

**DESTEK VEKTÖR MAKİNELERİ İLE  
DOKÜMAN SINIFLANDIRMA  
YÜKSEK LİSANS TEZİ**

Üzeyir FİDAN

DANIŞMAN

Yard. Doç. Dr. Engin TAŞ

İSTATİSTİK ANABİLİM DALI

Haziran, 2013

AFYON KOCATEPE ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

DESTEK VEKTÖR MAKİNELERİ İLE  
DOKÜMAN SINIFLANDIRMA

Üzeyir FİDAN

Danışman

Yard. Doç. Dr. Engin TAŞ

İSTATİSTİK ANABİLİM DALI

Haziran 2013

## TEZ ONAY SAYFASI

Üzeyir FİDAN tarafından hazırlanan “**Destek Vektör Makineleri ile Doküman Sınıflandırma**” adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 27/06/2013 tarihinde aşağıdaki jüri tarafından oy birliği/~~oy çokluğu~~ ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **İstatistik Anabilim Dalı’nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

**Danışman** : Yard. Doç. Dr. Engin TAŞ

İmza

**Başkan** : Prof. Dr. Memmedağa MEMMEDLİ

Eskişehir Anadolu Üniversitesi, Fen Fakültesi

**Üye** : Yard. Doç. Dr. Engin TAŞ

Afyon Kocatepe Üniversitesi, Fen Edebiyat Fakültesi

**Üye** : Yard. Doç. Dr. Fatih Onur HOCAOĞLU

Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü Yönetim Kurulu’nun

...../...../2013 tarih ve

..... sayılı kararıyla onaylanmıştır.

Enstitü Müdürü

Prof. Dr. Mevlüt DOĞAN

# BİLİMSEL ETİK BİLDİRİM SAYFASI

## Afyon Kocatepe Üniversitesi

**Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;**

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

**beyan ederim.**

27.06.2013

Üzeyir FİDAN

# ÖZET

Yüksek Lisans Tezi  
DESTEK VEKTÖR MAKİNELERİ İLE  
DOKÜMAN SINIFLANDIRMA

Üzeyir FİDAN

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

**Danışman:** Yard. Doç. Dr. Engin TAŞ

Bu tezde, çevrimiçi Destek Vektör Makineleri (DVM) kullanılarak doküman sınıflandırma çalışması yapılmıştır. Lasvm algoritması eşli çekirdek yöntemi ile çalışılacak şekilde adapte edilmiştir. İlk olarak, seçilen çekirdek fonksiyonun etkileri ve parametreler belirlenmiştir. Bunun için doğrusal bir karar sınırı ile sınıflanacak dokümanlar yüksek boyutlu bir uzaya gömülmüştür. Bu nedenle veri gömme prosedürü ve çekirdek hilesi detayları örneklerle açıklanmıştır. Optimal hiper parametreler belirlendikten sonra eşli eğitim ve test verilerinin farklı kombinasyonları ile deneyler yapılmıştır. DVM modellerinin performansları doğru sınıflandırma oranı ve ROC eğrisi altında kalan alan kriterleri ile değerlendirilmiştir. Sonuçlar, çevrimiçi eşli sınıflandırmanın ikili ve çok sınıflı, sınıflandırma işlemlerine iyi bir alternatif metot olduğunu göstermiştir. Bu doküman sınıflandırma işleminde, eldeki verilerin yüksek boyutlu olması nedeniyle doğrusal eşli çekirdekler, gauss eşli çekirdeklerine göre daha iyi sonuçlar vermiştir.

**2013, viii+60 sayfa**

**Anahtar Kelimeler:** Destek Vektör Makineleri, Eşli Çekirdek, Doküman Sınıflandırma, Çekirdek.

# ABSTRACT

M.Sc. Thesis

DOCUMENT CLASSIFICATION  
WITH SUPPORT VECTOR MACHINE

Üzeyir FİDAN

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Statistics

**Supervisor:** Assist. Prof. Dr. Engin TAŞ

In this thesis, document classification task is studied using an online Support Vector Machine. Lasvm algorithm is adapted as to work with pairwise kernels. At first, the effects of the choice of the kernel function and its parameters are considered. In order to classify documents with a linear decision bound, the data is mapped into a higher-dimensional space. Therefore the data mapping procedure and the kernel trick are explained in detail with several examples. After the determination of optimal hyperparameters, experiments are conducted on different combinations of pairwise training and testing data. Performances of the SVM models are compared according to the classification accuracy and area under the ROC curve. Results indicate that online pairwise classification is a good alternative to the methods used in binary and multi-class classification tasks. In this document classification task, linear pairwise kernels achieve better results than the gaussian pairwise kernels because of the high dimensionality of the data at hand.

**2013, viii+60 pages**

**Key Words:** Support Vector Machine, Pairwise, Document Classification, Kernel.

## TEŐEKKÜR

Akademik duruşu, bilgi birikimi ve alıŐkanlıđı ile bana örnek olan, bilimsel alıŐmalara olan ilgimin artmasını sađlayan, yüksek lisansa baŐlamama cesaret veren, bu alıŐmanın her sũrecinde tũm bilgilerini ve tecrũbelerini benimle paylaŐmaktan kaınmayan danıŐman hocam, sayın Yard.Do.Dr.Engin TAŐ'a sonsuz minnet ve teŐekkũrlerimi sunarım.

Yüksek lisans alıŐmamın her aŐamasında bana büyük bir sevgi ve hoŐgörũ ile yaklaşan, her an bıkmadan sabırla yanımda olan sevgili eŐim Melike Fidan'a,

Bu alıŐmanın tũm detaylarını incelerken bana desteklerini büyük bir titizlikle gösteren deđerli arkadaşlarım, Arif Emre Turan ve Ali Alaca'ya,

Bu günlere gelmemde büyük pay sahibi olan, maddi ve manevi desteklerini benden esirgemeyen aileme sonsuz teŐekkũrlerimi sunarım.

Üzeyir FİDAN  
AFYONKARAHİSAR, 2013

# İÇİNDEKİLER DİZİNİ

	Sayfa
ÖZET . . . . .	i
ABSTRACT . . . . .	ii
TEŞEKKÜR . . . . .	iii
İÇİNDEKİLER DİZİNİ . . . . .	iv
SİMGELER DİZİNİ . . . . .	vi
ŞEKİLLER DİZİNİ . . . . .	vii
ÇİZELGELER DİZİNİ . . . . .	viii
<b>1 GİRİŞ . . . . .</b>	<b>1</b>
<b>2 TEMEL KAVRAMLAR . . . . .</b>	<b>3</b>
2.1 Temel Kavramlar ve Tanımlar . . . . .	3
<b>3 ÇEKİRDEK FONKSİYONLAR . . . . .</b>	<b>8</b>
3.1 Özellik Uzayında İç Çarpımlar . . . . .	8
3.2 Çekirdek . . . . .	9
3.2.1 En Sık Kullanılan Çekirdek Türleri . . . . .	9
3.2.2 Çekirdek ve Parametrelerin Seçimi . . . . .	10
3.2.3 Eşli (Pairwise) Çekirdek . . . . .	13
3.3 Çekirdek Hilesi . . . . .	14
<b>4 DESTEK VEKTÖR MAKİNESİ . . . . .</b>	<b>17</b>
4.1 İstatistiksel Öğrenme Teorisi . . . . .	17
4.2 Veri Temsili ve Benzerlik . . . . .	18
4.3 Destek Vektör Makinesi . . . . .	19
4.4 Veri Gömme . . . . .	23
4.5 Basit Bir Örüntü Tanıma Algoritması . . . . .	26
4.6 Vapnik-Chervonenkis Boyut ve VC Sınır . . . . .	28
4.6.1 VC Boyut . . . . .	28
4.6.2 Parçalama . . . . .	30
4.6.3 VC Sınırı . . . . .	30
4.7 Kayıp Fonksiyonu . . . . .	30
4.8 Hiperdüzlem Sınıflayıcıları . . . . .	31



4.8.1	Karush-Kuhn-Tucker (KKT) Şartları . . . . .	32
4.8.2	Esnek Marj Sınıflandırması . . . . .	34
4.9	Lasvm Algoritması ve Uygulama Şeması . . . . .	35
4.9.1	Çevrimiçi Lasvm . . . . .	37
4.9.2	Uygulama Şeması . . . . .	40
<b>5</b>	<b>UYGULAMA ve SONUÇLARI . . . . .</b>	<b>41</b>
5.1	Doküman Sınıflandırma Problemi ve Yapılan Çalışmalar . . . . .	41
5.2	Verilerin Elde Edilmesi . . . . .	42
5.2.1	20 News Groups Haber Veri Seti . . . . .	42
5.3	Problemin Eşli Çekirdek Yöntemine Uyarlanması . . . . .	43
5.4	Uygulama . . . . .	45
5.4.1	Çeşitli Çekirdek Fonksiyonları ile Uygulama ve Sonuçları . . . . .	45
5.5	Sonuçlar ve Öneriler . . . . .	54
<b>6</b>	<b>KAYNAKLAR . . . . .</b>	<b>57</b>
	<b>ÖZGEÇMİŞ</b>	

## SİMGELER DİZİNİ

### Simgeler

---

$\kappa$	Çekirdek
$\mathbb{R}$	Reel Sayılar Kümesi
$\mathbb{C}$	Karmaşık Sayılar Kümesi
$\tau^*$	Lasvm parametresi
$\mathcal{V}$	Vektör Uzayı
$\mathbb{K}$	Cisim
$\mathcal{X}$	Veri Uzayı
$\mathcal{H}$	Hilbert Uzayı
$\varphi$	Veri Gömme Fonksiyonu
$sgn$	Signum (İşaret) Fonksiyonu
$K_{ij}$	Gram Matrisi
$\gamma$	Çekirdek Parametresi
$\sigma$	Çekirdek Parametresi
$C$	Ceza Parametresi
$R[f]$	Beklenen Hata
$R_{emp}[f]$	Deneysel Hata
$\phi$	Kapasite Sınırı
$w$	Destek Vektörü
$\tau$	Amaç Fonksiyonu
$\alpha$	Lagrange Çarpanları
$\partial$	1.Mertebeden Türev Fonksiyonu
$\xi_i$	Esneklik Katsayısı
$\rho$	Marj Parametresi
$b$	Eşik Değeri
$[P]_{s,r}$	Eşli Çekirdek Verileri

---

## ŞEKİLLER DİZİNİ

	Sayfa
Şekil 2.1	Belirlenen bir orjine göre oluşan konum vektörleri. . . . . 7
Şekil 3.1	$C = 100$ ve $C = 10$ olan durumlarda sınıflandırmanın görünümü. 10
Şekil 3.2	Polinomal çekirdeğin derecesi ile karar sınırının eğriliği arasındaki ilişki. . . . . 11
Şekil 3.3	$\gamma$ parametresinin değişimi ile karar sınırları arasındaki ilişki. . . . . 11
Şekil 3.4	Gauss Radyal çekirdeğin farklı parametre değerleri için oluşan sınıflandırmaları. . . . . 12
Şekil 3.5	$\mathbb{R}^3$ de doğrusal olarak iki sınıfa ayrılamayan noktalar kümesi. . . . . 15
Şekil 4.1	İstatistiksel öğrenme teorisi için sınıflandırma şekilleri. . . . . 17
Şekil 4.2	$\mathbb{R}^2$ de doğrusal sınıflandırılabilen veri noktaları. . . . . 19
Şekil 4.3	$\mathbb{R}^2$ de iki farklı karar sınırı . . . . . 20
Şekil 4.4	Bir polinomal çekirdek fonksiyonun olası sınıflandırmaları. . . . . 20
Şekil 4.5	Aynı iki sınıfın farklı şekillerde sınıflandırılmaları. . . . . 21
Şekil 4.6	Marj kavramı ve optimal sınıflandırma. . . . . 21
Şekil 4.7	DVM'nin geometrik yorumu. . . . . 22
Şekil 4.8	$\mathbb{R}^2$ de doğrusal sınıflandırılmayan nokta kümesi. . . . . 23
Şekil 4.9	Şekil (4.8) deki noktaların $\mathbb{R}$ den $\mathbb{R}^2$ ye gömülmüş hali. . . . . 24
Şekil 4.10	Şekil (4.9) deki $\mathbb{R}^2$ ye gömülü verilerin sınıflandırılmış hali. . . . . 24
Şekil 4.11	Veri noktalarının $\mathbb{R}^2$ uzayında doğrusal ayrılamama durumu. . . . . 25
Şekil 4.12	(4.11) noktaların $\mathbb{R}^3$ uzayına gömülmesi. . . . . 25
Şekil 4.13	(4.12) noktaların $\mathbb{R}^3$ uzayında profil görüntüsü. . . . . 26
Şekil 4.14	(4.5.1) ve (4.5.2) denklemlerinin geometrik yorumu. . . . . 27
Şekil 4.15	$\mathbb{R}^2$ de 3 noktanın iki sınıfa ayrılabilceği muhtemel tüm durumlar. 29
Şekil 4.16	$\mathbb{R}^2$ de 4 noktanın iki sınıfa ayrılamadığı durum görüntüsü. . . . . 29
Şekil 4.17	Optimal hiper düzlem ve marj. . . . . 33
Şekil 4.18	Uygulama şeması. . . . . 40

## ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 5.1	Orjinal haber grupları tablosu . . . . . 43
Çizelge 5.2	Eşli Çekirdek örnek tablosu . . . . . 44
Çizelge 5.3	Haber grupları ve kullanılacak kodları . . . . . 45
Çizelge 5.4	Haber gruplarının tüm olası kombinasyonları. . . . . 47
Çizelge 5.5	$\varepsilon$ parametresindeki değişimler. . . . . 49
Çizelge 5.6	$C$ parametresindeki değişimler. . . . . 50
Çizelge 5.7	50000 test dokümanı için deney sonuçları. . . . . 51
Çizelge 5.8	50000 eğitim dokümanı için deney sonuçları. . . . . 52
Çizelge 5.9	Eğitim ve test kümeleri aynı . . . . . 53
Çizelge 5.10	Eğitim ve test kümeleri farklı . . . . . 53
Çizelge 5.11	Eğitim ve test kümeleri aynı örneklem farklı . . . . . 53
Çizelge 5.12	Doğrusal çekirdek ile Gauss Radyal çekirdek performansları. . . 54

# 1 GİRİŞ

Günümüz akademik çalışmalarının tek bir disiplini ele alması kabul görmemektedir. Bilim dünyasında yaşanan hızlı gelişmeler farklı disiplinlerin bir arada çalışmasını gerektirmiştir. İstatistik bilimi burada temel rol oynamaktadır. İstatistik, insanoğlunun varoluşundan beri hayatın her alanında bir araç olmuştur fakat günümüzde istatistik araç olmaktan çok amaç haline gelmiştir. Bu sebeple istatistiğin kullanılması, bilim dallarının tartışmasız güvenilirliğini ve saygınlığını arttırmaktadır.

Teknolojik gelişmeler neticesinde istatistiğin en önemli alanlarından biri olan veri sınıflandırma ön plana çıkmıştır. Günümüz akademik ve teknolojik gelişmeleri, istatistiksel öğrenme ve veri sınıflandırmanın hayatımızı oldukça kolaylaştırdığını gün ışığına çıkarmaktadır. İstatistiksel öğrenme, teorik bir yaklaşım olduğundan ve uygulamada mühendislik çalışması gerektirdiğinden insan beyninin öğrenme yapısının modellenmesi işlemine çoğunlukla makine öğrenmesi denilmektedir. İnsan beyninin öğrenme, problem haline getirme ve sınıflandırma işlemlerini adım adım ele aldığımızda, varsayılan istatistiksel öğrenmenin esinlendiği noktayı gözlemleyebiliriz. Nitekim bilgisayarların düşünme ve sonuçlandırma işlemlerini insan beyninin işleyişinden bağımsız düşünmemiz mümkün değildir. İnsan beyni gözlemlerden oluşan verilerin sınıflandırılması ile bir sonuca ulaşır, bu sonucu bir model halinde öğrenir. Daha sonra karşılaştığı örnekleri bu model ile kıyaslayarak yeni sonuçlar elde eder. İstatistiksel öğrenme elde bulunan gözlemlerin sınıflandırılması ile bir model oluşturur ki, bu model beynin öğrenme prensibiyle büyük benzerlikler gösterir. Fakat günümüz teknolojisi ile hala beynimizin gösterdiği performansa yaklaşabilmiş durumda bile değiliz. Bilgisayarların insan beyninden daha güçlü olmadığını sadece daha hızlı işlem yapabildiğini unutmamamız gerekir. Bu netice ile bilgisayarları uygun algoritmalar ile çalıştırırsak çok daha hızlı bir öğrenmenin gerçekleştirilebilir olması gereklidir. Yıllarca süren akademik çalışmalarda amaç en düşük maliyet ile en hızlı ve yüksek doğruluk oranına sahip modellemenin gerçekleştirilebilmesidir.

Bu konudaki ilk çalışmalar 1936 yılında Fisher tarafından ilk örüntü tanıma algoritmasının yapılması ile gündeme gelmiştir(Fisher 1936). Örüntülerin öğrenilmesi ben-

zer problemler için bir karar modelinin oluşturulması anlamına geldiğinden istatistiksel öğrenmenin temelleri oluşmuştur. İstatistiksel öğrenme için birçok yöntemden bahsedilebilir fakat burada Destek Vektör Makinesi (DVM) üzerinde durulacaktır. DVM, öğrenme ve sınıflandırma için kullanılan bir eğitim algoritmasıdır. Bu algoritma ile yapılan eğitimin sonucunda oluşan model test edilir.

DVM ilk olarak V.Vapnik tarafından 1960 lı yıllarda ortaya atılmış olsa da, bu sürecin geliştirilmesi, uygulanabilmesi ve kabul görmesi 1995 yılına kadar sürmüştür. Yöntemin bu kadar uzun sürede kabul edilmesinin nedeni çok maliyetli olduğunun düşünülmesi olmuştur. V.Vapnik'e göre temel unsur eğitim kümesindeki verileri optimal bir hiperdüzlem ile iki guruba ayırabilmektir. Problem doğrusal sınıflandırma problemi olarak düşünülse de, doğrusal olarak sınıflandırılmayan veri kümeleri içinde geliştirilmiş bir algoritma yapısına sahiptir. Burada verilerin buldukları uzaydan daha yüksek boyutlu bir uzaya belli bir dönüşüm fonksiyonuyla aktarılması temel alınmıştır(Cortes and Vapnik 1995).

