSVM: A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>, erişim tarihi: 8.09.2015.
- Chatterjee, S., Hadi, A.S., 2012, Regression Analysis by Example, 5th Edition, John Wiley & Sons, p.32, 41, 68.
- Chelani, A.B., 2010, Prediction of daily maximum ground ozone concentration using support vector machine, Environmental Monitoring and Assessment, 162, 1, 169-176.
- Chelani, A.B., Gajghate, D.G., Hasan, M.Z., 2002, Prediction of ambient PM₁₀ and toxic metals using artificial neural networks, Journal of the Air & Waste Management Association, 52, 7, 805-810.
- Cherkassky, V., Ma, Y., 2002, Selecting the loss function for robust linear regression, Neural Computation.
- Cherkassky, V., Mulier, F.M., 2007, Learning From Data: Concepts, Theory, and Methods, 2nd Edition, Wiley-IEEE Press, p.17, 116, 122.
- Comrie, A.C., 1997, Comparing neural networks and regression models for ozone forecasting, Journal of the Air & Waste Management Association, 47, 6, 653-663.
- Cortes, C., Vapnik, V.N., 1995, Support-vector networks, Machine Learning, 20, 3, 273-297.
- Cristianini, N., Shawe-Taylor, J., 2000, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, p.28, 33, 103.
- Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems, 3, 4, 303-314.
- Çuhadaroğlu, B., Demirci, E., 1997, Influence of some meteorological factors on air pollution in Trabzon city, Energy and Buildings, 25, 3, 179-184.
- Demartines, P., Blayo, F., 1992, Kohonen self-organizing maps: Is the normalization necessary?, Complex System, 6, 105-123.
- Deng, N., Tian, Y., Zhang, C., 2013, Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions, Chapman & Hall/CRC, p.55, 76-77.

KAYNAKLAR DİZİNİ (devam)

- Diaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., vd., 2008, Hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile, *Atmospheric Environment*, 42, 35, 8331–8340.
- Dominick, D., Latif, M.T., Juahir, H., Aris, A.Z., Zain, S.M., 2012, An assessment of influence of meteorological factors on PM₁₀ and NO₂ at selected stations in Malaysia, *Sustainable Environment Research*, 22, 5, 305-315.
- Drucker, H., Burges, C.J.C, Kaufman, L., Smola, A.J., Vapnik, V.N., 1997, Support vector regression machines, *Advances in Neural Information Processing Systems 9*, M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), MIT Press, 155-161.
- EPA, 1992, Terms of Enviroment: Glossary, Abbreviations and Acronyms.
- EPA, 1994, Measuring air quality: The pollutant standards index, Office of Air Quality Planning and Standards.
- EPA, 2011 a,
http://www.epa.gov/cgibin/broker?polchoic=PM&_debug=0&_service=data&_program=dataprog.national_1.sas, erişim tarihi: 26.10.2015.
- EPA, 2011 b,
http://www.epa.gov/cgibin/broker?polchoic=CO&_debug=0&_service=data&_program=dataprog.national_1.sas, erişim tarihi: 26.10.2015.
- EPA, 2011 c,
http://www.epa.gov/cgibin/broker?polchoic=SO2&_debug=0&_service=data&_program=dataprog.national_1.sas, erişim tarihi: 26.10.2015.
- EPA, 2011 d,
http://www.epa.gov/cgibin/broker?polchoic=NOX&_debug=0&_service=data&_program=dataprog.national_1.sas, erişim tarihi: 26.10.2015.
- EPA, 2013, <http://www.epa.gov/airquality/particlepollution/>, erişim tarihi: 3.08.2015.
- EPA, 2014, <http://www.epa.gov/airquality/nitrogenoxides/>, erişim tarihi: 3.08.2015.
- Evans, J.D., 1996, *Straightforward Statistics for the Behavioral Sciences*, Brooks Cole Publishing Company.
- Fahimeh H., Azadeh, H., 2012, Influence of meteorological parameters on air pollution in Isfahan, 3rd International Conference on Biology, Environment and Chemistry, 46, 2.

KAYNAKLAR DİZİNİ (devam)