Bu tezde destek vektör makineleri ile doküman sınıflandırma problemi üzerine çalışılmıştır. Tez beş bölümden oluşmaktadır. Tezin ilk bölümü giriş için ayrılmıştır. İkinci bölümde bazı matematiksel ve istatistiksel ifadeler hakkında gerekli bilgiler verilmiştir. Üçüncü bölümde, çekirdek (kernel) fonksiyonu ve özellikleri anlatılmıştır. Dördüncü bölümde destek vektör makineleri, pairwise tekniği ve teorik alt yapısını oluşturan istatistiksel öğrenme teorisi üzerinde durulmuştur. Beşinci ve son bölümde “Doküman Sınıflandırma Problemi” hakkında bilgi verilip, uygun verilerle problem tasarlanmış ve deneysel çalışma anlatılmıştır. Sonuç olarak DVM'nin farklı çekirdek fonksiyonlar için teze konu olan probleme etkileri belirlenmiş ve benzer yapılacak çalışmalar için araştırmacılara kolaylık sağlamak amaçlı önerilere yer verilmiştir.

## 2 TEMEL KAVRAMLAR

### 2.1 Temel Kavramlar ve Tanımlar

**Tanım 2.1.1** (İç Çarpım):  $\mathcal{V}$ ,  $\mathbb{K}$  cismi üstünde bir vektör uzayı olsun.  $f : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{K}$  biçiminde,  $(u, v)$  deki değeri  $\langle u, v \rangle$  ile gösterilen ve aşağıdaki önermeleri doğrulayan bir  $f$  fonksiyonuna  $\mathcal{V}$  üstünde bir çarpım denir (Sabuncuoğlu 2004).

$$(i) \quad \forall u \in \mathcal{V}, \quad u \neq 0 \Rightarrow \langle u, u \rangle > 0$$

$$(ii) \quad \forall u, v \in \mathcal{V}, \quad \langle v, u \rangle = \overline{\langle u, v \rangle}$$

$$(iii) \quad \forall u, v, w \in \mathcal{V}, \quad \langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$$

$$(iv) \quad \forall \alpha \in \mathbb{K}, \forall u, v \in \mathcal{V}, \quad \langle \alpha u, v \rangle = \alpha \langle u, v \rangle$$

Tanım 2.1.1'te verilen  $\overline{\langle u, v \rangle}$  gösterimi,  $\langle u, v \rangle$  sayısının eşleniğini göstermektedir. Eğer  $\mathcal{V}$  reel vektör uzayı ise (ii) önermesi,  $\langle v, u \rangle = \langle u, v \rangle$  biçimine girer (Sabuncuoğlu 2004).

**Tanım 2.1.2:**  $\mathcal{V}$  vektör uzayı üstünde bir iç çarpım varsa bu vektör uzayına iç çarpım uzayı denir (Sabuncuoğlu 2004).

**Tanım 2.1.3:**  $\mathcal{V}$ ,  $\mathbb{K}$  cismi üstünde bir vektör uzayı olsun.

$$n : \mathcal{V} \rightarrow \mathbb{R}, \quad n : u \rightarrow \|u\|$$

biçimindeki  $n$  fonksiyonu aşağıdaki önermeleri sağlıyorsa, bu fonksiyona  $\mathcal{V}$  üstünde bir norm denir.  $\mathcal{V}$  vektör uzayına da, normlu vektör uzayı denir (Sabuncuoğlu 2004).

$$(i) \quad \forall u \in \mathcal{V}, \quad u \neq 0 \Rightarrow \|u\| > 0$$

$$(ii) \quad \forall c \in \mathbb{R}, \forall u \in \mathcal{V}, \quad \|cu\| = |c|\|u\|$$

$$(iii) \quad \forall u, v \in \mathcal{V}, \quad \|u + v\| \leq \|u\| + \|v\|$$

**Tanım 2.1.4:**  $\mathcal{V}$  doğrusal uzayında tanımlı toplama ve skaler ile çarpım işlemleri aşağıdaki özellikleri sağlar.  $\forall u, v, w \in \mathcal{V}$  ve  $\alpha, \beta \in \mathbb{R}$  olmak üzere,

- $u + v = v + u$

- $u + (v + w) = (u + v) + w$
- $u + 0 = 0 + u = u$
- $u + (-u) = 0$
- $\alpha(\beta u) = (\alpha\beta)u$
- $\alpha(u + v) = \alpha u + \alpha v$
- $(\alpha + \beta)u = \alpha u + \beta u$
- $1.(u) = u$
- $u.u^{-1}$

$\mathcal{V}$  doğrusal uzayında yukarıdaki özelliklerin yanı sıra Cauchy yakınsaklığı varsa  $\mathcal{V}$  uzayı Tam İç Çarpım Uzayı veya Hilbert Uzayı olarak isimlendirilir (Bierens et al. 2012).

Hilbert Uzayının tercih edilmesinin esas nedenlerinden biri de iç çarpımın tanımlanma şeklidir. Hilbert Uzayında iç çarpım;

$$\langle u, v \rangle = u^T G v \quad (2.1.1)$$

şeklinde ifade edilir. Bu şekildeki gösterimde simetrik ve pozitif tanımlı, kare  $G$  matrisi ile ifade edildiğinden Hilbert Uzay'ı matrislerle işlem yapma olanağı sağlar(Schölkopf 2002).

Hilbert Uzay'ına veri gömebilmek için bir transfer(gömme) fonksiyonu belirlemeliyiz.

**Tanım 2.1.5:**  $\mathcal{X}$  veri uzayı,  $\mathcal{H}$  Hilbert Uzayı olmak üzere,

$$\begin{aligned} \varphi : \mathcal{X} &\longrightarrow \mathcal{H} \\ \varphi : x &\longrightarrow \varphi(x) \end{aligned}$$

şeklinde verilen  $\varphi$  fonksiyonuna veri gömme fonksiyonu denir(Schölkopf 2002).

**Tanım 2.1.6:**  $\kappa : \mathcal{X}^2 \longrightarrow K$  belirli bir fonksiyon ( $K = \mathbb{R}$  veya  $K = \mathbb{C}$ ) ve girdiler  $x_1, x_2, \dots, x_m \in \mathcal{X}$  şeklindeki



$$K_{ij} = \kappa(x_i, x_j)$$

$m \times m$  tipindeki matrisse Gram (Çekirdek) Matrisi denir(Schölkopf 2002).

**Tanım 2.1.7:** Bir matris, eğer kompleks eşlenik transpozuna eşit ise, Hermit Matristir. Yani bir  $A$  matrisi için, eğer

$$A = A^H$$

ise  $A$  Hermit matristir(Jódar 1996, Villard 1996, Bellman et al. 1970).

**Tanım 2.1.8:** Bir  $A$ ,  $n \times n$  Hermit Matrisi, eğer her sıfırdan farklı  $n$  boyutlu  $x$  vektörü için,

$$\langle Ax, x \rangle > 0$$

şartı sağlanıyor ise pozitif tanımlı matris,

$$\langle Ax, x \rangle = 0$$

şartı sağlanıyor ise yarı pozitif tanımlı matris denir(Schölkopf 2002).

Aynı tipte iki pozitif tanımlı matrisin toplamı yine aynı tipte pozitif tanımlı bir matristir, pozitif tanımlı bir matrisin Hermit transpozu da pozitif tanımlıdır.

**Tanım 2.1.9:**  $\mathcal{X}$  boştan farklı bir küme,  $\forall m \in \mathbb{N}$  ve  $x_1, x_2, \dots, x_m \in \mathcal{X}$  olmak üzere  $\kappa$  fonksiyonu  $\mathcal{X} \times \mathcal{X}$  üzerinde tanımlı bir Gram Matris belirtiyorsa, bu matrisse pozitif tanımlı çekirdek denir. Genellikle basit olarak sadece çekirdek denir(Schölkopf 2002).

**Cauchy-Schwarz Teoremi 2.1.1:**  $\mathcal{V}$ ,  $\mathbb{K}$  cismi üzerinde tanımlanmış bir iç çarpım uzayı olsun. O takdirde  $x, y \in \mathcal{V}$  olmak üzere,

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

dir. Eşitlik durumunun var olması için gerek ve yeter şart,  $x$  ve  $y$  vektörlerinin lineer bağımlı olmasıdır (Buskes et al. 2000, Fujii, Masatoshi, et al. 1997, Callebaut, D. K. 1965).

**İspat:** Eğer  $x$  veya  $y$  sıfır vektörü ise, o zaman teoremin iddiası açıktır. Bundan dolayı  $x$  ve  $y$  vektörlerinin her ikisinde sıfırdan farklı kabul edilebilir.  $x = (x_1, x_2, \dots, x_n)$  ve  $y = (y_1, y_2, \dots, y_n)$  olsun. Diğer taraftan,  $(a - b)^2 \geq 0$  olduğundan,

$$a^2 - 2ab + b^2 \geq 0 \quad \text{yada} \quad a^2 + b^2 \geq 2ab \quad (2.1.2)$$

eşitsizliğini yazabiliriz.

$$a = \frac{|x_i|}{\|x\|} \quad \text{ve} \quad b = \frac{|y_i|}{\|y\|}$$

olarak alalım ve bunları 2.1.2 eşitsizliğinde yerine yazalım. O takdirde,

$$2 \frac{|x_i|}{\|x\|} \frac{|y_i|}{\|y\|} \leq \frac{|x_i|^2}{\|x\|^2} + \frac{|y_i|^2}{\|y\|^2} \quad (2.1.3)$$

eşitsizliğini elde ederiz. Şimdi 2.1.3 eşitsizliğinin her iki tarafını  $i = 1$  den  $n$  e kadar toplamı alırsak,

$$2 \sum_{i=1}^n \frac{|x_i|}{\|x\|} \frac{|y_i|}{\|y\|} \leq \sum_{i=1}^n \frac{|x_i|^2}{\|x\|^2} + \sum_{i=1}^n \frac{|y_i|^2}{\|y\|^2}$$

yazarız, halbuki burada

$$\sum_{i=1}^n \frac{|x_i|^2}{\|x\|^2} = 1 \quad \text{ve} \quad \sum_{i=1}^n \frac{|y_i|^2}{\|y\|^2} = 1$$

olduğundan eşitsizlik,

$$\sum_{i=1}^n |x_i| |y_i| \leq \|x\| \|y\| \quad (2.1.4)$$

durumuna gelir. Diğer taraftan

$$|\langle x, y \rangle| = \left| \sum_{i=1}^n x_i y_i \right| \leq \sum_{i=1}^n |x_i y_i| \quad (2.1.5)$$

olduğundan 2.1.4 ve 2.1.5 den,

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

elde ederiz böylece teoremin ilk kısmı ispat edilmiş olur. Eğer  $x$  ve  $y$  lineer bağımlı ise o zaman  $c \in \mathbb{K}$  olmak üzere  $x = cy$  yazabiliriz. Bu da bize eşitlik durumunun sağlandığını gösterir.

**Çekirdek Fonksiyonları için Cauchy-Schwarz Eşitsizliği 2.1.2:**  $\kappa$  pozitif tanımlı bir çekirdek ve  $x_1, x_2 \in \mathcal{X}$  olsun,

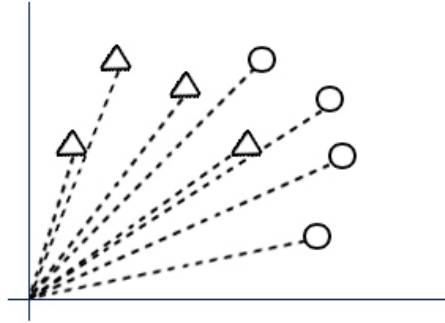
$$|\kappa(x_1, x_2)|^2 \leq \kappa(x_1, x_1)\kappa(x_2, x_2)$$

dir(Schölkopf 2002).

**İspat:**  $\kappa_{ij} = \kappa(x_i, x_j)$  girdilerin  $2 \times 2$  lik bir Gram matris olduğunu kabul edelim. Bu nedenle her ikisinin de öz değerleri negatif olmayacaktır.

$$0 \leq \kappa_{11}\kappa_{22} - \kappa_{12}\kappa_{21} = \kappa_{11}\kappa_{22} - \kappa_{12}\bar{\kappa}_{12} = \kappa_{11}\kappa_{22} - |\kappa_{12}|^2$$

**Tanım 2.1.10:** Destek Vektör Makinesi adını destek vektörleri denilen bir vektör ailesinden alır. Destek vektörleri belirlenmiş bir orjin noktasına göre sınıflandırılması planlanan noktaların konum vektörlerini temsil edebilecek niteliğe sahip en az sayıda oluşan vektörlerin ailesidir(Schölkopf 2002).



**Şekil 2.1:** Belirlenen bir orjine göre oluşan konum vektörleri.

Şekil (2.1) de açıkça görüldüğü gibi veriler, belirli bir noktaya göre oluşan konum vektörleri ile temsil edilebilir. Burada ki vektörler eğitim algoritması yardımıyla karar sınırını temsil edebilecek minimum sayıya düşürüldüğünde, geriye kalan vektörler destek vektörleri olarak adlandırılır. Bu destek vektörleri karar sınırlarını oluşturabilmek için gerekli olan parametrelerin elde edilmesini sağlayacaktır.

### 3 ÇEKİRDEK FONKSİYONLAR

Çekirdek, veri noktaları arasındaki benzerliği nokta çarpım fonksiyonu ile hesaplayıp, bu benzerlik ölçüsünden elde edilen değerleri veri noktalarının yeni kimliği olarak belirlemeye yarayan bir fonksiyondur. Çekirdek fonksiyonu tüm metrik aksiyomları sağlar. Uygulamada çoğu gerçek sınıflandırma problemi doğrusal olarak ayrılmamaktadır. Belirtilmesi gerekir ki doğrusal olmayan bir sınıflandırma fonksiyonu elbette 0 eğitim hatasını garanti etmez (Erastö 2001).

#### 3.1 Özellik Uzayında İç Çarpımlar

Bir transfer fonksiyonu veriyi özellik uzayına aktararak boyut arttırma işlemi gerçekleştirir. Verileri birer özellik vektörü temsili olarak alınırsa, bu özellik vektörleri arasındaki iç çarpım bir çekirdek matrisi halinde benzerlikleri depolamaktadır. Bu şekilde iç çarpımlar ile oluşturulan matrise Gram Matrisi denir. Gram Matrisi oluşturan vektörlerin elde edildiği iç çarpım bir metrik fonksiyonudur (Gunn 1998).

Destek vektörleri reel ve kompleks uzayda da tanımlanabilir, fakat sınıflandırma problemleri her zaman bu kısıtlı uzaylarda çözümlenemez. Bu sebeple problemler küçük boyutta sınıflandırılabilir olsa da, verileri özellik uzayına gömme işlemi bir çeşit genelleme olarak görülebilir. Verileri özellik uzayına aktardığımız fonksiyon  $\varphi$ , iç çarpım fonksiyonu da  $\kappa$  olmak üzere;

$$\kappa(x, x') = \langle x, x' \rangle = \langle \varphi(x), \varphi(x') \rangle \quad (3.1.1)$$

şeklinde yapılan işlem iç çarpımın özellik uzayına aktarılmadan, sanki özellik uzayında yapılmış gibi sonuç vermesini sağlar (Schölkopf 2002). Bu özellik sayesinde yapılan işlemin maliyeti artmadan kolaylıkla yapılabilmesini görürüz. Çekirdek fonksiyonun  $\varphi$  dönüşümü altında çalışabilmesi için aşağıdaki özellikleri sağlaması gereklidir.

Seçilen çekirdek fonksiyonu;

- Simetrik olmalıdır.
- Pozitif tanımlı olmalıdır.
- Cauchy-Schwarz Eşitsizliğini sağlamalıdır (Gunn 1998).

## 3.2 Çekirdek

Çekirdek tekniği doğrusal olarak ayrılamayan girdi uzayı için zorunlu olarak yardımına başvurulmuş bir yöntemdir (Herbrich 2002). Çekirdek fonksiyonu olarak adlandırılan fonksiyonların esas amacı veri noktalarının aralarındaki benzerliği belirleyip, bu sonuçlar yardımıyla sınıflandırmanın yapılmasını sağlamaktır.

Çekirdek fonksiyonu metrik özellikleri sağlayacak şekilde seçilir. İç çarpım fonksiyonunu ele alırsak iki veri noktasının tüm bileşenlerinin çarpımı ile elde edilen değer aslında bu iki veri noktasının birbirine benzerliğinin değeridir. Bu sayede çıkan sonuç ne kadar büyürse verilerin birbirine o büyüklükte benzediğini bulmuş oluruz.

### 3.2.1 En Sık Kullanılan Çekirdek Türleri

Literatür de birçok çekirdek türü kullanılmıştır. Fakat bunlardan en sık kullanılanları Doğrusal, Polinomal ve Gauss Radyal çekirdek çeşitleridir.

Doğrusal çekirdek,

$$\kappa(x, x') = \langle x, x' \rangle \quad (3.2.1)$$

eşitliği ile verilen fonksiyona denir.

Polinomal çekirdek,

$$\kappa(x, x') = \langle x, x' \rangle^d \quad (3.2.2)$$

eşitliği ile verilen fonksiyona denir. Bu eşitlikte  $d$  sayısı oluşan polinomun derecesini belirtmektedir. Polinomal çekirdek fonksiyonu,

$$\kappa(x, x') = (\langle x, x' \rangle + 1)^d \quad (3.2.3)$$

şekliyle tanımlanır.