- Fausett, L., 1994, Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, p.3, 22, 48, 80, 88, 295-296.
- Feng, Y., Zhang, W., Sun, D., Zhang, L., 2011, Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification, Atmospheric Environment, 45, 11, 1979-1985.
- Freeman, J.A., Skapura, D.M., 1991, Neural Networks: Algorithms, Applications, and Programming Techniques, Addison-Wesley Publishing Company, p.100, 102.
- García Nieto, P.J., Combarro, E.F., Del Coz Díaz, J.J., Montañés, E., 2013, A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study, Applied Mathematics and Computation, 219, 17, 8923–8937.
- Gardner, M.W., Dorling, S.R., 1999, Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London, Atmospheric Environment, 33, 5, 709-719.
- Genton, M.G., 2001, Classes of kernels for machine learning: A statistics perspective, Journal of Machine Learning Research, 2, 299-31.
- Girma, H., 2009, A Tutorial on Support Vector Machine, University of Ljubljana, <https://pdfs.semanticscholar.org/52cf/2494f7375f528b53ab7be979110671b4c2dd.pdf>, erişim tarihi: 21.10.2015.
- Gómez-Perales, J.E., Colvile, R.N., Nieuwenhuijsen, M.J., Fernández-Bremauntz, A., Gutiérrez-Avedoy, V.J., vd., 2004, Commuters' exposure to PM_{2.5}, CO, and benzene in public transport in the metropolitan area of Mexico City, Atmospheric Environment, 33, 8, 1219–1229.
- Grivas, G., Chaloulakou, A., 2006, Artificial neural network models for prediction of PM₁₀ hourly concentrations, in the greater area of Athens, Greece, Atmospheric Environment, 40, 7, 1216–1229.
- Gunn, S.R., 1998, Support vector machines for classification and regression, Technical report, University of Southampton, 66 p.
- Gurney, K., 1997, An Introduction to Neural Networks, Kindle Edition, UCL Press, p.43-44.
- Gutierrez-Osuna, R., 2002, Competitive Learning, http://research.cs.tamu.edu/prism/lectures/pr/pr_116.pdf, erişim tarihi: 20.12.2015.

KAYNAKLAR DİZİNİ (devam)

- Hajek, M., 2005, Neural Networks,
<http://www.cs.ukzn.ac.za/notes/NeuralNetworks2005.pdf>, erişim tarihi: 20.12.2015.
- Hajek, P., Olej, V., 2015, Predicting common air quality index – The case of Czech Microregions, *Aerosol and Air Quality Research*, 15, 2, 544–555.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., vd., 2009, The WEKA data mining software: An update, *SIGKDD Explorations*, 11, 1, 10-18.
- Hamel, L., 2009, *Knowledge Discovery With Support Vector Machines*, John Wiley & Sons, p.101, 178.
- Han, J., Kamber, K., Pei, J., 2011, *Data Mining: Concepts and Techniques*, 3rd Edition, The Morgan Kaufmann, p.82-83, 87, 93, 105.
- Han, X., Naeher, L.P., 2006, A review of traffic-related air pollution exposure assessment studies in the developing world, *Environment International*, 32, 1, 106-120.
- Hastie, T., Tibshirani, R., Friedman, J., 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, p.242.
- Haykin, S., 2009, *Neural Networks and Learning Machines*, 3rd Edition, Pearson Education, p.137.
- He, X., Xu, S., 2010, *Process Neural Networks: Theory and Applications*, Springer Science & Business Media, p.128.
- Hewson, E.W., 1956, Meteorological factors affecting causes and controls of air pollution, *Journal of the Air Pollution Control Association*, 5, 4, 235-241.
- Hoffman, M., 2006, *Support Vector Machines - Kernels and the Kernel Trick*, http://www.cogsys.wiai.unibamberg.de/teaching/ss06/hs_svm/slides/SVM_Seminarbericht_Hofmann.pdf, erişim tarihi: 23.11.2015.
- Hoffman, T., Schölkopf, B., Smola, A.J., 2008, Kernel methods in machine learning, *The Annals of Statistics*, 36, 3, 1171–1220.
- Hopfield, J.J., 1982, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of National Acad Sciences*, 79, 8, 2554–2558.
- Hornik, K., Stinchcombe, M., White, H., 1989, Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 5, 359–366.
- Hsu, C.W., Chang, C.C., Lin, C.J., 2003, *A Practical Guide to Support Vector Classification*, <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, erişim tarihi: 8.09.2015.

KAYNAKLAR DİZİNİ (devam)

- Hyndman, R.J., Athanasopoulos, G., 2014, Forecasting: Principles and Practice, OTexts, p.46.
- Igel, C., 2014, Machine Learning: Kernel-based Methods Lecture Notes, <http://image.diku.dk/igel/teaching/KernelBasedMachineLearning.pdf>, erişim tarihi: 15.03.2015.
- Jacobson, M.Z., 2005, Air Pollution and Global Warming: History, Science, and Solution, Cambridge University Press, p.127.
- Jain, A.K., Mao, J., Mohiuddin, K.M., 1996, Artificial neural networks: A tutorial, Computer, 29, 3, 31-44.
- Kaminski, W., Skrzypski J., Strumillo, P., 1999, Forecasting of air pollution in urban areas by means of artificial neural networks, WIT Transactions on The Built Environment: Urban Transport V, L.J. Sucharov (Ed.), 44.
- Kecman, V., 2001, Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models, The MIT Press, p.140, 145,165,170-171.
- Khan, W.U., Makwana, A., 2005, Soft Computing, D. Garg, A. Singh (Eds.), Allied Publishers, p.33.
- Kohonen, T., 1982, Self-organized formation of topologically correct feature maps, Biological Cybernetics, 43, 1, 59-69.
- Kriesel, D., 2005, A Brief Introduction to Neural Networks, e-kitap, 8, 86, http://www.dkriesel.com/en/science/neural_networks, erişim tarihi: 18.12.2015.
- Kröse, B., Smagt, P., 1996, An Introduction to Neural Networks, 8th Edition, e-kitap, The University of Amsterdam, 17, <https://www.infor.uva.es/~teodoro/neuro-intro.pdf>, erişim tarihi: 18.12.2015.
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., vd., 2003, Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki, Atmospheric Environment, 37, 32, 4539–4550.
- Lin, K.P., Pai, P.F., Yang, S.L., 2011, Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms, Applied Mathematics and Computation, 217, 12, 5318–5327.