Gauss Radyal çekirdek,

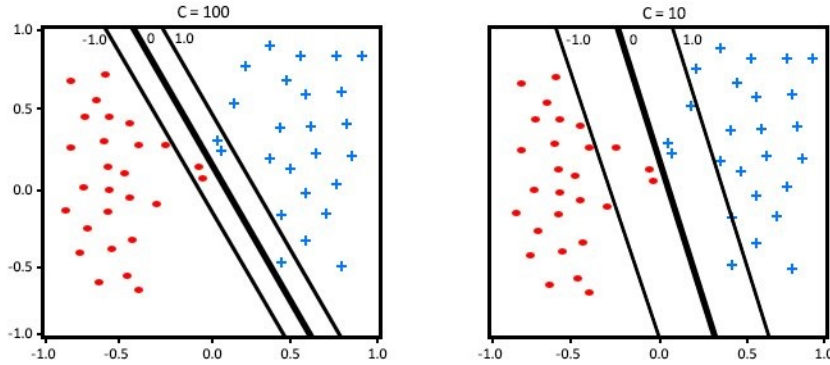
$$\kappa(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \quad (3.2.4)$$

eşitliği ile tanımlanır (Vapnik 2002, Schölkopf 2002, Erastö 2001, Cortes and Vapnik 1995).

### 3.2.2 Çekirdek ve Parametrelerin Seçimi

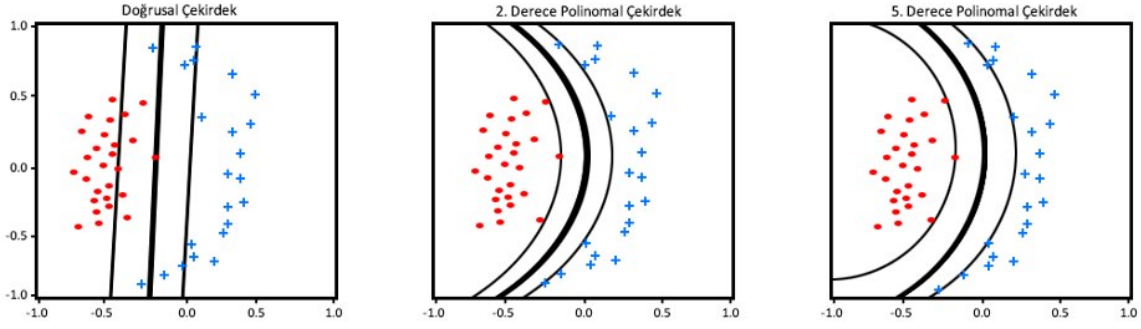
Destek Vektör Makinelerinde diğer bir önemli hususta çekirdek ve parametrelerin belirlenmesidir. Yumuşak marj parametresi  $C$  ve çekirdek ile ilgili DVM parametrelerini hiper parametreler olarak adlandırabiliriz. DVM de marj ile ilgili hataları cezalandırma puanı atamak için  $C$  parametresi kullanılır.  $C$  değeri çok düşük alındığında marj hatası oluşur. Bu sebeple verilerin geri kalanına büyük bir marj sağlamak için hiper düzlem pozisyonunu değiştirir. Eğer  $C$  parametre değeri çok büyük seçilirse çok yumuşak bir marj ile sınıflandırma yapılacağı için hata oranı artacaktır. Bu sebeple çekirdek parametrelerinin aynı zamanda karar sınırı üzerinde önemli bir etkiye sahip olduğu unutulmamalıdır. Fakat optimal bir  $C$  değerinin varlığından söz edilemez. Dolayısı ile  $C$  parametresi probleme uygun olarak belirlenmelidir.  $C$  parametresinin optimal değeri, ancak deneysel çalışmalar ile belirlenebilir.

$C$  parametresinin farklı değerleri için oluşan sınıflandırma Şekil(3.1) de verilmiştir.



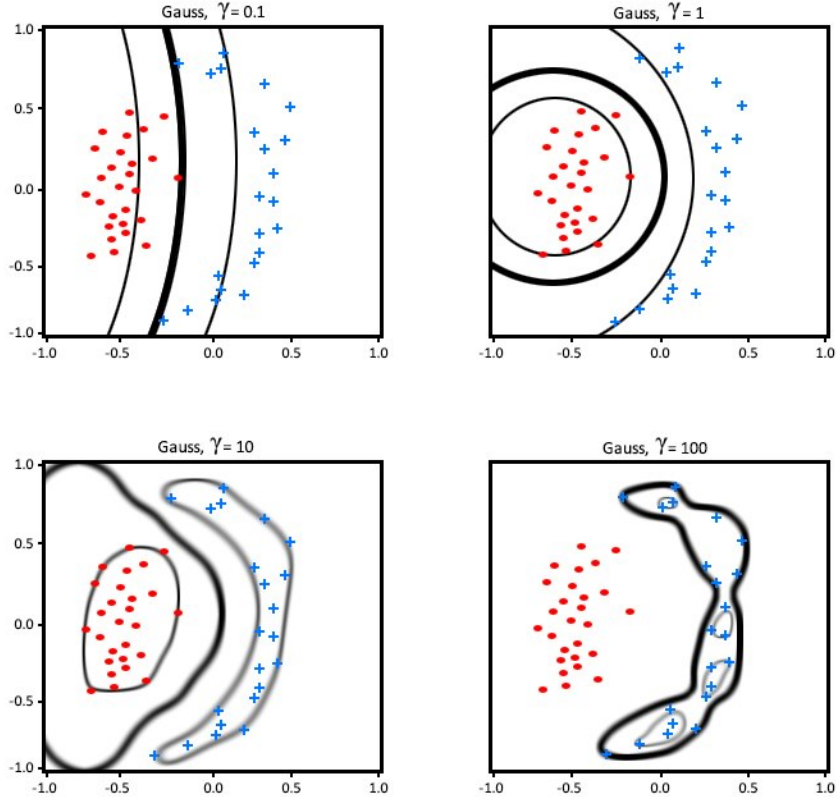
Şekil 3.1:  $C = 100$  ve  $C = 10$  olan durumlarda sınıflandırmanın görünümü.

Polinom çekirdeğin derecesi ve Gauss çekirdeğin parametreleri, elde edilen sınıflandırıcının esnekliğini kontrol eder. Özellikleri arasında doğrusal olmayan bir ilişki var olsa da derecesi düştükçe polinom fonksiyonu doğrusal fonksiyon özellikleri gösterir. 5 dereceli bir polinom daha çok eğrisellik gösterse de, 2 dereceli bir polinom fonksiyonu büyük bir esneklik ile iki sınıfa ayırma işlemini Şekil(3.2) gerçekleştirir. Gauss çekirdeğine dikkat edecek olursak,  $\kappa(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$  eğer  $x$  ve  $x'$  arasındaki mesafe  $\sigma$  değerinden



**Şekil 3.2:** Polinomal çekirdeğin derecesi ile karar sınırının eğriliği arasındaki ilişki.

çok büyükse, fonksiyonun değeri  $\infty$  a yakınsar. Öte yandan  $x$  ve  $x'$  arasındaki mesafe  $\sigma$  değerinden çok küçükse, fonksiyonun değeri sıfıra yakınsar. Bu sebeple  $\gamma = \frac{1}{\sigma}$  sayısı verilerle karar sınırı arasındaki uyumu belirleyen parametredir.  $\gamma$  daki değişimin etkisi Şekil(3.3) de verilmiştir.

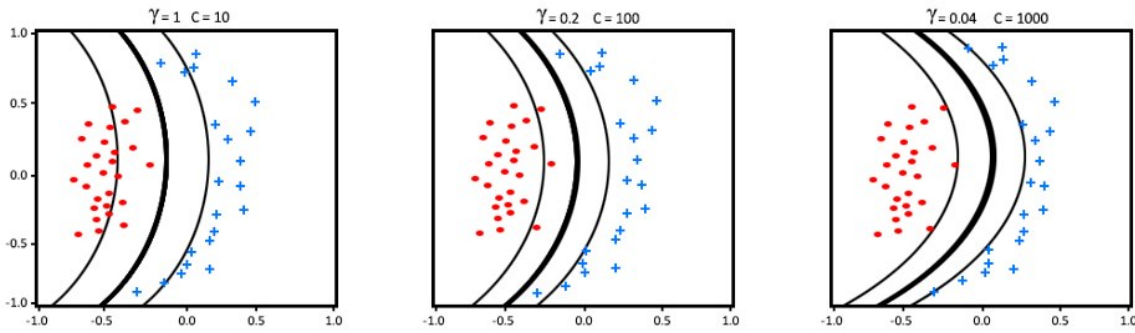


**Şekil 3.3:**  $\gamma$  parametresinin değişimi ile karar sınırları arasındaki ilişki.

Parametrelerle farklı sonuçlar elde edilebildiği gibi, çekirdekler ile de farklı sonuçlar elde etmek mümkündür. İlk defa çalışılan verilerle birkaç farklı çekirdek ile deney

düzenlenmeli ve sonuçlar değerlendirilmelidir. Yapılan çalışmalara göre, ilk olarak doğrusal çekirdek ile performans değerlendirilmeli sonrasında ise doğrusal olmayan çekirdeklere yönelmek gerekmektedir. Doğrusal çekirdek maliyet ve sonuçları bakımından yol gösterici bir rol üstlenir bunun yanında biyolojik çalışmalarda çok iyi sonuçlarda alınmıştır (Ben-Hur and Jason 2010).

DVM karar sınırının bağımlılığı, bağımlı sınıflandırıcının hiper parametrelerinin doğruluk oranları DVM sınıflandırıcının doğruluk oranlarına çevirir. Doğrusal bir sınıflandırıcı ile çalışırken belirlenmesi gereken tek hiper parametre DVM yumuşak marj sabitidir.



**Şekil 3.4:** Gauss Radyal çekirdeğin farklı parametre değerleri için oluşan sınıflandırmaları.

Polinom ve Gauss çekirdekleri için nokta arama alanı iki boyutludur. Veri noktalarını iki boyutlu uzayda keşfetmek için standart olarak ızgara arama yöntemi uygulanır. Izgara arama yöntemi, veri noktalarını genellikle logaritmik ölçekte seçer ve sınıflandırıcının doğruluğu ızgara üzerinde her bir nokta için tahmin edilir. Kısaca belirtmek gerekirse model seçimi doğrudan veri ile ilgilidir (Ben-Hur and Jason 2010).

Esas sorun farklı çekirdeklerin problemin çözümüne olan etkilerini belirleyip, en ideal çekirdek fonksiyonunu seçmektir. Bu yeni bir sorun değildir. Tek bir problem çerçevesi içinde çekirdekleri kıyaslamak daha doğru olacaktır. VC Boyut çekirdeklerin üst sınırlarını karşılaştıracak potansiyel bir yöntemdir. Bununla birlikte, doğrusal sınıflandırmamayan bir veri setini kapsayan kürenin çapının tahmin edilmesi gerekir. Son bir uyarı olarak, şu anda var olmayan bir en ideal çekirdek seçimi belirlenme metodu bulunsa bile, deneysel çalışmalar sonucu elde edilen optimal sonucun önüne geçemeyecektir (Gunn 1998).



### 3.2.3 Eşli (Pairwise) Çekirdek

İkili sınıflandırma problemleri için birçok öğrenme algoritması bulunmaktadır. Çok sınıflı problemler, ikili sınıflandırma problemlerini esas alarak geliştirilmiştir. En çok kullanılan yöntem, bire karşı hepsi yaklaşımıdır. Bu yaklaşım ikili sınıflandırma üzerine kuruludur. Başlangıç olarak pozitif bir sınıf belirlenir ve diğer tüm sınıflar negatif olarak sınıflanır. Sonrasında oluşan bu negatif sınıf verilerine tekrar aynı yöntem uygulanarak yeni bir pozitif sınıf elde edilir, tüm sınıflar ayrılıncaya kadar bu yönteme devam edilir. Sonuç olarak ikili sınıflandırmadan esinlenerek çok sınıflandırma gerçekleştirilmiş olur (Park and Johannes 2007).

Eşli sınıflandırma, bire karşı hepsi yaklaşımına alternatif bir yöntemdir. Aslında burada anlatılan temel fikir  $c$  sınıflı bir problemin,  $c(c-1)/2$  ikili sınıflandırma problemine dönüştürülmesidir. Eşli çekirdek tekniği destek vektör makinelerinde bire karşı hepsi tekniğine göre daha hızlı bir yöntemdir (Park and Johannes 2007).

Standart ikili sınıflandırma probleminin hedefi  $f : \mathcal{X} \rightarrow \{+1, -1\}$  fonksiyonunu öğrenmektir. Eşli çekirdek ise  $f : \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \rightarrow \{+1, -1\}$  fonksiyonunu öğrenmeye çalışır. Farz edelim ki, elemanları pozitifler için  $+1$  ve negatifler için  $-1$  elemanlarından oluşan bir  $f = |\mathcal{X}^{(1)}| \times |\mathcal{X}^{(2)}|$  matrisi olsun.  $x_{i_1}^{(1)} \in \mathcal{X}^{(1)}$  ve  $x_{i_2}^{(2)} \in \mathcal{X}^{(2)}$  olmak üzere  $x_{i_1}$  ile  $x_{i_2}$  arasındaki ilişki durumunu göstermek içinde  $[P]_{i_1, i_2}$  notasyonunu kullanırsak,  $x_{i_1}$  ve  $x_{i_2}$  arasında bağ varsa (aynı sınıftalar ise)  $[P]_{i_1, i_2} = +1$ , bağ yok ise (aynı sınıfta değiller ise)  $[P]_{i_1, i_2} = -1$  olur. Burada ikili örneklerin sınıflandırılması için ihtiyacımız olan, çiftler arasındaki ilişkiyi belirleyen çekirdek fonksiyonudur.  $(x_{i_1}^{(1)}, x_{i_2}^{(2)})$  ve  $(x_{j_1}^{(1)}, x_{j_2}^{(2)})$  gibi iki veri çifti arasındaki benzerliği tanımlayalım. Burada doğal olarak söylenebilir ki, iki veri çifti arasında benzerlik ilişkisi var ise her iki eleman içinde benzerlik ilişkisi vardır. Diğer bir deyişle  $x_{i_1}^{(1)}$  ve  $x_{j_1}^{(1)}$  birbirine benzer elemanlar ise,  $x_{i_2}^{(2)}$  ve  $x_{j_2}^{(2)}$  de birbirine benzerdir. Bu ilişki eşli çekirdek yöntemini tanımlamamız için bize yol gösterir. O zaman bu benzerliği matematiksel anlamda ifade etmek istersek,

$$\kappa((x_{i_1}^{(1)}, x_{i_2}^{(2)}), (x_{j_1}^{(1)}, x_{j_2}^{(2)})) = [\kappa^{(1)}]_{i_1, j_1} [\kappa^{(2)}]_{i_2, j_2} \quad (3.2.5)$$

elde ederiz. Burada  $\kappa^{(1)}$ ,  $\mathcal{X}^{(1)}$  veri setinin çekirdek matrisini ve  $\kappa^{(2)}$  de  $\mathcal{X}^{(2)}$  veri setinin çekirdek matrisini göstermektedir. Yukarıdaki eşitlikte verilen çekirdek Mercer Çekirdeği

olarak da adlandırılır. Bu çekirdeğin simetrik olması durumunda,  $(x_{i_1}, x_{i_2})$  ve  $(x_{j_1}, x_{j_2})$  eşli çekirdeği aşağıdaki gibi verilebilir.

$$\kappa^{SYM}((x_{i_1}, x_{i_2}), (x_{j_1}, x_{j_2})) = [\kappa]_{i_1, j_1} [\kappa]_{i_2, j_2} + [\kappa]_{i_1, j_2} [\kappa]_{i_2, j_1} \quad (3.2.6)$$

Böylece eşli çekirdeğin  $(x_{i_1}^{(1)}, x_{i_2}^{(2)}), (x_{j_1}^{(1)}, x_{j_2}^{(2)})$  çifti için tahmin fonksiyonu aşağıdaki şekilde elde edilir.

$$[P]_{i_1, i_2} = \sum_{j_1, j_2} \alpha(x_{j_1}^{(1)}, x_{j_2}^{(2)}) \kappa((x_{i_1}^{(1)}, x_{i_2}^{(2)}), (x_{j_1}^{(1)}, x_{j_2}^{(2)})) \quad (3.2.7)$$

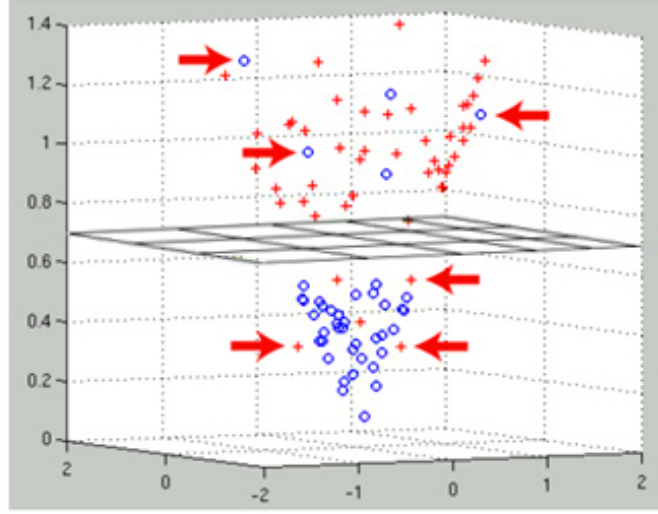
Burada  $\alpha$  lar modelin parametreleridir (Kashima et al. 2010, Kashima et al. 2009, Park and Johannes 2007, Vert, Qiu and Noble 2007, Trevor and Tibshirani 1998).

### 3.3 Çekirdek Hilesi

Her zaman mevcut veri seti doğrusal ayrılamayacaktır, dolayısı ile bu durumlarda verileri daha yüksek boyutlu bir özellik uzayına aktarmamız gerektiğinden bahsetmiştik. Fakat bu aktarma(gömme) işlemi bazen maliyeti çok yükseltebilir. Çekirdek hilesinin amacı, oluşan maliyetin minimize edilmesi anlamında sisteme büyük bir kazanç sağlamaktır.

$\mathbb{R}^3$  uzayında doğrusal sınıflandırılmayan bir girdi kümesinin daha yüksek boyutlu bir uzaya gömülmesi işleminde maliyetin azaltılmasını için çekirdek hilesini inceleyelim. Kabul edelim ki girdi verilerimiz  $\mathbb{R}^3$  uzayında Şekil(3.5)deki gibi bulunsun.

Matematiksel olarak çekirdek hilesini bir noktayı örnekleyerek açıklayalım. Veri setinden bir noktayı ele alalım.  $x_i = (1, 2, 3)$  olsun. Bu noktanın koordinatlarını bir satır vektörü olarak ele alıp, doğrudan dokuz boyutlu bir uzaya gömme işlemini aşağıdaki notasyonla tanımlayalım.