KAYNAKLAR DİZİNİ (devam)

- Lu, W., Wang, W., Leung, A.Y.T., Lo, S.M., Yuen, R.K.K., vd., 2002, Air pollutant parameter forecasting using support vector machines, Proceedings of the 2002 International Joint Conference on Neural Networks, 1, 630-635.
- Lu, W.Z., Wang, D., 2008, Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme, The Science of the Total Environment, 395, 2, 109-116.
- Lu, W.Z., Wang, W.Z., 2005, Potential assessment of the support vector machine method in forecasting ambient air pollutant trends, Chemosphere, 59, 5, 693-701.
- McCulloch, W., Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics, 5, 4, 115-133.
- Mehrotra, K., Mohan, C.K., Ranka, S., 1996, Elements of Artificial Neural Networks, The MIT Press, p.9, 85.
- Meteoroloji Genel Müdürlüğü, 2016, <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?m=DENIZLI>, erişim tarihi: 12.12.2016.
- Mitchell, T.M., 1997, Machine Learning, McGraw-Hill, p.88.
- Miyazaki, T., Yamaoka, S., 1991, Meteorological factors causing high dust concentrations, Energy and Buildings, 16, 1, 691-698.
- Montgomery, D.C., Peck, E.A., Vining, G.G., 2012, Introduction to Linear Regression Analysis, 5th Edition, John Wiley & Sons, p.1, 20, 72, 80.
- Murphy, A.H., 1995, The coefficients of correlation and determination as measures of performance in forecast verification, Weather and Forecasting, 10, 4, 681-688.
- Müller, K.R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., vd., 1997, predicting time series with support vector machines, International Conference on Artificial Neural Networks, Springer Berlin Heidelberg, 999-1004.
- Negnevitsky, M., 2005, Artificial Intelligence: A Guide to Intelligent Systems, 2nd Edition, Addison-Wesley, p.189, 207.
- NOAA, 2015, http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=)), erişim tarihi: 3.06.2015.
- Noback, C.R., Strominger, N.L., Demarest, R.J., Ruggiero, D.A., 2005, The Human Nervous System: Structure and Function, Springer Science & Business Media, p.67.

KAYNAKLAR DİZİNİ (devam)

- Olson, D.L., Delen, D., 2008, *Advanced Data Mining Techniques*, Springer, p.111.
- Omaye, S.T., 2002, Metabolic modulation of carbon monoxide toxicity, *Toxicology*, 180, 2, 139-150.
- Ortiz-García, E.G., Salcedo-Sanz, S., Pérez-Bellido, Á.M., Portilla-Figueras, J.A., Prieto, L., 2010, Prediction of hourly O₃ concentrations using support vector regression algorithms, *Atmospheric Environment*, 44, 35, 4481-4488.
- Osowski, S., Garanty, K., 2007, Forecasting of the daily meteorological pollution using wavelets and support vector machine, *Engineering Applications of Artificial Intelligence*, 20, 6, 745–755.
- Özdamar, K., 2004, *Paket Programlar İle İstatistiksel Veri Analizi (Çok Değişkenli Analizler)*, 5. Baskı, Kaan Kitabevi, s.91.
- Özkan, Y., 2013, *Veri Madenciliği Yöntemleri*, 2. Basım, Papatya Yayıncılık, s.40-41.
- Öztemel, E., 2012, *Yapay Sinir Ağları*, 3. Basım, Papatya Yayıncılık, s.137.
- Prasad, K., Gorai, A.K., Goyal, P., 2016, Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time, *Atmospheric Environment*, 128, 246-262.
- Priddy, K.L., Keller, P.E., 2005, *Artificial Neural Networks: An Introduction*, SPIE Press, p.5.
- Rojas, R., 1996, *Neural Networks: A Systematic Introduction*, Springer, p.374.
- Rom, W.N., 2011, *Environmental Policy and Public Health: Air Pollution, Global Climate Change, and Wilderness*, Jossey-Bass, p.42.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986, Learning representations by back-propagating errors, *Nature*, 323, 533-536.
- Sánchez, A.S., García Nieto, P.J., Fernández, P.R., Del Coz-Díaz, J.J., Iglesias-Rodríguez, F.J., 2011, Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain), *Mathematical and Computer Modelling*, 54, 5, 1453–1466.
- Schölkopf, B., 2000, The kernel trick for distances, *Proceedings of the 13th International Conference on Neural Information Processing Systems*, 283-289.

KAYNAKLAR DİZİNİ (devam)

- Schölkopf, B., Smola, A.J., 2002, Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press, p.7, 200.
- Sewell, M., 2008, VC Dimension, <http://www.svms.org/vc-dimension/vc-dimension.pdf>, erişim tarihi: 8.11.2015.
- Sewell, M., 2009, Structural Risk Minimization, <http://www.svms.org/srm/srm.pdf>, erişim tarihi: 8.11.2015.
- Sezgin, N., Tekin, R., Çalışkan, A., 2012, Örüntü tanımada Hopfield ağınn kullanılması, Yaşam Bilimleri Dergisi, 2, 1, 257-266.
- Shanker, M., Hu, M.Y., Hung, M.S., 1996, Effect of data standardization on neural network training, Omega, 24, 4, 385–397.
- Sheather, S.J., 2009, A Modern Approach to Regression With R, Springer, p.134.
- Singh, K.P., Gupta, S., Kumar, A., Shukla, S.P., 2012, Linear and nonlinear modeling approaches for urban air quality prediction, Science of the Total Environment, 426, 244–255.
- Sivanandam, S.N., Sumathi, S., Deepa, S.N., 2006, Introduction to Neural Networks Using Matlab 6.0, McGraw-Hill, p.187, 260-261.
- Smith, B.R., 1997, Neural network enhancement of closed-loop controllers for ill-modeled systems with unknown nonlinearities, Ph.D. thesis, Faculty of the Virginia Polytechnic Institute and State University, 139 p.
- Smola, A.J., 1996, Regression estimation with support vector learning machines, M.Sc. thesis, Technische Universität München, 78 p.
- Smola, A.J., Bartlett, P., Schölkopf, B., Schuurmans, D., 1999, Advances in Large-Margin Classifiers, The MIT Press.
- Smola, A.J., Schölkopf, B., 2003, A tutorial on support vector regression, Statistics and Computing, 14, 3, 199–222.
- Snell, R.S., 2010, Clinical Neuroanatomy, Lippincott Williams & Wilkins, p.34.
- Soman, K.P., Loganathan, R., Ajay V., 2009, Machine Learning With SVM and Other Kernel Methods, PHI Learning, p.135.

KAYNAKLAR DİZİNİ (devam)

- Sotomayor-Olmedo, A., Aceves-Fernández, M.A., Gorrostieta-Hurtado, E., Pedraza-Ortega, C., Ramos-Arreguín, J.M., vd., 2013, Forecast urban air pollution in Mexico City by using support vector machines: A kernel performance approach, *International Journal of Intelligence Science*, 3, 3, 126-135.
- Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., 2007, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environmental Modelling & Software*, 22, 1, 97–103.
- T.C. Çevre ve Şehircilik Bakanlığı, 2015, <http://www.havaizleme.gov.tr/>, erişim tarihi: 3.06.2015.
- Tapkan, P., Özbakır, L., Baykasoğlu, A., 2011, Weka ile veri madenciliği süreci ve örnek uygulama, *Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi*, 247-262.
- Tay, F.E.H., Cao, L., 2001, Application of support vector machines in financial time series forecasting, *Omega*, 29, 4, 309–317.
- Taylor, B.J., 2006, *Methods and Procedures for the Verification and Validation of Artificial Neural Networks*, Springer Science & Business Media, p.71.
- Turalıoğlu, F.S., Nuhoglu, A., Bayraktar, H., 2005, Impacts of some meteorological parameters on SO₂ and TSP concentrations in Erzurum, Turkey, *Chemosphere*, 59, 11, 1633–1642.
- Ul-Saufie, A.Z., Yahay, A.S., Ramli, N.A., Rosaida, N., Hamid, H.A., 2013, Future daily PM₁₀ concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA), *Atmospheric Environment*, 77, 621-630.
- Vallero, D., 2007, *Fundamentals of Air Pollution*, Academic Press, p.58, 581.
- Vapnik, V.N., 1979, *Estimation of Dependences Based on Empirical Data*, Nauka.
- Vapnik, V.N., 1995, *The Nature of Statistical Learning Theory*, Springer
- Vapnik, V.N., 1999, An overview of statistical learning theory, *IEEE Transactions on Neural Networks*, 10, 5, 988-999.
- Vapnik, V.N., 2000, *The Nature of Statistical Learning Theory*, 2nd Edition, Springer, p.138.
- Vapnik, V.N., Chervonenkis, A., 1964, On a perceptron class, *Avtomat i Telemekh*, 25, 1, 112–120.

KAYNAKLAR DİZİNİ (devam)