**Şekil 3.5:**  $\mathbb{R}^3$  de doğrusal olarak iki sınıfa ayrılamayan noktalar kümesi.

$$\varphi(x) = x^T \cdot x \quad (3.3.1)$$

Buradan,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot [1 \ 2 \ 3] = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

elde edilir ki oluşan sonucu (3.3.2) deki gibi yorumlanabilir.

$$\varphi(x) = [1 \ 2 \ 3] \rightarrow [1 \ 2 \ 3 \ 2 \ 4 \ 6 \ 3 \ 6 \ 9] \quad (3.3.2)$$

Veri noktasının iç çarpım fonksiyonu altındaki görüntüsü (3.3.3) deki gibi hesaplanabilir.

$$\kappa(x, x) = \langle x, x \rangle \quad (3.3.3)$$

$$\kappa(x, x) = 1.1 + 2.2 + 3.3 = 14 \quad (3.3.4)$$

Fakat bu uzayda doğrusal sınıflandırma mümkün olmadığından bu iç çarpım sonucu 3.3.4 sınıflandırmaya yaramayacaktır. Şimdi ise elde edilen yeni veri noktasının iç çarpım fonksiyonu altındaki görüntüsü hesaplanmalıdır.

$$\kappa(x, x) = \langle \varphi(x), \varphi(x) \rangle \quad (3.3.5)$$

$$\langle \varphi(x), \varphi(x) \rangle = 1^2 + 2^2 + 3^2 + 2^2 + 4^2 + 6^2 + 3^2 + 6^2 + 9^2 = 196 \quad (3.3.6)$$

Girdi verisi üç boyutlu bir uzayda iken olan maliyet 3.3.4 ile dokuz boyuta gömüldükten sonraki maliyet 3.3.6 arasında oluşan fark aşikardır. Bu sebeple  $\mathbb{R}^3$  de doğrusal sınıflandırılmayan verileri  $\mathbb{R}^9$  da doğrusal sınıflandırabilirsek bile maliyetin yüksek olması sistemin kullanılabilirliğini azaltmaktadır.

Çekirdek hilesi ile;

$$\langle x, x \rangle^2 = (1^2 + 2^2 + 3^2)^2 = 196 \quad (3.3.7)$$

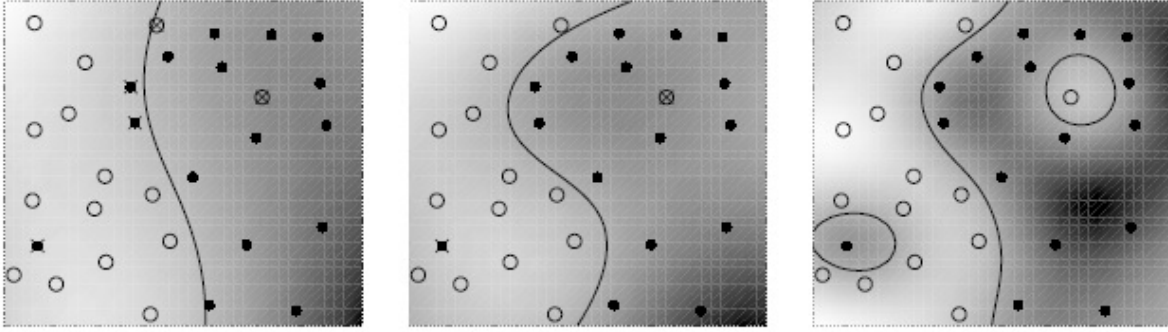
iç çarpım fonksiyonunu denklem 3.3.7 ile hesaplanıp aynı sonuca ulaşılabilir. Çekirdeklerin sonucu aynı olmasına karşın maliyet hala  $\mathbb{R}^3$  de uygulanan iç çarpım kadar olduğu görülür (Liu, et al. 2004, Schölkopf 2001).

## 4 DESTEK VEKTÖR MAKİNESİ

*”Hiçbir şey iyi bir teoriden daha pratik değildir.” (Vapnik 1998)*

### 4.1 İstatistiksel Öğrenme Teorisi

İstatistiksel öğrenme teorisi gözlemsel verilere dayanarak bir sınıflandırma fonksiyonu bulmayı amaçlamaktadır. Fakat tahmin fonksiyonu birden çok olabilir. Burada istatistiksel öğrenme teorisi en ideal sınıflandırmayı sağlayacak fonksiyonu tahmin etme amacındadır.



**Şekil 4.1:** İstatistiksel öğrenme teorisi için sınıflandırma şekilleri.

Şekil (4.1) de soldaki sınıflandırmada doğrusal bir sınıflandırma yapılmış gibi görülsede karar sınırına çok yakın hatalı noktalar oluşmuştur. Bu durumda sınıflandırmanın hata yüzdesi yüksek olacaktır. Sağdaki sınıflandırma çok hassaslaştırılmış ve sadece bu probleme özgü bir karar sınırı belirlenmiştir. Bu durum verilen problem için hatayı minimize etse de, genelleme yapmak mümkün değildir. Dolayısı ile elde edilen karar sınırı farklı bir test kümesi için uygulandığında aykırı noktalara gösterilen yüksek tolerans nedeniyle hata seviyesi oldukça yüksek olacaktır. Fakat ortadaki şekil incelendiğinde soldaki şekilde yapılan karar sınırına yakın hatalar engellenmiştir, böylelikle doğru sınıflandırma oranı yükseltilmiş olur. Sağ şekil de ki kadar da hassas bir sınıflandırma yapmayı kolaylaştırmıştır. Kısacası ortadaki şekil de istatistiksel öğrenme teorisine göre istenilen optimal şartlar yerine getirilmiş olur. Böylece istatistiksel öğrenme teorisinin amacını da açıklamış oluruz, yani amaç hatayı sıfır yapmak değil, minimize ederek daha genel bir model elde etmektir (Schölkopf 2002).

Sınıflandırmayı matematiksel olarak incelersek, elde edilen örnekler iki sınıfa ayrılmış

durumdayken yeni bir örnek ile karşı karşıya kaldığımızda bu iki sınıftan birine atamak, öğrenme teorisinin en temel problemlerinden biridir. Bu işlemler yapılırken girdi ve çıktı sınıflarını belirlemek yani etiketleme yapmak gereklidir.

$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad (4.1.1)$$

$\mathcal{X}$  boştan farklı bir küme,  $x_i$  ler de bu kümedeki gözlemler ve  $\{\pm 1\}$  kümesi de gözlemlerin çıktı etiketlerini temsil etmektedir. Çıktı sınıfının ikiye ayrıldığı bu gösterimde girdiler için ise bir sınırlama yoktur.

İstatistiksel Öğrenme Teorisi  $x_i$  leri, belirlenen  $y_i$  lerle eşleyerek sınıflandırmayı esas almaktadır. Bu yapıyı oluşturmak adına öncelikle öğrenme işlemi gerçekleştirilmek için öğrenme (eğitim) algoritması, sonra da performans değerlendirmek (test) için oluşacak olan model gereklidir. Burada eğitim algoritmasının eksiksiz ve tarafsız uygulanması çok büyük öneme sahiptir. Bu durumu anlamlandırabilmek için İstatistiksel Öğrenme Teorisinin, teorik yapısının çok iyi irdelenmesi gerekmektedir.

Schölkopf (2008) de, istatistiksel öğrenme teorisinin günümüzde makine eğitim algoritmaları üzerinde çok önemli bir yer edineceğine vurgu yapmıştır (Gretton 2008).

Bousquet (2004)e göre istatistiksel öğrenme teorisinin temel amacı; eldeki veri kümesinden bir karar modeli oluşturup bu model doğrultusunda karar verme çerçevesi oluşturmaktır. Öğrenme algoritmalarında performansı; öğrenme, genelleme ve çıkarımsama üzerine yoğunlaşarak değerlendirmiştir (Zhou 2004, Bousquet et al. 2004).

Evgeniou (2000) de, istatistiksel öğrenmenin bir paradigma olduğunu ve bu paradigmaya göre, eğitim yaparak gerçekleşen öğrenme, girdiler ile çıktılar arasındaki en ilişkili fonksiyonu bulmamızı sağlar demiştir (Evgeniou et al. 2000).

## 4.2 Veri Temsili ve Benzerlik

Deneysel veriler;

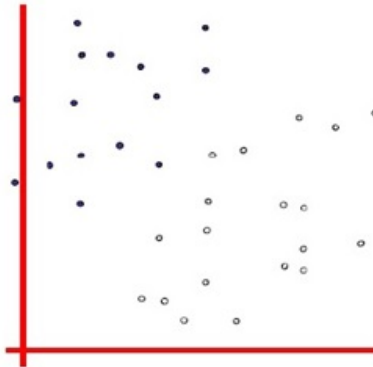
$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad (4.2.1)$$

şeklinde ifade edilebilir. Matematiksel kolaylık için çıktı sınıfı iki farklı küme olarak belirlendi fakat istenirse çıktı kümesi çok sayıda belirlenebilir. Bu gösterime binary(ikili) sınıflandırma denir.

Deneyisel verileri sınıflarken, elde edilen yeni  $x_i \in \chi$  ler için  $y_i \in \pm 1$  çıktılarını tahmin etmeliyiz. Aslında  $(x,y)$  eğitim girdilerini temsil edebilecek, yani bu girdiye benzer  $y$  çıktı sınıfını temsil edecek bir benzerlik ölçüsü belirlemeye çalışıyoruz. Benzerliği tanımlayabilmek için en çok tercih edilen benzerlik fonksiyonunu incelemeliyiz (Schölkopf 2002).

### 4.3 Destek Vektör Makinesi

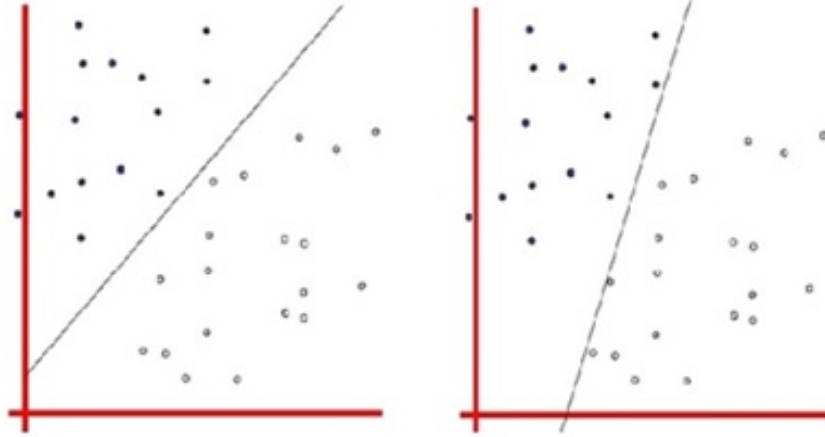
Destek vektörleri yardımı ile veri noktalarının sınıflandırılmasını sağlayan sisteme Destek Vektör Makinesi denir. Destek Vektör Makinesi nokta çarpımlarla benzerlikleri belirleyen bir çekirdek sınıflandırma yöntemidir. Bu durum bazı yüksek boyut özellikleri bir nokta çarpımla hesaplar ve çekirdek fonksiyona transfer etmemizi sağlar. Bu yöntem sağladığı iki avantaj ile ön plana çıkar, ilk olarak doğrusal sınıflandırma için tasarlanmış yöntemler kullanılarak doğrusal olmayan karar sınırları oluşturmamızı sağlar, ikinci olarak da çekirdek fonksiyonların kullanımı ile sabit boyutlu bir vektör uzayı temsili ile sınıflandırma yapmamızı sağlar (Ben-Hur and Jason 2010).



**Şekil 4.2:**  $\mathbb{R}^2$  de doğrusal sınıflandırılabilen veri noktaları.

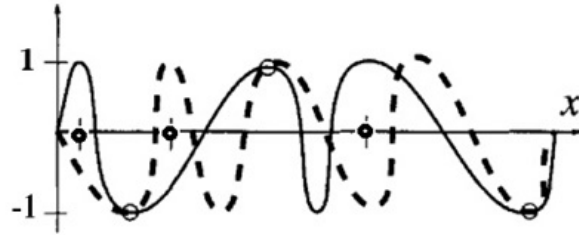
Şekil (4.2) de doğrusal olarak sınıflandırılabilen veri noktalarının düzlemdeki görüntüleri verilmiştir. Verilen girdi noktalarını bir doğru yardımıyla farklı şekillerde ayırmak

mümkündür. Burada adı geçen doğru, sınıflandırmayı sağlayan karar sınırını temsil etmektedir. Karar sınırı birden çok olduğundan optimal karar sınırını belirlemek gerekmektedir. Bu durumda asıl amaç sınıflandırmanın istatistiksel öğrenme teorisine uygun şartlarda sağlanmasıdır.



**Şekil 4.3:**  $\mathbb{R}^2$  de iki farklı karar sınırı

Şekil (4.3) de iki farklı karar sınırı yardımıyla aynı veri noktalarının sınıflandırıldığını görebiliriz. Bu durumdaki gibi farklı şekillerde sınıflandırma yapmak mümkündür. Şimdi de bir fonksiyon grafiği üzerinde aynı durumu inceleyelim.

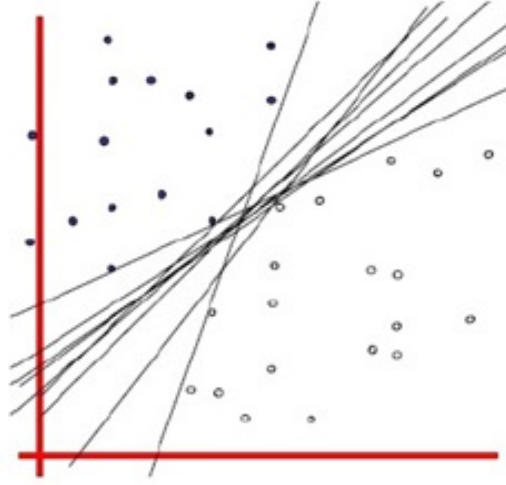


**Şekil 4.4:** Bir polinomal çekirdek fonksiyonunun olası sınıflandırmaları.

Şekil (4.4) de  $[0, \infty) \rightarrow [-1, 1]$  aralığında tanımlı bir polinomal sınıflandırma fonksiyonu verilmiştir. Fonksiyon eğrileri veri noktalarını farklı şekillerde de olsa doğru sınıflandırmıştır. Kısacası verilerin doğru sınıflandırılabilmesi için gerekli fonksiyon tek türlü olmak zorunda değildir. Daha genel bir sınıflandırma örneği aşağıda verilmiştir.

Şekil (4.5)de birden çok şekilde sınıflandırılmanın yapılabileceğini temsili olarak görmüş oluruz. Fakat burada karşımıza çıkan asıl problem optimal sınıflandırmayı sağlayanın

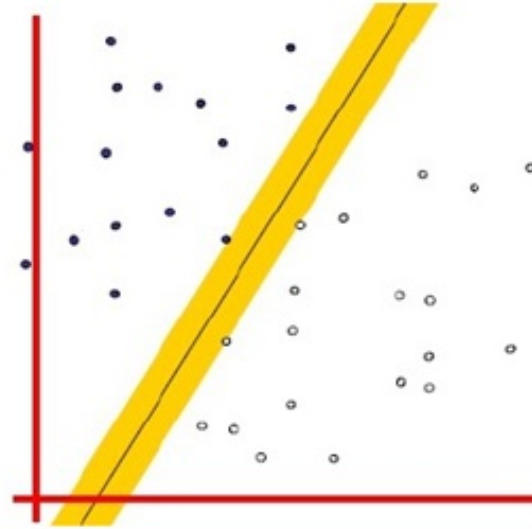




**Şekil 4.5:** Aynı iki sınıfın farklı şekillerde sınıflandırılmaları.

doğruların hangisi olduğudur. Destek Vektör Makinesi problemin optimal çözümünü aramakta ve geniş bir marj ile sınıflandırmayı amaçlamaktadır.

Marj kelimesi literatür karşılığı olarak kenar, boşluk gibi anlamlar içermektedir. Sınıflandırmayı tek bir doğru ile değil de, iki doğru ile yaparsak sınıfları birbirinden daha belirgin ayırmış oluruz. DVM için marj tanımı bu iki doğrunun arasında kalan bölge anlamındadır.

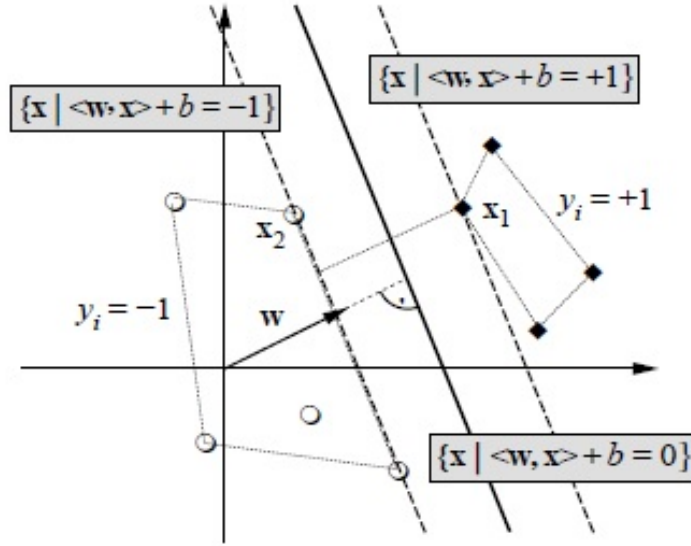


**Şekil 4.6:** Marj kavramı ve optimal sınıflandırma.

Şekil (4.6)de veri noktalarının iki sınıfa ayrıldığı bu doğru çiftine karar sınırları denir.

Her sınıfın karar sınırı o sınıfın veri noktalarını diğer veri noktalarından ayırmaya çalışır. Karar sınırlarına eşit uzaklıktaki sınıflandırmanın sağlandığı esas sınır, optimal hiperdüzlem veya hiperdüzlem sınıflayıcı adını alır. Optimal hiperdüzlem veri noktalarından doğrudan etkilenmesede, karar sınırlarına göre belirlendiğinden dolayı olarak etkilenmektedir (Schölkopf 2002).