- Vapnik, V.N., Chervonenkis, A., 1974, Theory of Pattern Recognition: Statistical Problems of Learning, Nauka.
- Vapnik, V.N., Golowich, S.E., Smola, A.J., 1997, Support vector method for function approximation, regression estimation, and signal processing, Advances in Neural Information Processing Systems 9, M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), MIT Press, 281-287.
- Vapnik, V.N., Lerner, A., 1963, Recognition of patterns with help of generalized portraits, Avtomat i Telemekh, 24, 6, 774–780.
- Vapnik, V.N., Levin, E., Cun, Y.L., 1994, Measuring the VC-dimension of a learning machine, Neural Computation, 6, 5, 851-876.
- Verma, S.S., Desai, B., 2008, Effect of meteorological conditions on air pollution of Surat city, Journal of International Environmental Application & Science, 3, 5, 358-367.
- Wang, W., Men, C., Lu, W., 2008, Online prediction model based on support vector machine, Neurocomputing, 71, 4, 550–558.
- Weston, J., 2006, Support Vector Machine (and Statistical Learning Theory) Tutorial, http://www.cs.columbia.edu/~kathy/cs4701/documents/jason_svm_tutorial.pdf, erişim tarihi: 8.11.2015.
- Witten, I.H., Frank, E., Hall, M.A., 2011, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, p.48, 180, 404-405, 506.
- World Bank, Institute for Health Metrics and Evaluation, 2016, The cost of air pollution: Strengthening the economic case for action, World Bank, 102 p.
- Yang, J.Y., Ip, W.F., Vong, C.M., Wong, P.K., 2011, Effect of choice of kernel in support vector machines on ambient air pollution forecasting, 2011 International Conference on System Science and Engineering.
- Yazıcı, H., Akçay, M., Çay, Y., Sekmen, Y., Yılmaz, İ.T., vd., 2010, Hava kirliliğinin doğal gaz kullanımı ile değişimi, Denizli ili örneği, Teknik-Online Dergi, 9, 3, 205-215.,
- Yeganeh, B., Motlagh, M.S.P., Rashidi, Y., Kamalan, H., 2012, Prediction of CO concentrations based on a hybrid partial least square and support vector machine model, Atmospheric Environment, 55, 357-365.
- Yegnanarayana, B., 1999, Artificial Neural Networks, p.26-27, 88.
- Zaki, M.J., Meira Jr, W., 2014, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, p.525.

KAYNAKLAR DİZİNİ (devam)

- Zhang, H., Wang, Y., Hu, J., Ying, Q., Hu, X.M., 2015, Relationships between meteorological parameters and criteria air pollutants in three megacities in China, *Environmental Research*, 140, 242-254.
- Zhang, X., 1999, Using class-center vectors to build support vector machines, *Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 3-11.
- Zheng, L., Jia, H., Yu, S., Yu, M., 2010, Prediction of NO_x concentration from coal combustion using LS-SVR, *4th International Conference on Bioinformatics and Biomedical Engineering*.
- Zurada, J.M., 1992, *Introduction to Artificial Neural Systems*, West Publishing Company, p.55, 57, 66-67.

EK AÇIKLAMALAR

Ek Açıklama-A: Geriyayılım Algoritması

Ek Açıklama-B: Destek Vektör Sınıflandırıcı İçin Nümerik Bir Örnek

Ek Açıklama-C: Destek Vektör Regresyon İçin Açıklayıcı Bir Örnek

Ek Açıklama-D: Örnek Veri Seti

Ek Açıklama-E: Örnek Normalleştirilmiş Veri Seti

Ek Açıklama-F: Çok Katmanlı Algılayıcı Mimarisi

Ek Açıklama-A: Geriyayılım Algoritması

Adım 0: Ağırlıklar için rassal başlangıç değerleri ata.

Adım 1: Durdurma kriteri sağlanana dek Adım 2-9 arasını yinele.

Adım 2: Her biri girdi-çıkıtsı ikilisi için Adım 3-8 arasını yinele

İleri besleme:

Adım 3: x_i girdi sinyali bir sonraki katmana aktar.

Adım 4: Gizli katman nöronları için toplam sinyali hesapla.

$$z_j = \sum w_{ij} x_j - b_j$$

Gizli katman nöronları için çıktı değerlerini hesapla ve çıktı katmanına aktar.

$$y_j = f(z_j)$$

Adım 5: Çıktı katmanı nöronları için toplam sinyali hesapla.

$$z_k = \sum w_{jk} x_k - b_k \quad (y_j = x_k)$$

Sinir ağının çıktılarını hesapla.

$$y_k = f(z_k)$$

Geriyayılım:

Adım 6: Çıktı katmanının hata değerini hesapla ve gizli katmana ilet.

$$\delta_k = -(d_k - y_k) f'(z_k)$$

Çıktı katmanına iletilen ağırlıkların değişimini hesapla.

$$\Delta w_{jk} = +\eta \delta_k x_k$$

Çıktı katmanı nöronlarındaki eşik değer değişimini hesapla.

$$\Delta b_k = \eta \delta_k$$

Adım 7: Çıktı katmanından gizli katmana yayılan hatayı hesapla.

$$\delta_{kj} = \sum_k \delta_k w_{jk}$$

Gizli katmanın hata değerini hesapla ve bir önceki katmana ilet.

$$\delta_j = \delta_{kj} f'(z_k)$$

Gizli katmana iletilen ağırlıkların değişimini hesapla.

$$\Delta w_{ij} = \eta \delta_j x_j \quad (x_j = x_i)$$

Gizli katman nöronlarındaki eşik değer değişimini hesapla.

$$\Delta b_j = \eta \delta_j$$

Ağırlık ve Eşik Değerlerin Güncellenmesi:

Adım 8: Çıktı katmanına iletilen ağırlıkları günceller

$$w_{jk(t+1)} = w_{jk(t)} + \eta \delta_k x_k$$

Çıktı katmanı nöronlarındaki eşik değerleri güncelle

$$b_{k(t+1)} = b_{k(t)} + \eta \delta_k$$

Gizli katmana iletilen ağırlıkları güncelle

$$w_{ij(t+1)} = w_{ij(t)} + \eta \delta_j x_j$$

Gizli katman nöronlarındaki eşik değerleri güncelle

$$b_{j(t+1)} = b_{j(t)} + \eta \delta_j$$

Adım 9: Durdurma kriterini sına.