Hiperdüzlem sınıflayıcının belirlenmesini, geometrik olarak Şekil (4.7) de inceleyebiliriz .



**Şekil 4.7:** DVM'nin geometrik yorumu.

Şekil (4.7) de oluşan karar sınırları ve marj matematiksel olarak aşağıdaki gibi ifade edilir.

$$\langle w, x_1 \rangle + b = +1 \quad (4.3.1)$$

$$\langle w, x_2 \rangle + b = -1 \quad (4.3.2)$$

$$\implies \langle w, (x_1 - x_2) \rangle = 2 \quad (4.3.3)$$

$$\implies \left\langle \frac{w}{\|w\|}, (x_1 - x_2) \right\rangle = \frac{2}{\|w\|} \quad (4.3.4)$$

## 4.4 Veri Gömme

Literatür de map olarak verilen veri gömme işlemi, doğrusal ayırlamama durumunda verilerin aralarındaki ilişkiye etki etmeden daha yüksek boyutlu bir uzaya aktarılması anlamı taşımaktadır.

Veri gömme işlemi bir  $\varphi$  dönüşüm fonksiyonu ile sağlanır. Bu fonksiyon genelde doğrusal değildir.  $\varphi$  fonksiyonu problemin gerçekleştiği uzaydan özellik uzayına verileri aktarır. Bu sayede doğrusal sınıflandırmanın olanaksız olduğu uzaylardan aktarılan veriler özellik uzayında doğrusal sınıflandırılabilirler (Schölkopf 2002).

İkili sınıflandırmaya yarayan doğrusal çekirdek fonksiyonları her zaman en ideal çekirdek olmayabilir. Destek Vektör Makineleri doğrusal karar sınırının belirlenemediği durumlarda, verileri daha yüksek boyutlu bir özellik uzayına aktararak, bu yeni uzayda hiper düzlemsel bir sınıflandırma gerçekleştirir (Walgampaya and Kantardzic 2006).

Özellik uzayında iç çarpım fonksiyonunu genel bir benzerlik ölçüsü olarak kullanabilmek için öncelikle veri gömme işleminin gerçekleşmesi gerekmektedir.

$$\varphi : \mathcal{X} \longrightarrow \mathcal{H} \quad (4.4.1)$$

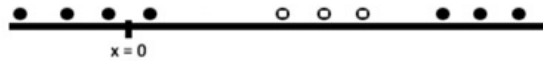
$$\varphi : x \longrightarrow \mathbf{x} = \varphi(x) \quad (4.4.2)$$

İç çarpım fonksiyonunu özellik uzayına gömülen verilere uygulamak için de;

$$k(x, x') = \langle x, x' \rangle = \langle \varphi(x), \varphi(x') \rangle \quad (4.4.3)$$

tanımlanabilir (Schölkopf 2002).

Şimdi doğrusal olarak verilen 3 beyaz ve 7 siyah noktadan oluşan örneği inceleyelim.



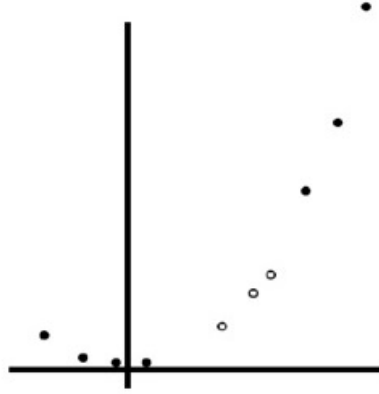
**Şekil 4.8:**  $\mathbb{R}^2$  de doğrusal sınıflandırılmayan nokta kümesi.

$\mathbb{R}$  uzayında bulunan bu siyah noktalar ile beyaz noktaları doğrusal olarak birbirinden ayırmak mümkün değildir. Dolayısı ile  $\mathbb{R}$  uzayında doğrusal ayrılamayan bu noktaları,

$$\varphi : \chi \longrightarrow \mathcal{H} \quad (4.4.4)$$

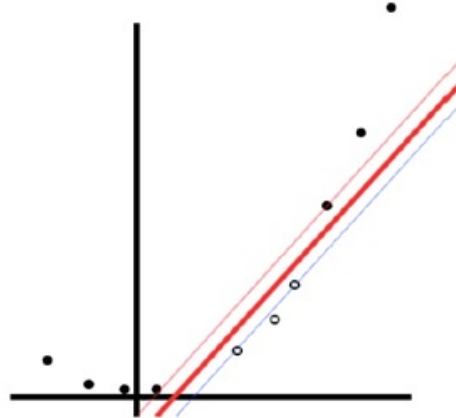
$$\varphi : x \longrightarrow \varphi(x) = (x, x^2) \quad (4.4.5)$$

fonksiyonu yardımıyla  $\mathbb{R}^2$  uzayına gömme işlemi gerçekleştirilirse veri noktaları Şekil (4.9) deki gibi olur.



**Şekil 4.9:** Şekil (4.8) deki noktaların  $\mathbb{R}$  den  $\mathbb{R}^2$  ye gömülmüş hali.

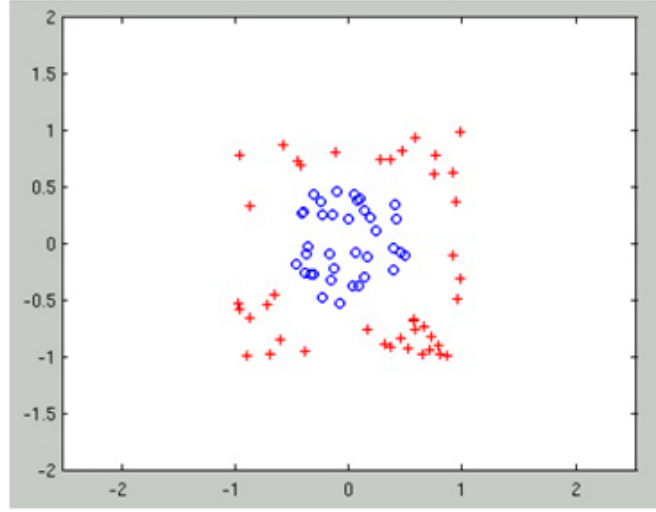
Girdi verileri geometrik olarak parabolik bir yapıya ulaştığından artık beyaz noktalar ile siyah noktaları doğrusal olarak ayırmak mümkündür.



**Şekil 4.10:** Şekil (4.9) deki  $\mathbb{R}^2$  ye gömülü verilerin sınıflandırılmış hali.

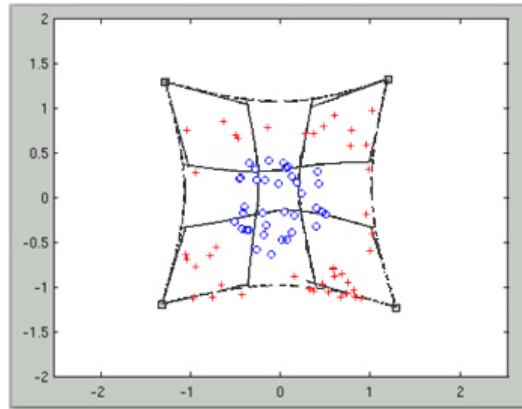
Şekil (4.10) de de görüldüğü üzere doğrusal olarak ayrılamayan girdi verilerini daha yüksek boyutlu bir özellik uzayına gömerek doğrusal sınıflandırmak mümkün kılınmıştır.

Şimdi de  $\mathbb{R}^2$  de doğrusal olarak ayıramayan girdi verilerinin de  $\mathbb{R}^3$  uzayına gömülmesi ile nasıl sınıflandırıldığını inceleyelim. Burada o ve + sembolleri ile temsil edilmiş girdi verilerini farklı sınıflara ayırmaya çalıştığımızı düşünelim.  $\mathbb{R}^2$  uzayında bu girdi verilerini doğrusal olarak ayırmak mümkün olmayacaktır. Şimdi özel bir  $\varphi$  fonksiyonu yardımıyla verileri  $\mathbb{R}^3$  e gömersek,



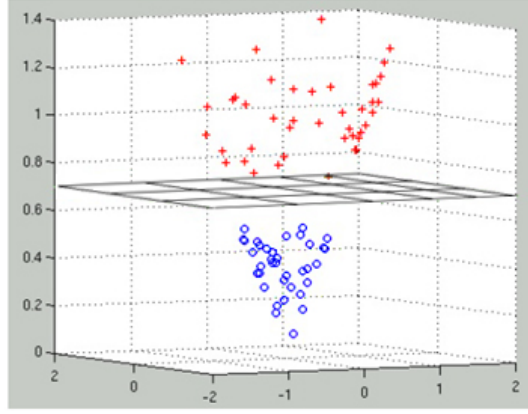
**Şekil 4.11:** Veri noktalarının  $\mathbb{R}^2$  uzayında doğrusal ayıramama durumu.

Şekil (4.12) de dikkat edilirse sanki "o" ile ifade edilen veri noktalarından düzlemin büküldüğünü ve üç boyuta taşındığını düşünebiliriz.



**Şekil 4.12:** (4.11) noktaların  $\mathbb{R}^3$  uzayına gömülmesi.

Verileri üç boyutta incelenirse,  $\mathbb{R}^3$  uzayında girdi verilerinin doğrusal olarak ayrıldığı Şekil (4.13)de gözlemlenebilir.



**Şekil 4.13:** (4.12) noktaların  $\mathbb{R}^3$  uzayında profil görüntüsü.

$m, n \in \mathbb{N}$  ve  $m < n$  olmak üzere  $m$  boyutta doğrusal olarak ayrılamayan veri noktalarının da bir  $\varphi$  fonksiyonu yardımıyla  $n$  boyutta bir uzaya gömülerek doğrusal ayrılabilceğini söyleyebiliriz.

#### 4.5 Basit Bir Örüntü Tanıma Algoritması

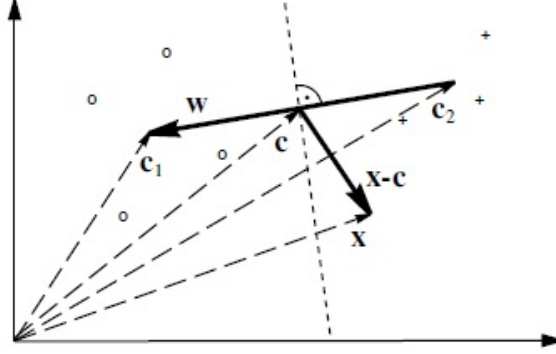
Verilerin  $\mathcal{H}$  özellik (Hilbert) uzayına gömülü olduğunu varsayalım. İç çarpım fonksiyonu yardımı ile uzaklıkları kullanarak benzerlik ölçüsü elde edebiliriz. Böylece  $\mathcal{H}$  özellik uzayını basit olarak iki sınıfa ayırabiliriz.

Sınıflandırmayı kolaylık olması açısından binary(ikili) olarak yaparsak, pozitif ve negatif olmak üzere iki sınıf elde ederiz. Pozitif sınıfların merkezini  $c_+$  ve negatif sınıfların merkezini  $c_-$  ile temsil edersek;

$$c_+ = \frac{1}{m_+} \sum_{\{i|y_i=+1\}} x_i \quad (4.5.1)$$

$$c_- = \frac{1}{m_-} \sum_{\{i|y_i=-1\}} x_i \quad (4.5.2)$$

Burada  $m_+$  pozitif etiketli (çıktı) örneklerin sayısını,  $m_-$  negatif etiketli örneklerin sayısını temsil etmektedir.  $m_+, m_- > 0$  olduğu için her iki sınıfı da boştan farklı kabul etmiş olduk. Aşağıdaki geometrik yapı iç çarpım açısından formüle edilebilir.  $c_-$  ile  $c_+$  nin orta noktasını  $c$  olarak tanımlayalım.



**Şekil 4.14:** (4.5.1) ve (4.5.2) denklemlerinin geometrik yorumu.

$\overrightarrow{c_+c_-}$  vektörüne de  $w$  ismi verilirse, geometrik olarak sınıflar betimlenmiş olur.  $\overrightarrow{w}$  vektörünün normali çizildiğinde  $\mathcal{H}$  özellik uzayı iki sınıfa ayrılmış olur.

Model kurulurken kullanılan veriler dışında yeni bir veri noktasını ele alalım. Yeni noktanın hangi sınıfa ait olduğunu belirleyebilmek için aşağıdaki adımlar izlenmelidir.

- Seçilen nokta  $x$  ile temsil edilsin,  $\overrightarrow{xc}$  vektörü belirlenir.
- Oluşan bu  $\overrightarrow{xc}$  vektörünün  $\overrightarrow{w}$  vektörü ile yaptığı açının kosinüsünün işaretine bakılır.

Böylece basit bir örüntü tanıma algoritması elde etmiş oluruz. Şimdi de bu algoritmayı matematiksel olarak inceleyelim.

$$\begin{aligned}
 y &= \text{sgn}\langle(\mathbf{x} - c), \mathbf{w}\rangle \\
 &= \text{sgn}\langle(\mathbf{x} - (c_+ + c_-)/2), (c_+ - c_-)\rangle \\
 &= \text{sgn}\langle\langle\mathbf{x}, c_+\rangle - \langle\mathbf{x}, c_-\rangle + b\rangle
 \end{aligned}$$

Karar fonksiyonu için ise,

$$\begin{aligned}
y &= \operatorname{sgn}\left(\frac{1}{m_+} \sum_{\{i|y_i=+1\}} \langle \mathbf{x}, \mathbf{x}_i \rangle - \frac{1}{m_-} \sum_{\{i|y_i=-1\}} \langle \mathbf{x}, \mathbf{x}_i \rangle + b\right) \\
&= \operatorname{sgn}\left(\frac{1}{m_+} \sum_{\{i|y_i=+1\}} k(x, x_i) - \frac{1}{m_-} \sum_{\{i|y_i=-1\}} k(x, x_i) + b\right)
\end{aligned}$$

elde edilir.

## 4.6 Vapnik-Chervonenkis Boyut ve VC Sınır

Vapnik ve Chervonenkisin ortaya attığı bu teori aslında sınıflandırma fonksiyonları için bir çeşit sınırlama ve kapasite belirleme ölçüsüdür. Karar sınırı kurallarının sabit bir koleksiyonu ile çalışmak, pratik nedenlerle daha verimli kavramsal bir çerçeve sunar.

### 4.6.1 VC Boyut

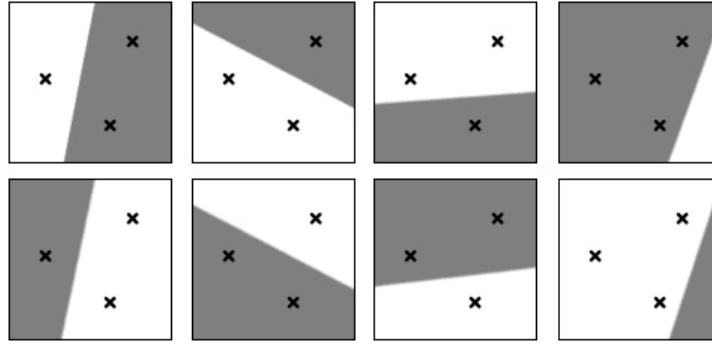
VC Boyut bir çeşit kapasite sınırıdır. Şöyle ki modeli sınıflara ayıran ve etiketlenmesine neden olan bir fonksiyondur. Etiketlerin  $\pm 1$  olması demek  $m$  girdi için  $2^m$  etiketleme demektir (Schölkopf 2002).

VC Boyut aslında fonksiyon kümelerinin bir özelliğidir ve fonksiyonların çeşitli sınıfları için tanımlanabilir. Burada sadece iki sınıf için karşılık gelen fonksiyonları dikkate alırsak, her  $x$  için,  $f(x) \in \{-1, 1\}$ . Şimdi eğer  $m$  puanı verilen bir fonksiyonun olası tüm etiketlemeleri  $2^m$  ise, bu veriler  $2^m$  parçaya ayrılabilir denir. VC Boyutu fonksiyonlar için parçalanabilir eğitim noktalarının sayısı olarak tanımlanır. Mesela VC Boyutu  $m$  ise, parçalanabilen  $m$  nokta kümesi vardır, fakat bu veri kümesi her zaman  $m$  noktaya parçalanabilir ifadesi doğru olmayacaktır (Burges 1998).

Öğrenen makineler tarafından uygulanan fonksiyonlar kümesinin sonlu bir VC boyutuna sahip olmasının gerekli ve yeterli olduğu kanıtlanmıştır. Düzgün yakınsama oranı üzerindeki dağılımdan bağımsız sınırların, VC boyutuna, eğitim hatalarının sayısına ve gözlem sayısına bağlı olduğu bulunmuştur. VC boyutu öğrenen makinenin kapasitesini ölçen sabit bir değerdir (Vapnik 2002).

İkili bir sınıflandırma yapılmak istendiğinde, VC boyutu sınıflandırıcının maksimum ayırabildiği sınıfların ( $2^m$ ) sayısıdır.  $\mathbb{R}^2$  de üç nokta sınıflandırılmak istendiğinde şekil



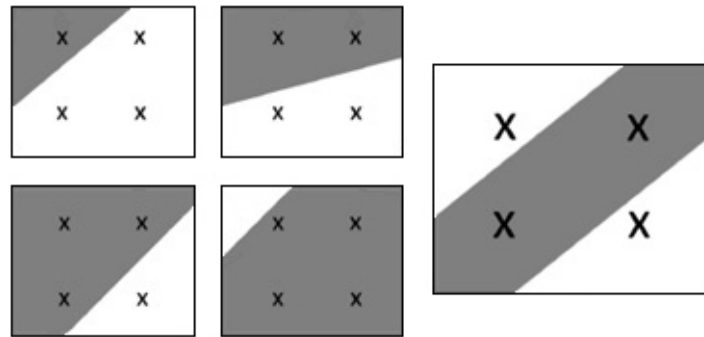


**Şekil 4.15:**  $\mathbb{R}^2$ de 3 noktanın iki sınıfa ayrılabilceği muhtemel tüm durumlar.

(4.15) de ki durumlar oluşur.

Bir eğitim kümesi büyüklüğü için, eğitim hatası ve test hatası arasındaki fark yüksek kapasiteli makineler için daha büyük olacaktır. VC boyutuna dayanan öğrenme teorisi, eğitim kümesi büyüklüğünün bir fonksiyonu olarak, eğitim hatası ile test hatası arasındaki farkın davranışının tek bir birim tarafından karakterize edilebildiğini öngörmektedir. Bu birim makinenin kapasitesini belirten VC boyutudur(Vapnik 1994).

İki boyutlu bir girdi uzayında sınıflandırma fonksiyonunun VC boyutu üçtür. n boyutlu bir girdi uzayında, hiperdüzlem sınıflandırma fonksiyonunun VC boyutu n +1 e eşittir; yani  $m = n+1$  dir. O halde n boyutlu bir uzayda en fazla n+1 noktanın tüm sınıflandırmaları gerçekleşebilir. Örneğin;  $\mathbb{R}^2$  de dört nokta sınıflandırılmak istendiğinde ise şekil (4.16) deki durumlar oluşur.



**Şekil 4.16:**  $\mathbb{R}^2$ de 4 noktanın iki sınıfa ayrılamadığı durum görüntüsü.

Şekil (4.16) e dikkat edilirse sağdaki gibi bir sınıflandırmanın mümkün olmadığını görebiliriz.

## 4.6.2 Parçalama

Verilen fonksiyon sınıfı, uzayı  $m$  noktaya parçalamak için yeterince zengin olmayabilir. Gereken sınıflamanın yapılabilmesi için VC Boyut en büyük  $m$  ye göre tanımlanmalıdır. Bu bir makine öğrenmesi kapasitesinin bir sayı karşılığı gibi düşünülebilir.

Bu nedenle bu fonksiyon daha doğru kapasite tedbirleri ile tanımlanmalıdır. Fakat bunları değerlendirmek çok zor kabul edilir. Yine de VC teorisinin kuramsal yapısında temel rol oynarlar. Bir başka ilginç kapasite ölçüsü VC Boyut un ölçekte hassas olan versiyonu VC Sınırı düşünülebilir (Schölkopf 2002).

## 4.6.3 VC Sınırı

Eğer fonksiyon sınıfının VC Boyutu  $h < m$  ise makine öğrenmesi kullanılabilir.  $R[f]$ , beklenen hata,  $R_{emp}[f]$ , deneysel hata olmak üzere,

$$R_{emp}[f] = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} |f(x_i) - y_i| \quad (4.6.1)$$

$$R[f] = \int \frac{1}{2} |f(x) - y| dP(x, y) \quad (4.6.2)$$

$$R[f] \leq R_{emp}[f] + \phi(h, m, \delta) \quad (4.6.3)$$

Burada  $\phi$  yi belli bir kapasite sınırı olarak düşünebiliriz.

$$\phi(h, m, \delta) = \sqrt{\frac{1}{m} \left( h \left( \ln \frac{2m}{h} + 1 \right) + \ln \frac{4}{\delta} \right)} \quad (4.6.4)$$

Bu ölçüler kapasiteyi belirleyip, problemin hangi sınırlar içerisinde çözümünün olabileceğinin ortaya çıkarılmasını sağlar (Schölkopf 2002).

## 4.7 Kayıp Fonksiyonu

Kayıp fonksiyonu aslında bir çeşit hata belirleme fonksiyonudur. Sınıf sayısı ve sınıflandırma yöntemi farklı kayıp fonksiyonları ortaya çıkarmıştır.

**Tanım 4.7.1:** (Kayıp Fonksiyonu)  $x$  girdi verisinin üçlü bileşenleri  $(x, y, f(x)) \in X \times Y \times Y$  olsun. Bir  $c$  fonksiyonu her  $x \in X$  ve her  $y \in Y$  olmak üzere,

$$c : X \times Y \times Y \longrightarrow [0, \infty), \quad c(x, y, f(x)) \in \{0, 1\} \quad (4.7.1)$$

olacak şekilde tanımlı fonksiyona kayıp fonksiyonu denir (Schölkopf 2002).

İkili sınıflandırmada yanlış sınıflandırma yapılırsa kayıp için 1 ceza puanı bulunurken, aksi durumlarda ceza puanı bulunmaz. Ama bu tanımdaki  $c$  fonksiyonu farklı sınıfları ve hata türlerini ayırt edemez.

$$c(x, y, f(x)) = \begin{cases} 0 & y = f(x) \\ 1 & y \neq f(x) \end{cases}$$

$f(x)$  reel değerli bir fonksiyon ise  $y \in \{-1, 1\}$  dir. Bu durumda sınıfların etiketlerini belirlemek için  $f(x)$  fonksiyonunun işaretine bakılır, yani sınıf etiketlerini belirleyen fonksiyon  $sgn(f(x))$  dir.

İkili sınıflandırma yapabileceğimiz bir veri kümesi ele alalım, kümelerin +1 ve -1 olarak ikiye ayrıldığını kabul edelim. Bu veri kümesinde hata fonksiyonu özel olarak Sıfır-Bir Kayıp Fonksiyonu olarak adlandırılır.

$$c(x, y, f(x)) = \frac{1}{2} |f(x) - y| \quad (4.7.2)$$

Fonksiyonunda sınıflandırma doğru ise  $f(x) = y$  olacağından  $c = 0$  (hata değeri) olacaktır. Aksi halde doğru sınıf +1 iken -1 bulunması durumunda veya doğru sınıf -1 iken +1 olması durumunda  $c = 1$  olacaktır. Bu nedenle  $c$  fonksiyonun değerlerinde sıfır sayısının çokluğu sınıflandırmanın doğruluk düzeyini belirlememize olanak sağlar (Schölkopf 2002).

## 4.8 Hiperdüzlem Sınıflayıcıları

Şimdiye kadar anlatılan girdi modellerinin öğrenme algoritmalarının istatistiksel etkinlikleri kontrol edilebilirdir. O halde tek ihtiyacımız olan kapasitesi hesaplanabilir fonksiyonların sınıflandırılmasıdır.  $\mathcal{H}$  iç çarpım uzayında,

$$\langle w, x \rangle + b = 0, \quad w \in \mathcal{H}, b \in \mathbb{R} \quad (4.8.1)$$

karar fonksiyonuna karşılık gelen,

$$f(x) = \text{sgn}(\langle w, x \rangle + b) \quad (4.8.2)$$

hiper düzlem sınıflayıcıları tanımlanmıştır. Buradan yola çıkarak veri sınıflayan tüm hiper düzlemler arasında eşsiz bir optimal hiper düzlem olduğunu söyleyebiliriz. Burada herhangi bir eğitim noktası ve hiper düzlem arasındaki marjı maksimum yaparak sınıflandırma gerçekleşir.

$$\max_{w \in \mathcal{H}, b \in \mathbb{R}} \min\{\|x - x_i\| : x \in \mathcal{H}, \langle w, x \rangle + b = 0, i = 1, 2, \dots, m\} \quad (4.8.3)$$

Hiper düzlemler arası mesafe ayrılan sınıf kapasitelerinin marjı arttıkça azalır. Amaç en büyük marjı elde edebileceğimiz normal vektörü bulmaktır.

$$\min_{w \in \mathcal{H}, b \in \mathbb{R}} \tau(w) = \frac{1}{2} \|w\|^2 \quad (4.8.4)$$

$$y_i(\langle w, x_i \rangle + b) \geq 1 \quad \forall i = 1, 2, \dots, m \quad (4.8.5)$$

Burada ki  $\tau$  fonksiyonu sınıflandırma probleminin amaç fonksiyonudur. Bu verilerle kısıtlamalı optimizasyon problemini düşünersek,

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i (y_i(\langle x_i, w \rangle + b) - 1) \quad (4.8.6)$$

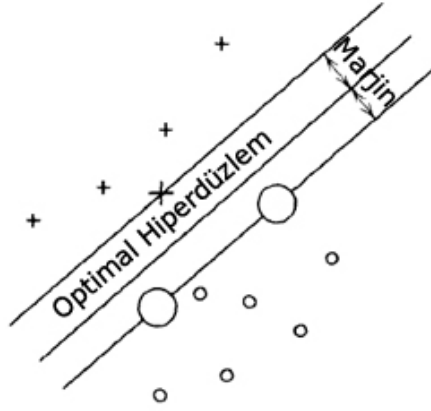
Bu eşitliklerde  $\alpha$  ile gösterilen ifadeler Lagrange çarpanları,  $w$  ve  $b$  denklemin ilkel değişkenleridir. Bu eşitlik  $w$  ve  $b$  değişkenleri için minimize edilirken,  $\alpha_i$  ler için de maksimize edilebilirdir (Schölkopf 2002).

Sonuç olarak, hiper düzlem sınıflayıcının amacı, maksimum marj ile verilerin en ideal sınıflara ayrılmasını sağlayarak sınıflandırma işlemini gerçekleştirmektir.

#### 4.8.1 Karush-Kuhn-Tucker (KKT) Şartları

Optimizasyon teorisinin tamamlayıcılık koşulları Karush, Kuhn ve Tucker tarafından ortaya atılmış ve ismini de buradan almıştır.

$$\frac{\partial}{\partial b} L(w, b, \alpha) = 0, \quad \frac{\partial}{\partial w} L(w, b, \alpha) = 0 \quad (4.8.7)$$



**Şekil 4.17:** Optimal hiper düzlem ve marj.

İlkel değişkenler açısından  $L$ nin türevleri yok olur. Buradan,

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad ve \quad w = \sum_{i=1}^m \alpha_i y_i x_i \quad (4.8.8)$$

elde edilir. Sıfırdan farklı  $\alpha_i$  ler le bu sistem,

$$w = \sum_{i=1}^m \alpha_i y_i x_i \quad (4.8.9)$$

destek vektörü olarak adlandırılır. KKT koşulları ile,

$$\alpha_i [y_i (\langle x_i, w \rangle + b) - 1] = 0, \quad \forall i = 1, 2, \dots, m \quad (4.8.10)$$

elde edilir. Burada yapılan işlemlerin avantajı çözümün kısıtlamalardan dolayı diğer örneklere bağımlı olmamasıdır. Bazen Lagrange çarpanlarında  $w$  ve  $b$  ilkel değişkenlerinden biri yok edilebilir. Böylece hiçbir kısıtlama kalmamış olur ve

$$\max_{\alpha \in \mathbb{R}^m} w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (4.8.11)$$

$$\forall \alpha_i \geq 0, \quad \sum_{i=1}^m \alpha_i y_i = 0 \quad (4.8.12)$$

elde edilir. Dolayısı ile destek vektörü,

$$w = \sum_{i=1}^m \alpha_i y_i x_i \quad (4.8.13)$$

olur. Bu destek vektörleri için yazılabilecek karar fonksiyonu,

$$f(x) = \operatorname{sgn} \left( \sum_{i=1}^m y_i \alpha_i \langle x, x_i \rangle + b \right) \quad (4.8.14)$$

şeklini alır ki burada yine b ilkel değişkenini yok sayabiliriz. Sonuç olarak KKT Şartları altında Destek Vektör Makinesi için bir karar fonksiyonu yazılabilir.

$$f(x) = \operatorname{sgn} \left( \sum_{i=1}^m y_i \alpha_i \langle \varphi(x), \varphi(x_i) \rangle + b \right) \quad (4.8.15)$$

$$= \operatorname{sgn} \left( \sum_{i=1}^m y_i \alpha_i k(x, x_i) + b \right)$$

ve karar fonksiyonunun Lagrange çarpanları ile KKT şartları altında bir optimizasyon problemine dönüştürmüş oluruz.

#### 4.8.2 Esnek Marj Sınıflandırması

Yüksek gürültü problemi sınıflarda büyük bir örtüşmeye neden olursa uygulamada sınıflayıcı bir hiper düzlem bulunamayabilir. Bahsedilen gürültü düzeyinden dolayı sınıf ihlali gerçekleşirse veri noktalarının doğru sınıflandırmasına izin vermek için daha esnek bir yapı gereklidir. Kısıtlamaları zayıflatıp sisteme esneklik kazandırarak,

$$\forall \xi_i \geq 0, \quad y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad \forall i = 1, 2, \dots, m \quad (4.8.16)$$

elde edilir.

İyi bir esnek sınıflandırıcı,  $\|w\|$  değerinin minimize edilmesi ile kapasite kontrolü yapar ve  $\sum_{i=1}^m \xi_i$  esneklik katsayılarının toplamıyla belirlenir. Sonuç olarak bu şekilde esnek marjlı bir sınıflandırmanın gerçekleşmesi amaç fonksiyonunu minimize etmiş olur.

$$\tau(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^m \xi_i \quad (4.8.17)$$

elde edilir.

Buradaki  $c > 0$  olması marj maksimizasyonu ve eğitim hatalarının minimizasyonu arasındaki değişimi belirler. Çekirdek fonksiyonu yardımı ile Lagrange çarpanları yeniden düzenlenebilir fakat bu durum yeniden bir maksimizasyon problemine dönüşür.

$$0 \leq \alpha_i \leq c, \quad i = 1, 2, \dots, m \quad \text{ve} \quad \sum_{i=1}^m \alpha_i y_i = 0 \quad (4.8.18)$$

Standart sınıflandırma yöntemlerinden tek farkı,  $\alpha(i)$  Lagrange çarpanları için C nin bir üst sınır belirtmesidir. Bu yöntemle girdi verilerinin herhangi birinin bireysel etkisini sınırlanmış oluruz.

$\alpha_i \leq c$  olduğu sürece b eşik değeri ihmal edilebilir.  $\xi_i = 0$  alınırsa,

$$\sum_{i,j=1}^m \alpha_i y_i k(x_i, x_j) = y_i \quad (4.8.19)$$

elde edilir.

Geometrik olarak b nin seçimi hiper düzlemin hareketini belirler. Olası hiper düzlemin esnek bir marj gerçekleştirme için v parametrizasyonu kullanılır. Bundan dolayı örneklerin alt ve üst sınırları daha belirgin bir hale getirmek için c parametresi  $v \in (0, 1]$  ile değiştirilir. Bu hata terimi ile ilkel bir amaç fonksiyonu kullanılır. Yine aynı şekilde  $\alpha_i \leq c$  olduğu sürece b eşik değeri ihmal edilebilir. Buradan,

$$c \sum_{i=1}^m \xi_i \quad \text{yerine} \quad \left( \frac{1}{vm} \sum_{i=1}^m \xi_i \right) - \rho \quad (4.8.20)$$

kullanılabilir.

Sınıfların ayrılması için gereken kısıtlamalar bir  $\rho$  marj parametresi içerir.

$$y_i \left( \langle w, x_i \rangle + b \right) \geq \rho - \xi_i \quad \forall i = 1, 2, \dots, m \quad (4.8.21)$$

Böylece kendiliğinden optimizasyon probleminin değişkeni haline gelir.

#### 4.9 Lasvm Algoritması ve Uygulama Şeması

Bu tez çalışmasında, deneysel çalışmalar ile problemin çözümünde 2005 yılında Bordes ve Bottou tarafından geliştirilen Lasvm algoritmasının eşli çekirdek tekniğine uyarlanmış

hali kullanılmıştır (Bordes and Bottou 2005).

Bilgisayarlar çok karmaşık istatistiksel modeller hesaplamak için yeteneğimizi geliştirmiştir. Ucuz, yaygın ve ağa bağlı bilgisayarların artık daha büyük bir ölçüde gözlem toplamak için yeteneğimizi arttırmış bulunmaktadır. Veri boyutları bilgisayarların hızında büyümesini gerekli kılmıştır. Son on yılda, işlemciler 100 kat daha hızlanmış ve sabit diskler de 1000 kez daha büyümüştür (Bordes et al. 2005).

Eğitim örnekleri çok yüksek boyutlu olduğunda teorik olarak öğrenme sistemleri de bol zamana ihtiyaç duyacaktır. Bu sebeple işlem maliyeti sınırlayıcı bir faktör haline gelir. Herhangi bir etkin öğrenme algoritması her örneğe belli bir önem derecesi vermelidir. Bu sorun deneysel bir cevap gerektirir. Destek Vektör Makineleri (DVM) problemin çözümü için çekirdek sınıflandırıcı sunar. Açıkça eğitim örneklerinden kendi içsel durumlarının ifadesi için çekirdek sınıflandırıcı kabul edilebilirdir (Bordes et al. 2005).

Fakat çalışmanın zaman maliyeti çözüm tekniğini kabul edilemez hale getirmektedir. Bordes ve arkadaşları, DVM çözümüne yakınsayan Lasvm adında yeni bir çevrimiçi öğrenme algoritması önermektedir. Farklı veri setleri üzerinde deneysel kanıtlar güvenilir bir eğitim seti üzerinden yüksek doğruluk oranlarına ulaşır. Daha az bellek kullanır ve daha hızlı bir algoritmadır. Tüm örneklerin işlemde her tekrarında bilgilendirici eğitim örnekleri seçmek için iki kriter inceler. Deneysel kanıtlar sınıf etiketlerinin kullanımını yapmadan bilgilendirici örnekler seçerek büyük ölçüde eğitim süresini azaltmak ve eşdeğer veya daha üstün doğruluk ile çok daha kompakt sınıflandırmanın mümkün olduğunu göstermektedir. Önceki çalışmalardan farklı olarak Lasvm, geleneksel yumuşak marjlı DVM formülasyonuna dayanır gürültülü veri setlerini de dikkate alarak oluşturulmuştur (Bordes et al. 2005).

Kullanılmakta olan hali hazırdaki sayısal algoritmalar SVM-QP problemi çözmek için geliştirilmiştir iyi algoritmalarındandır. En iyi bilinen yöntemler Eşlenik Gradyan ve Sıralı Minimal Optimizasyon yöntemidir (Vapnik and Chervonenkis 1982). Her iki yöntem de iyi seçilmiş yönleri izleyen aramalar yaparak çalışır. Her bir yön arama  $\alpha$  başlangıç vektörü, DVM problem kısıtlanması çözen ve belirtilen  $u$  yönünde uzanmaktadır. Böyle



bir arama yeni bir  $\alpha + \lambda^*u$  vektörü verir.

$$\lambda^* = \arg \max w(\alpha + \lambda u) \quad 0 \leq \lambda \leq \phi(\alpha, u) \quad (4.9.1)$$

$\phi(\alpha, u)$  üst sınırı  $\alpha + \lambda u$  kabul edilebilir. Hesaplamalardan sonuçla,  $K_{ij} = K(x_i, x_j)$  ve  $g = (g_1, g_2, \dots, g_n)$ ,  $w(\alpha)$  'nın gradiyent vektörü olmak üzere,

$$\lambda^* = \min \left( \phi(\alpha, u), \frac{\sum_i g_i u_i}{\sum_{i,j} u_i u_j K_{i,j}} \right) \quad (4.9.2)$$

elde edilir.