Ek Açıklama-B: Destek Vektör Sınıflandırıcı İçin Nümerik Bir Örnek

Aşağıda +1 ve -1 sınıflarına ait üç adet gözlem değeri verilmiştir. Bu gözlemlerden yararlanılarak SVM ile verileri birbirlerinden ayıran karar sınırı araştırılmak isteniyor.

+1 Sınıfı: (0, 0) ve (0, 1)

-1 Sınıfı: (1, 1)

Cözüm:

Gözlem değerleri (x_i) ve sınıfların (y) matrisel ifadesi aşağıdaki gibi yazılır.

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Değerler Denklem (6.28)'de belirtilen Lagrange fonksiyonunda probleme uygun bir biçimde açık olarak yazılırsa aşağıdaki ifade elde edilir.

$$\begin{aligned} L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} & (\alpha_1 \alpha_1 y_1 y_1 x_1^T x_1 + \alpha_1 \alpha_2 y_1 y_2 x_1^T x_2 + \alpha_1 \alpha_3 y_1 y_3 x_1^T x_3 \\ & + \alpha_2 \alpha_1 y_2 y_1 x_2^T x_1 + \alpha_2 \alpha_2 y_2 y_2 x_2^T x_2 + \alpha_2 \alpha_3 y_2 y_3 x_2^T x_3 \\ & + \alpha_3 \alpha_1 y_3 y_1 x_3^T x_1 + \alpha_3 \alpha_2 y_3 y_2 x_3^T x_2 + \alpha_3 \alpha_3 y_3 y_3 x_3^T x_3) \end{aligned}$$

$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ olduğundan bu gözlemi içeren çarpımlar sıfır olacaktır. Böylece Lagrange fonksiyonu daha basit bir şekilde aşağıdaki gibi yazılabilir.

$$\begin{aligned} L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} & (\alpha_2 \alpha_2 y_2 y_2 x_2^T x_2 + \alpha_2 \alpha_3 y_2 y_3 x_2^T x_3 \\ & + \alpha_3 \alpha_2 y_3 y_2 x_3^T x_2 + \alpha_3 \alpha_3 y_3 y_3 x_3^T x_3) \end{aligned}$$

Değerlerin yerine konulmasıyla Lagrange fonksiyonu şöyle yazılır.

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} \left\{ \begin{aligned} & \alpha_2^2 (1)^2 [0 \quad 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \alpha_2 \alpha_3 (1)(-1) [0 \quad 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ & + \alpha_3 \alpha_2 (-1)(1) [1 \quad 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \alpha_3^2 (-1)^2 [1 \quad 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned} \right\}$$

Gerekli işlemler yapıldığında Lagrange fonksiyonu aşağıdaki gibi elde edilir.

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(\alpha_2^2 - 2\alpha_2\alpha_3 + 2\alpha_3^2)$$

Denklem (6.21)'de KKT koşulu olarak belirtilen $\sum_{i=1}^n \alpha_i y_i = 0$ ile aşağıdaki ifade elde edilir.

$$\sum_{i=1}^3 \alpha_i y_i = (1)\alpha_1 + (1)\alpha_2 + (-1)\alpha_3 = 0$$

$$\alpha_1 + \alpha_2 = \alpha_3$$

Elde edilen bu değer Lagrange fonksiyonunda yerine yazılmasıyla aşağıdaki ifade bulunur.

$$L_D(\alpha) = 2\alpha_3 - \frac{1}{2}(\alpha_2^2 - 2\alpha_2\alpha_3 + 2\alpha_3^2)$$

α_2 ve α_3 Lagrange çarpanlarının değerini bulmak için Lagrange fonksiyonunun çarpanlara kısmi türevi alınarak sıfıra eşitlenir.

$$\frac{\partial L(\alpha)}{\partial \alpha_2} = -2\alpha_2 + 2\alpha_3 = 0$$

$$\frac{\partial L(\alpha)}{\partial \alpha_3} = 2 + \alpha_2 - 2\alpha_3 = 0$$

Bu eşitliklerin çözülmesiyle $\alpha_1 = 0$, $\alpha_2 = 2$, $\alpha_3 = 2$ bulunur. $\alpha_2, \alpha_3 \geq 0$ olduğundan x_2 ve x_3 gözlemleri birer destek vektördür.