#### 4.9.1 Çevrimiçi Lasvm

Bu bölümde, Lasvm adlı bir çevrimiçi DVM algoritması verilmiştir. Bu algoritmayı incelemek için iki yol vardır. Mevcut çekirdek genişlemesinde biriken vektörler online işlem sürecinde kaldırılabilir. Lasvm algoritması, SMO algoritmasının sıralı yön aramaların bir yeniden yapılanmasıdır ve SMO algoritmasının sonuçları ile yakınsaktır.

Temel çekirdek algılayıcı ile karşılaştırıldığında, Lasvm algoritması incelikle gürültülü verileri de işler. Geleneksel DVM sınıflandırıcısı ile karşılaştırıldığında, Lasvm maliyet yararları ve çevrimiçi öğrenme algoritmalarına esneklik getirir. Sonuç olarak, deneysel kanıtlar Lasvm eğitim örnekleri üzerinden tek sıralı geçtikten sonra DVM doğruluğu eşleştiğini gösterir.

Burada Process ve Reprocess olarak adlandırılan iki tür yordam elde edilir (Bordes et al. 2005).

Lasvm algoritması eldeki veri bilgilerinin üç temel unsurunu korur.

- Potansiyel destek vektörleri indisleri kümesi S.
- Mevcut çekirdek katsayıları  $\alpha$ .
- Tanımlanmış kısmi türevler  $g_i$ .

Yalnızca  $i \in S$  olduğunda,  $\alpha_i$  ve  $g_i$  değişkenleri anlamlı değerler içerir. Eğer  $i \notin S$  ise  $\alpha_i$  katsayılarının kümesinin boş olduğu varsayılır.

### Yordam 1 (Lasvm Process)

1)  $k \in S$  ise atla.

$$2) \alpha_k \leftarrow 0, \quad g_k \leftarrow y_k - \sum_{s \in S} \alpha_s K_{ks}, \quad S \leftarrow S \cup \{k\}$$

3) Eğer  $y_k = +1$  ise

$$i \leftarrow k, \quad j \leftarrow \arg \min_{s \in S} g_s, \quad \alpha_s > A_s$$

değilse,

$$j \leftarrow k, \quad i \leftarrow \arg \max_{s \in S} g_s, \quad \alpha_s > B_s$$

4) Eğer (i,j) bir  $\tau^*$  (pozitif tolerans katsayısı) ihlali yoksa atla.

$$5) \lambda \leftarrow \min \left\{ \frac{g_i - g_j}{K_{ii} + K_{jj} - 2K_{ij}}, \quad B_i - \alpha_i, \quad \alpha_j - A_j \right\}$$

$$\alpha_i \leftarrow \alpha_i + \lambda, \quad \alpha_j \leftarrow \alpha_j - \lambda$$

$$g_s \leftarrow g_s - \lambda(K_{is} - K_{js}), \quad \forall s \in S$$

Lasvm yeni gelen bir veriyi sınıflandırmak için çevrimiçi kullanılabilir. Dolayısı ile anlık sınıflandırma yapmak mümkündür. Process algoritması veriler geldikleri gibi iteratif olarak çalışır.

### Yordam 2 (Lasvm Reprocess)

1)  $i \leftarrow \arg \max_{s \in S} g_s, \quad \alpha_s > B_s$

$$j \leftarrow \arg \min_{s \in S} g_s, \quad \alpha_s > A_s$$

2) Eğer (i,j) bir  $\tau^*$  (pozitif tolerans katsayısı) ihlali yoksa atla.

$$3) \lambda \leftarrow \min \left\{ \frac{g_i - g_j}{K_{ii} + K_{jj} - 2K_{ij}}, B_i - \alpha_i, \alpha_j - A_j \right\}$$

$$\alpha_i \leftarrow \alpha_i + \lambda, \quad \alpha_j \leftarrow \alpha_j - \lambda$$

$$g_s \leftarrow g_s - \lambda(K_{is} - K_{js}), \quad \forall s \in S$$

$$4) i \leftarrow \arg \max_{s \in S} g_s, \quad \alpha_s > B_s$$

$$j \leftarrow \arg \min_{s \in S} g_s, \quad \alpha_s > A_s$$

$$\forall s \in S \quad \text{iken} \quad \alpha_s = 0$$

$$\text{Eğer } y_s = -1 \text{ ve } g_s \geq g_i, \quad S = S - \{s\}$$

$$\text{Eğer } y_s = +1 \text{ ve } g_s \leq g_j, \quad S = S - \{s\}$$

$$5) b \leftarrow (g_i + g_j)/2, \quad \delta \leftarrow g_i - g_j$$

Adım 1 de durum değişkenlerine ilk değerleri atanır. Sonra adım 2 de önceden tanımlanmış çevrimiçi Lasvm algoritması dönüşümlü Process ve Reprocess dizisi uygulanır. Son olarak adım 3 te çekirdeğin tüm  $\tau^*$  ihlal çiftlerini kaldırmak için Reprocess çalıştırılarak çekirdek hesaplanması kolaylaştırılır.

### Algoritma (LASVM)

#### 1) Başlatma Şartı :

Her sınıfın bir kaç çekirdeği ile S belirlenir.

$\alpha_s \leftarrow 0$  alınır ve başlangıç gradyanı hesaplanır.

#### 2) Çevrimiçi tekrarlamalar:

Dizilerin tekrar sayısı belirlenir.

- Bir  $k_t$  örneği alınır.
- $k_t$  için Process yordamını çalıştır.
- Bir kere Reprocess yordamını çalıştır.

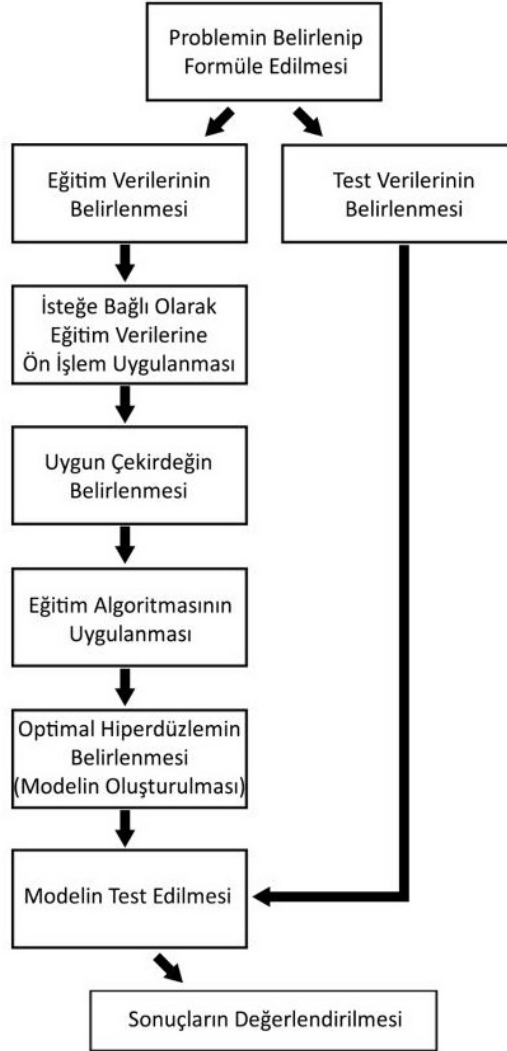
#### 3) Bitirme Şartı :

$\delta \leq \tau^*$  şartı sağlanana kadar Reprocess algoritmasını tekrarla (Bordes et al. 2005).

Bu tez çalışmasında kullanılan Lasvm algoritmasının eşli çekirdek tekniğine dönüştürülmemiş ilk hali (İnt.Kyn.2 2005) den elde edilmiştir. Tez çalışması için algoritma eşli çekirdek ile çalışacak şekilde düzenlenmiştir.

#### 4.9.2 Uygulama Şeması

Destek vektör sınıflandırmasının çalışma prensibi aslında iki temel adıma dayanmaktadır. Girdilerin en uygun algoritmayla eğitilmesi ve çıktılarla oluşturulan modelin test verileri ile test edilmesidir. Fakat bu çalışma prensibinin sadece bir özetidir. Bilimsel kaynaklarda bu işlem süreci çok farklı adımlamalarla ifade edilmiştir. Çalışmamızda kullandığımız bu adımlamaların en önemli süreçlerini şema olarak aşağıdaki şekilde verebiliriz.



Şekil 4.18: Uygulama şeması.

## 5 UYGULAMA ve SONUÇLARI

### 5.1 Doküman Sınıflandırma Problemi ve Yapılan Çalışmalar

Doküman sınıflandırma problemi için literatürde çeşitli çalışmalar bulunmaktadır. Bu çalışmalarda kullanılan teknik ve ara işlemler önemli rol oynamaktadır. Doküman sınıflandırma çalışmalarında daha çok veri setinden elde edilen kelime köklerinden bir sözlük oluşturulup, bu sözlüğe göre sınıflandırma gerçekleştirilmektedir.

Sınıflandırma problemleri günümüz teknolojik gelişmeleri için ihtiyaç duyulan tekniklere dönüşmüştür. Bu sebeple V.Vapnik ile başlayan sınıflandırma çalışmaları çok çeşitli bilim dallarında kullanılmaya başlanmıştır (Vapnik 1994). Bu gelişmeler başta İstatistik ve mühendislik bilim dallarında olmak üzere birçok bilimsel çalışmanın da yapılmasına olanak sağlamıştır.

Vladimir Vapnik ve Corinna Cortes 1995’de yaptıkları çalışmada metin sınıflandırmanın temelini oluşturan karakter tanıma problemi ile ilgilenmişlerdir (Cortes and Vapnik 1995). Verileri Amerika Ulusal Posta Servisi’nden elde ettiklerinden çok örnekli bir veri kümesi elde etmişlerdir. Bu çalışmalarında özellikle yüksek boyutlu veri kümeleri ile çalıştıklarından esnek marj ve hiper düzlemsel sınıflayıcıları ön planda tutmuşlardır.

M.Fatih Amasyalı ve Aytunç Beken 2009 yılında, Türkçe kelimelerin anlamsal benzerliklerini kullanarak yaptıkları çalışmada birçok sınıflandırma tekniği kullanıp bunları karşılaştırmışlardır. Bu çalışmada DVM tekniği veri sayısının değişmesine rağmen yüksek doğruluk oranıyla diğer yöntemlerden üstünlüğünü göstermiştir (Amasyalı ve Aytunç 2009) .

Thorsten Joachims 1998a makalesinde, metin sınıflandırmada DVM yöntemiyle hem teorik hem de deneysel olarak olumlu sonuçlar almıştır. Teorik olarak, DVM yönteminin yüksek boyutlu uzaylarda çalışabilme imkanı, ilgisiz veri noktalarının etkisinin az olması ve seyrek verilerle(eksik veri kümeleri ile) de işlem yapılabilmesi yönünden çok güçlü olduğu sonucuna varmıştır (Joachims 1998).

Simon Tong ve Daphane Koller 2001a makalelerinde, metin sınıflandırma konusunda daha güncel problemler belirlemişlerdir. İnternet aramaları, mail filtreleme gibi konular üzerine yaptıkları çalışmalarda DVM tekniğinin sonuçlarını değerlendirmişlerdir. Özellikle internet aramaları için tamamen bağımsız bir veri seti yerine veri tabanında var olan örneklerin değerlendirilerek daha iyi bir performans sergilemenin mümkün olduğunu görmüşlerdir (Tong and Daphne 2001).

Pairwise Çekirdek tekniği standart çekirdek sınıflandırma yöntemlerine alternatif olarak düşünülmüş bir yöntemdir. Pairwise tekniğinde oluşturulan veri çiftleri ile literatür çalışmalarında başarılı sonuçlar elde edilmiştir. Son yıllarda popülerliğini arttıran yöntem daha çok biyolojik sınıflandırma çalışmalarında tercih edilmiştir.

Franck Rapaport ve arkadaşları 2007 yılında yaptıkları çalışmada geliştirdikleri yeni çekirdek fonksiyonu ile pairwise tekniğini kullanarak genler arasında oluşan bağları ve protein protein eşleşmelerini konveks optimizasyon problemi ile çözmüşlerdir (Rapaport et al. 2007).

Hisashi Kashima ve arkadaşları 2010 yılında eşli çekirdek tekniğine, kartezyen çekirdek yöntemini kazandırmışlardır (Kashima et al. 2010).

## **5.2 Verilerin Elde Edilmesi**

Veriler elde edilirken daha önceden yapılmış çalışmalar esas alınarak araştırmalar yapılmıştır.

### **5.2.1 20 News Groups Haber Veri Seti**

20 Haber gurubu veri seti 20 farklı haber grupları arasından düzgün olarak bölümlenmiş yaklaşık 20000 belgeden oluşmaktadır. Veriler filtrelenerek Ken Lang tarafından öğrenmeye en elverişli hale getirilmiştir. 20 haber koleksiyonu, metin sınıflandırma ve metin kümeleme gibi makine öğrenme teknikleri, metin uygulamalarında deneyler için belirlenen popüler bir veri haline gelmiştir (Lang 2004).

Veriler 20 farklı haber gurubu başlığında, her yeni gelen farklı başlığa olacak şekilde

düzenlenmiştir. Haberlerden bazıları birbiri ile çok yakından ilişkilidir (comp.sys.ibm.pc.-hardware / comp.sys.mac.hardware), bazıları ise birbirinden oldukça ilgisizdir (misc.forsale / soc.religion.christian). Konularına göre bölümlenmiş 20 haber gruplarının bir listesi aşağıda verilmiştir.

**Çizelge 5.1:** Orjinal haber grupları tablosu

comp.graphics	rec.autos	sci.crypt
comp.os.ms-windows.misc	rec.motorcycles	sci.electronics
comp.sys.ibm.pc.hardware	rec.sport.baseball	sci.med
comp.sys.mac.hardware	rec.sport.hockey	sci.space
comp.windows.x		
misc.forsale	talk.politics.misc	talk.religion.misc
	talk.politics.guns	alt.atheism
	talk.politics.mideast	soc.religion.christian

Burada veri seti 6 ana küme (Bilgisayar, spor, bilim, satılık ilanları, siyaset, din) ve bu kümelerin 20 alt kümesi olarak gruplanmışlardır (İnt.Kyn.1 2008).

### 5.3 Problemin Eşli Çekirdek Yöntemine Uyarlanması

Herhangi bir veri setine keyfi olarak eşli çekirdek tekniği uygulanamaz. Bunun nedeni verilerin teknik olarak algoritmaya hazır hale getirilme gereksinimidir. Veriler genellikle noktasal kimlikleri ile depolanırlar. Bu sebeple Pairwise tekniği için her veri noktasının ikili düğümleri belirlenmelidir. Başka bir şekilde ifade etmek istersek,  $x_i$  noktasının tekil ölçümleri ile çekirdek hesaplaması yapmak yerine öncelikle  $x_i$  ve  $x_j$  arasında bir bağ oluşup oluşmadığı belirlenmelidir. Eğer  $x_i$  ile  $x_j$  veri noktaları kendi aralarında bağ yapıyorlarsa +1, bağ yapmıyorlarsa -1 değerini alır. Veri setinin özelliklerine göre bağ yapma durumunu aynı kümeye ait olma durumu olarak da ifade edebiliriz.

Eşli çekirdek tekniğine göre düzenlenen verileri, veri çifti şeklinde ifade etmek daha doğru olacaktır. Sonuç itibarıyla oluşan bu veri çiftleri arasındaki ilişki varsayılan nok-

tasal çarpım fonksiyonları ile hesaplamak mümkün olmayacaktır. Bu sebeple alternatif çekirdekler geliştirilmiştir.

Bu çalışmada veri kümesi hali hazırda kullanılan DVM teknikleri kullanılarak doğrusal, Gauss Radyal çekirdek fonksiyonları ile modellenmiş ve test edilmiştir. Çıkan sonuçlar çekirdek fonksiyonlarının gösterdikleri performansları değerlendirmek üzere çapraz tablolar ile verilmiştir. Buna ek olarak eşli çekirdek fonksiyonları için veri kümesi düzenlenmiş ve analiz edilmiştir.

Verilerin eşli çekirdek tekniğine uygun olarak veri çiftlerine dönüştürülmesini Çizelge (5.2) deki gibi küçük bir örnek veri kümesi ile ifade edilebilir. Burada dikkat edilmesi gereken önemli bir unsurda, veri setinin büyüklüğüdür. Çünkü büyüklüğü  $n$  adet veriden oluşan veri kümesi, eşli veri çiftlerine dönüştürüldüğünde  $n^2$  adet veri büyüklüğüne ulaşacaktır. Ayrıca veri kümesinde bulunan noktalar  $\{x_1, x_2, \dots, x_n\}$  iken, veri çiftlerine dönüştürülmüş veri setinin noktaları  $\{(x_1, x_1), (x_1, x_2), \dots, (x_n, x_n)\}$  şekline gelmektedir.

**Çizelge 5.2:** Eşli Çekirdek örnek tablosu

$x_i$	$x_j$	Bağ Yapma Durumu	Oluşan Veri Çifti ve Etiketi
$x_1$	$x_1$	+1 (aynı sınıfa ait)	$(x_1, x_1) \longrightarrow +1$
$x_2$	$x_1$	-1 (farklı sınıfa ait)	$(x_2, x_1) \longrightarrow -1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_n$	$x_1$	+1	$(x_n, x_1) \longrightarrow +1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{n-1}$	$x_n$	-1	$(x_{n-1}, x_n) \longrightarrow -1$
$x_n$	$x_n$	+1	$(x_n, x_n) \longrightarrow +1$



## 5.4 Uygulama

Destek Vektör Makinesi, vektör ailelerinin sınıflandırma sistemi olduğundan tek bir sınıflandırma işlemi için bile tüm destek vektörlerinin sisteme yeni katılan nokta ile işleme girmesi gerekir. Veri kümesi ne kadar büyürse, maliyet o oranda artmaktadır. Maliyeti azaltmak için birçok ara işlem yapılabilir. Kullanılan algoritma ara bellekte tuttuğu çekirdek değerleri sayesinde tüm işlemleri tekrar yapmak yerine sadece gerekli destek vektörlerini kullanarak maliyeti düşürmemize yardım eder.