Denklem (6.29) kullanılarak w^* aşağıdaki gibi hesaplanır.

$$\begin{aligned} w^* &= \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \alpha_3 y_3 x_3 \\ &= 0(1) \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 2(1) \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 2(-1) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ 0 \end{bmatrix} \end{aligned}$$

Denklem (6.32)'in kullanılmasıyla da b^* aşağıdaki gibi hesaplanır.

$$\begin{aligned} b^* &= y_{sv} - w^* \cdot x_{sv} \\ &= 1 - \begin{bmatrix} -2 \\ 0 \end{bmatrix} \cdot [0 \quad 1] \\ &= 1 \end{aligned}$$

Son olarak karar sınırını belirleyecek olan fonksiyon Denklem (6.33) kullanılarak aşağıdaki gibi bulunur.

$$\begin{aligned} f(x) &= \text{sgn}(w^* \cdot x + b^*) \\ f(x) &= \text{sgn}\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix} \cdot x + 1\right) \end{aligned}$$

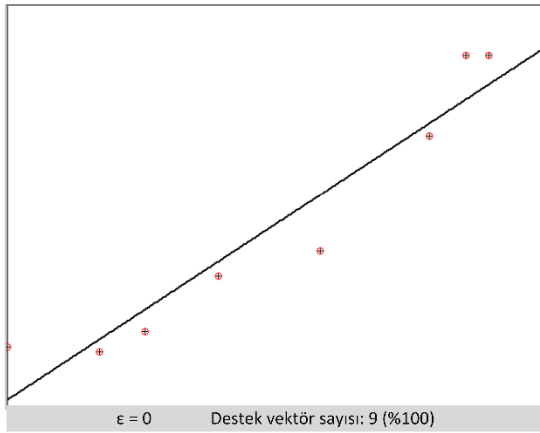
Örnek olarak $x_4 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ gözlemi için sınıf şöyle belirlenir.

$$\begin{aligned} f(x) &= \text{sgn}\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix} \cdot [2 \quad 3] + 1\right) \\ &= \text{sgn}(-3) = -1 \end{aligned}$$

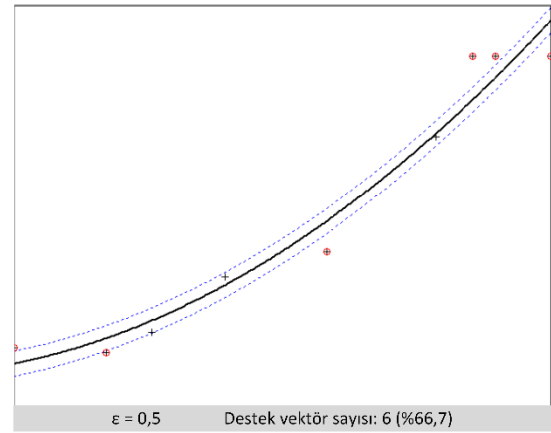
Sonuç olarak x_4 gözlem değerinin -1 sınıfına ait olduğu bulunmuştur.

,

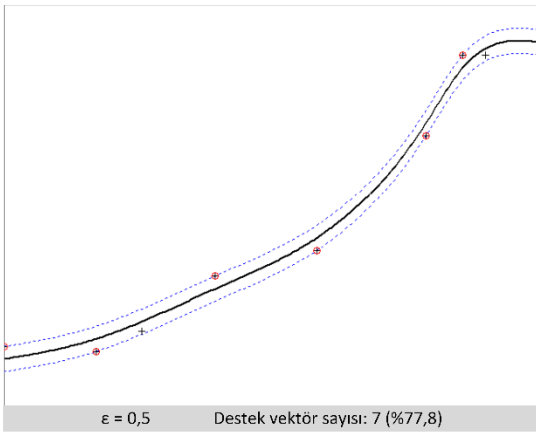
Ek Açıklama-C: Destek Vektör Regresyon İçin Açıklayıcı Bir Örnek



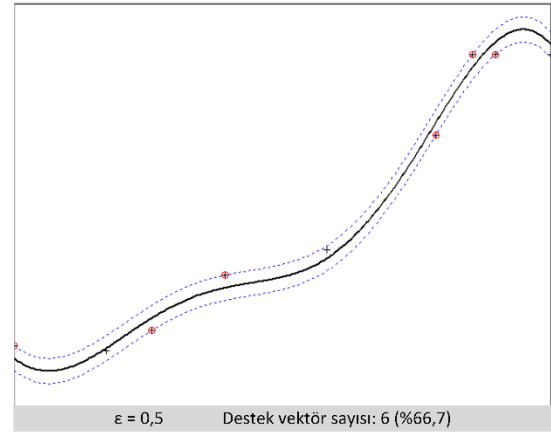
a) Doğrusal



b) Polinom Çekirdek



b) Spline Çekirdek



d) RBF Çekirdek

+ Gözlem Değerleri

⊕ Destek Vektörleri

— Regresyon Fonksiyonu

- - - ε Tüpü (Yardımcı hiper düzlemler)

Ek Açıklama-D: Örnek Veri Seti

Row No.	X1_SO2	X_2PM10	X_3_TEMP	X_4MAXTEMP	X_5MINTEMP	X_6DEWP	X_7SLP	X_8STP	X_9VISIB	X_10WDSP	X_11MXSPD	X_12PRCP	Y_SO2
1	62	144	3	6	1	-6	1020.800	968.500	18.830	2.110	3.600	0	47
2	47	74	1	5	-2	-8	1025.600	972.800	19.960	2.370	3.600	0	18
3	18	54	2	6	-1	-7	1029.900	976.900	19.960	1.900	3.090	0	41
4	41	73	1	5	-3	-9	1030.500	977.400	19.960	1.080	2.110	0	16
5	16	42	1	6	-3	-9	1027.100	974.200	19.960	0.930	1.490	0	10
6	10	45	1	6	-3	-8	1024.200	971.500	18.510	1.180	2.110	0	17
7	17	32	2	6	-3	-7	1022.600	970	17.540	0.980	1.490	0	7
8	7	22	7	13	-3	-4	1016.200	964.500	19.960	2.320	5.710	0	22
9	22	77	12	16	9	4	1012	961.900	16.900	2.730	4.120	5.080	5
10	5	184	14	17	11	2	1009.400	959.600	19.960	2.680	3.600	0	6
11	6	73	13	17	9	1	1007.800	958.200	19.960	3.240	6.690	3.050	19
12	19	55	11	15	6	3	1012.200	961.900	19.960	2.110	3.600	2.030	13
13	13	33	8	10	6	5	1012	961.300	13.840	0.670	0.980	45.970	12
14	12	34	9	12	7	4	1011.500	960.800	18.830	0.870	1.490	7.370	8
15	8	31	8	13	5	3	1011.600	960.900	18.510	1.180	2.620	0.760	44
16	44	30	10	13	7	3	1015.300	964.500	19.960	1.030	2.620	0	44
17	44	27	10	14	7	4	1018.900	968.200	19.960	1.180	2.110	0.250	7
18	7	68	11	16	8	4	1016.400	966	19.960	1.180	2.620	0	34
19	34	37	11	15	7	6	1016.900	966.300	19.960	0.820	1.490	0	78
20	78	75	11	15	7	3	1016.800	966.200	19.960	0.980	1.490	0	3

Ek Açıklama-E: Örnek Normalleştirilmiş Veri Seti

Row No.	X1_SO2	X_2PM10	X_3_TEMP	X_4MAXTEMP	X_5MINTEMP	X_6DEWP	X_7SLP	X_8STP	X_9WSIB	X_10WDSP	X_11MXSPD	X_12PRCP	Y_SO2
1	0.195	0.157	0.214	0.191	0.308	0.235	0.638	0.609	0.849	0.417	0.300	0	0.147
2	0.147	0.078	0.167	0.170	0.231	0.176	0.757	0.727	0.916	0.470	0.300	0	0.054
3	0.054	0.056	0.190	0.191	0.256	0.206	0.864	0.839	0.916	0.374	0.251	0	0.128
4	0.128	0.077	0.167	0.170	0.205	0.147	0.878	0.852	0.916	0.209	0.155	0	0.048
5	0.048	0.043	0.167	0.191	0.205	0.147	0.794	0.765	0.916	0.178	0.095	0	0.029
6	0.029	0.046	0.167	0.191	0.205	0.176	0.722	0.691	0.830	0.229	0.155	0	0.051
7	0.051	0.031	0.190	0.191	0.205	0.206	0.682	0.650	0.774	0.188	0.095	0	0.019
8	0.019	0.020	0.310	0.340	0.205	0.294	0.524	0.500	0.916	0.460	0.505	0	0.067
9	0.067	0.082	0.429	0.404	0.513	0.529	0.419	0.429	0.736	0.543	0.351	0.065	0.013
10	0.013	0.202	0.476	0.426	0.564	0.471	0.355	0.366	0.916	0.532	0.300	0	0.016
11	0.016	0.077	0.452	0.426	0.513	0.441	0.315	0.328	0.916	0.646	0.601	0.039	0.058
12	0.058	0.057	0.405	0.383	0.436	0.500	0.424	0.429	0.916	0.417	0.300	0.026	0.038
13	0.038	0.033	0.333	0.277	0.436	0.559	0.419	0.413	0.557	0.126	0.046	0.591	0.035
14	0.035	0.034	0.357	0.319	0.462	0.529	0.407	0.399	0.849	0.166	0.095	0.095	0.022
15	0.022	0.030	0.333	0.340	0.410	0.500	0.409	0.402	0.830	0.229	0.205	0.010	0.137
16	0.137	0.029	0.381	0.340	0.462	0.500	0.501	0.500	0.916	0.198	0.205	0	0.137
17	0.137	0.026	0.381	0.362	0.462	0.529	0.591	0.601	0.916	0.229	0.155	0.003	0.019
18	0.019	0.072	0.405	0.404	0.487	0.529	0.529	0.541	0.916	0.229	0.205	0	0.105
19	0.105	0.037	0.405	0.383	0.462	0.588	0.541	0.549	0.916	0.156	0.095	0	0.246
20	0.246	0.080	0.405	0.383	0.462	0.500	0.538	0.546	0.916	0.188	0.095	0	0.006

Ek Açıklama-F: Çok Katmanlı Algılayıcı Mimarisi