Çeşitli kernel fonksiyonları doğruluk oranını ve maliyeti etkiler. Bu sebeple herhangi bir veri seti için yapılan uygulamalarda farklı kernel fonksiyonları kullanılmalı ve veri seti için optimal sonuçlar veren çekirdek fonksiyon belirlenmelidir.

### 5.4.1 Çeşitli Çekirdek Fonksiyonları ile Uygulama ve Sonuçları

DeneySEL çalışmada haber grupları 1 ile 20 arasında numaralandırıp, deneyler yapıldığından bu çalışmada aksi belirtilmedikçe haber grupları isimleri ile değil Çizelge (5.3) deki kodları ile verilecektir.

**Çizelge 5.3:** Haber grupları ve kullanılacak kodları

Sınıf	Kod	Sınıf	Kod
alt.atheism	1	rec.sport.hockey	11
comp.graphics	2	sci.crypt	12
comp.os.ms-windows.misc	3	sci.electronics	13
comp.sys.ibm.pc.hardware	4	sci.med	14
comp.sys.mac.hardware	5	sci.space	15
comp.windows.x	6	soc.religion.christian	16
misc.forsale	7	talk.politics.guns	17
rec.autos	8	talk.politics.mideast	18
rec.motorcycles	9	talk.politics.misc	19
rec.sport.baseball	10	talk.religion.misc	20

Veri çiftleri temsili olarak  $P_s = (x_i, x_j)$  ve  $P_r = (x_m, x_n)$  olsun. Çalışmada Eşli çekirdek tekniği için aşağıdaki iki farklı yöntem kullanılmıştır.

**TPPK Kernel :**

$$\kappa(P_s, P_r) = \kappa(x_i, x_m) \cdot \kappa(x_j, x_n) + \kappa(x_i, x_n) \cdot \kappa(x_j, x_m) \quad (5.4.1)$$

**MLPK Kernel :**

$$\kappa(P_s, P_r) = [\kappa(x_i, x_m) - \kappa(x_i, x_n) - \kappa(x_j, x_m) + \kappa(x_j, x_n)]^2 \quad (5.4.2)$$

Elde edilen deneysel sonuçları değerlendirmek için Accuracy (ACC) doğru sınıflandırma oranı ve Receiver Operating Characteric (ROC) yani doğru pozitif sınıflandırmalarının yanlış pozitiflere sınıflandırmalara oranı yöntemleri kullanılmıştır. 20 Haber Grubu veri setinde optimal sınıflandırmayı belirleyebilmek için veri setinden  $\text{Comp}\{2,3,4,5,6\}$  ve  $\text{Rec.Sports}\{8,9,10,11\}$  olmak üzere iki ana kategori belirleyip bu kategorilerden Comp grubunu pozitif (+), Rec.Sports grubunu negatif (-) olarak sınıflandırdığımızda Çizelge (5.4) deki sonuçlar elde edilmiştir. Bu deneysel çalışmada DVM parametrelerinden literatür çalışmalarında en ideal sonuçların alındığı iddia edilen  $\varepsilon = 1$ ,  $C = 10$  kabul edilmiştir.  $n = 50000$  gözlemlik haber metinleri rassal olarak seçilmiştir. Comp grubunda beş alt grup, Rec.Sports grubunda dört alt grup bulunduğundan tüm olası kombinasyonlarla deneyler yapılarak optimal eşleşmeler belirlenmiştir.

**Çizelge 5.4:** Haber gruplarının tüm olası kombinasyonları.

Eğitim Kümesi		Test Kümesi		TPPK		MLPK	
(+)	(-)	(+)	(-)	ACC	ROC	ACC	ROC
2,3	8,9	4,5,6	10,11	64.47	0.703994	65.03	0.71164
2,3	8,10	4,5,6	9,11	67.94	0.7561	66.67	0.749
2,3	8,11	4,5,6	9,10	67.795	0.750011	66.65	0.74488
2,3	9,10	4,5,6	8,11	68.865	0.762968	68.954	0.7775
2,3	9,11	4,5,6	8,10	67.235	0.750162	68.741	0.75412
2,3	10,11	4,5,6	8,9	60.75	0.668157	61.786	0.67921
2,4	8,9	3,5,6	10,11	67.485	0.745719	66.581	0.71469
2,4	8,10	3,5,6	9,11	73.09	0.814879	74.192	0.821569
2,4	8,11	3,5,6	9,10	71.32	0.793971	70.96	0.78865
2,4	9,10	3,5,6	8,11	71.82	0.797739	72.81	0.81659
2,4	9,11	3,5,6	8,10	70.195	0.791092	69.84	0.80008
2,4	10,11	3,5,6	8,9	59.19	0.657117	60.53	0.649583
2,5	8,9	3,4,6	10,11	66.205	0.73193	65.20	0.74265
2,5	8,10	3,4,6	9,11	73.345	0.81543	74.13	0.824761
2,5	8,11	3,4,6	9,10	71.215	0.801055	71.89	0.81256
2,5	9,10	3,4,6	8,11	71.65	0.796949	70.560	0.804693
2,5	9,11	3,4,6	8,10	71.96	0.797982	72.236	0.78698
2,5	10,11	3,4,6	8,9	59.435	0.671429	58.483	0.687962
2,6	8,9	3,4,5	10,11	61.825	0.671761	60.046	0.68432
2,6	8,10	3,4,5	9,11	65.98	0.727378	65.243	0.735692
2,6	8,11	3,4,5	9,10	64.24	0.710563	65.71	0.711789
2,6	9,10	3,4,5	8,11	67.695	0.753157	68.435	0.761789
2,6	9,11	3,4,5	8,10	68.035	0.751441	66.928	0.76175
2,6	10,11	3,4,5	8,9	59.765	0.651721	60.514	0.66813
3,4	8,9	2,5,6	10,11	64.685	0.710676	65.8	0.72489
3,4	8,10	2,5,6	9,11	68.9	0.767525	69.523	0.759346
3,4	8,11	2,5,6	9,10	68.715	0.762986	67.25	0.77951
3,4	9,10	2,5,6	8,11	68.795	0.763213	67.930	0.75924
3,4	9,11	2,5,6	8,10	68.97	0.765991	69.12	0.76814

Çizelge 5.4 (Devam)

Eğitim Kümesi		Test Kümesi		TPPK		MLPK	
(+)	(-)	(+)	(-)	ACC	ROC	ACC	ROC
3,5	8,10	2,4,6	9,11	72.14	0.802796	71.947	0.79437
3,5	8,11	2,4,6	9,10	71,425	0.796258	70.573	0.862479
3,5	9,10	2,4,6	8,11	71.13	0.790727	71.658	0.791256
3,5	9,11	2,4,6	8,10	70,095	0.784521	71.54	0.81436
3,5	10,11	2,4,6	8,9	59.27	0.656839	60.813	0.662746
3,6	8,9	2,4,5	10,11	63.71	0.697012	64.763	0.70439
3,6	8,10	2,4,5	9,11	66.895	0.742094	65.075	0.75681
3,6	8,11	2,4,5	9,10	66.835	0.737838	67.254	0.748315
3,6	9,10	2,4,5	8,11	68.41	0.760781	70.453	0.78785
3,6	9,11	2,4,5	8,10	64.99	0.744611	66.01	0.75586
3,6	10,11	2,4,5	8,9	59.28	0.645587	59.25	0.64584
4,5	8,9	2,3,6	10,11	63.885	0.695939	62.184	0.694513
4,5	8,10	2,3,6	9,11	67.79	0.751715	68.125	0.75485
4,5	9,10	2,3,6	8,11	67.065	0.745538	67.45	0.74874
4,5	9,11	2,3,6	8,10	67.01	0.741044	67.123	0.7475
4,5	10,11	2,3,6	8,9	56.785	0.613866	58.125	0.63452
4,6	8,9	2,3,5	10,11	68.085	0.75394	70.187	0.759826
4,6	8,10	2,3,5	9,11	72.175	0.805351	70.256	0.78787
4,6	9,10	2,3,5	8,11	71.57	0.800498	71.125	0.79145
4,6	10,11	2,3,5	8,9	58.885	0.652984	61.01	0.671234
5,6	9,11	2,3,4	8,10	72.545	0.813163	73.02	0.800042
5,6	10,11	2,3,4	8,9	59.65	0.677626	60.85	0.676754

**Çizelge 5.5:**  $\varepsilon$  parametresindeki deęişimler.

C	$\varepsilon$	TPPK	
		ACC	ROC
10	0.001	72.055	0.803623
10	0.01	72.055	0.803623
10	0.1	72.055	0.803623
10	1	72.175	0.805351
10	10	71.02	0.792166
10	100	71.02	0.792166
10	1000	71.02	0.792166

Çizelge (5.4) deki sonuçlardan, çalışmada en ideal sonuçların alındığı eşleşme 5 ve 6 pozitif eğitim sınıfı, 8 ve 10 negatif eğitim sınıfı olarak, 2, 3 ve 4 pozitif test sınıfı, 9 ve 11 ise negatif test sınıfı olarak belirlenmiştir.

Dikkat edilecek olursa Çizelge (5.4) de TPPK ve MLPK pairwise çekirdek yöntemlerinin sonuçlarının birbirine çok yakın olduğu görülmektedir. Bu sebeple parametrelerin belirlenmesinde çekirdek yöntemlerden birini seçmemiz daha pratik ve yeterli olacaktır. Çizelge (5.5) de TPPK pairwise çekirdek yöntemi uygulanarak yapılan deneylerde öncelikle  $C = 10$  parametresi sabit alınmış ve  $\varepsilon$  parametresindeki deęişimler incelenmiştir.

Çizelge (5.5) den sonuçla  $\varepsilon = 1$  parametre deęeri ile doğru sınıflandırmanın tepe noktasına ulaşıldığını görebiliriz. Deneysel çalışmada  $\varepsilon$  parametresinin 0.1 ile 1 aralığındaki deęerleri için daha hassas deneyler yapılmış fakat sonucu etkilemediğinden Çizelge (5.5) de verilmemiştir.

Ayrıca bu sonuçlardan yola çıkarak  $\varepsilon = 1$  parametre deęerini sabit alıp  $C$  parametresindeki deęişimler de incelenmiştir.  $C$  parametresinin sistemin esnekliğini belirleyen bir ceza parametresi olduğundan deęişimlerin problemin çözümüne etkileri önemli bir hal almaktadır. Bu deęişimlerin sınıflandırmanın doğruluk oranlarına etkileri de Çizelge (5.6) de verilmiştir.

**Çizelge 5.6:**  $C$  parametresindeki deęişimler.

$\varepsilon$	$C$	TPPK	
		ACC	ROC
1	0.001	53.17	0.555565
1	0.01	50	0.563659
1	0.1	52.07	0.561714
1	1	64.23	0.70711
1	10	66.17	0.735244
1	100	66.11	0.735114
1	1000	66.11	0.735114

Elde edilen neticelerden, 20 Haber Grubu veri seti için 50000 rassal seçimle elde edilmiş örneklem ile deneyler yapıldığında,  $\varepsilon = 1$  ve  $C = 10$  parametre deęerlerinin en ideal sonuçları verdiği belirlenmiştir. Bu sebeple çalışmanın bu aşamasından sonra belirlenen parametre deęerleri ile deneyler yapılmıştır.

Yapılan deneyler ile optimal eşleşen gruplar ve çekirdek parametreleri belirlendiğinden artık farklı sayıda eğitim verileri ve test verileri ile deneyler yapılarak sonuçları belirlenebilir.

Çizelge (5.7) de, farklı sayıda eğitim verilerinin 50000 test verisi için gösterdikleri performanslar verilmiştir. Çizelge (5.7) deki deneysel sonuçlar, farklı sayıda veri ile DVM nin eğitimi yapılmış ve sabit olarak 50000 veri ile test edilerek hesaplanmıştır. Kullanılan TPPK ve MLPK pairwise çekirdekleri ile genel olarak eğitim verilerinin sayısı arttıkça ACC ve ROC deęerlerinin yükseldiği söylenebilir.

**Çizelge 5.7:** 50000 test dokümanı için deney sonuçları.

Eğitim Verisi Sayısı (n)	TPPK		MLPK	
	ACC	ROC	ACC	ROC
1000	54.43	0.558848	57.31	0.599182
2500	57.21	0.606352	60.61	0.640378
5000	63.52	0.690406	63.74	0.674516
10000	66.77	0.739657	65.23	0.711881
15000	71.01	0.789158	68.79	0.749666
20000	73.77	0.815749	68.86	0.761805
25000	73.7	0.822586	69.16	0.773641
30000	74.84	0.834969	69.64	0.777622
35000	75.53	0.844959	69.25	0.780297
40000	76.13	0.855329	70.71	0.79143
45000	77.53	0.860832	69.41	0.792881
50000	77.97	0.866653	73.31	0.801751
55000	78.11	0.870087	72.12	0.799266
60000	78.73	0.8773	74.19	0.808676
65000	79.45	0.883978	74.53	0.806576
70000	79.34	0.883091	74.01	0.803936
75000	79.7	0.886823	74.25	0.810331
80000	79.42	0.884319	73.66	0.808572
85000	80.33	0.890223	75.26	0.817195
90000	79.38	0.88517	75.86	0.812368
95000	80.52	0.892089	75.08	0.815112
100000	80.93	0.89387	76.58	0.824198
150000	80.95	0.893071	77.26	0.817268
200000	80.8	0.896823	79.18	0.816557
250000	81.33	0.90317	78.08	0.826178

**Çizelge 5.8:** 50000 eğitim dokümanı için deney sonuçları.

Test Veri Sayısı (n)	TPPK		MLPK	
	ACC	ROC	ACC	ROC
1000	76.6	0.859928	79.25	0.831924
2500	78.2	0.872204	78.71	0.84116
5000	77.94	0.867332	78.41	0.854765
10000	78.2	0.870905	77.31	0.850945
15000	79.5	0.889198	80.12	0.889145
20000	77.74	0.871054	78.19	0.87444
25000	77.45	0.874532	78.16	0.862245
30000	78.2	0.884705	79.64	0.900455
35000	77.75	0.870544	76.25	0.86784
40000	79.1	0.894522	77.71	0.887822
45000	78.55	0.8744105	77.41	0.86547
50000	77.554	0.871404	78.16	0.88884
55000	76.8	0.85524	76.64	0.866742
60000	74.5	0.844342	75.25	0.85124
65000	72.35	0.814645	73.53	0.82111
70000	73.5	0.820574	70.01	0.821844
75000	72.7	0.824522	70.25	0.83457
80000	72.45	0.81254	69.66	0.804178
85000	70.25	0.804512	67.26	0.812482
90000	68.4	0.784705	67.86	0.79455
95000	68.75	0.78451	64.08	0.75451
100000	68.65	0.75754	64.58	0.751245
150000	69.56	0.764527	63.26	0.7256
200000	67.9	0.734175	63.18	0.71465
250000	67.4	0.72454	62.08	0.69745

Çizelge (5.8) deki deneysel sonuçlar, 50000 veri ile DVM nin eğitimi yapılmış ve farklı sayılarda veriler ile test edilerek hesaplanmıştır. Kullanılan TPPK ve MLPK pairwise çekirdekleri ile genel olarak test verilerinin sayısı eğitim verileri ile yaklaşık olarak aynı



seviyelerdeyken ACC ve ROC deęerlerinin yükseldięi söylenebilir. Dięer taraftan, test verilerinin sayısı eğitim verilerinin sayısından arttıķa giderek ACC ve ROC deęerlerinin azaldıęı söylenebilir.

Ana kümeden seçilen ve 50000 örneklemden oluşan eğitim kümesi ile test kümesi tamamen aynı kümeler olduęunda, doğrusal çekirdek ile oluşan deneysel sonuçlardan alınan çıktıları Çizelge (5.9) de verilmiştir.

**Çizelge 5.9:** Eğitim ve test kümeleri aynı

$\varepsilon$	C	TPPK		MLPK	
		ACC	ROC	ACC	ROC
1	10	100	1	99.975	1

Ana kümeden seçilen ve 50000 örneklemden oluşan eğitim kümesi ile test kümesi tamamen farklı kümeler olduęunda, doğrusal çekirdek ile oluşan deneysel sonuçlardan alınan çıktıları Çizelge (5.10) de verilmiştir.

**Çizelge 5.10:** Eğitim ve test kümeleri farklı

$\varepsilon$	C	TPPK		MLPK	
		ACC	ROC	ACC	ROC
1	10	55.995	0.586245	54.095	0.587334

Ana kümeden seçilen 50000 örneklemden oluşan eğitim kümesi ile test kümesi aynı fakat farklı örneklemler olduęunda doğrusal çekirdek ile oluşan deneysel sonuçlardan alınan çıktıları Çizelge (5.11) de verilmiştir.

**Çizelge 5.11:** Eğitim ve test kümeleri aynı örneklem farklı

$\varepsilon$	C	TPPK		MLPK	
		ACC	ROC	ACC	ROC
1	10	83.435	0.911404	83.225	0.90299

20 Haber Grubu veri setinin sınıflandırılmasında doğrusal çekirdek ve Gauss Radyal çekirdek fonksiyonlarının probleme etkileri farklı olabilir. Bu sebeple TPKK Kernel

pairwise çekirdeği ile yapılan deneylerde bu iki çekirdek fonksiyon karşılaştırılması gerekmektedir. Bu deneylerden elde edilen sonuçlar Çizelge (5.12) de verilmiştir.

Yapılan bu deneyler Gauss Radyal çekirdeğin sonuçlarının, problemin boyutu ile ilgili olduğu sonucuna varılabilir.

**Çizelge 5.12:** Doğrusal çekirdek ile Gauss Radyal çekirdek performansları.

	Eğitim (n)	Test (n)	$\epsilon$	C	TPPK	
					ACC	ROC
Doğrusal Çekirdek	1000	10000	1	10	55.13	0.579344
	10000	10000	1	10	71.1	0.795753
	20000	10000	1	10	77.6	0.840326
	50000	10000	1	10	80.63	0.894072
	100000	10000	1	10	83.32	0.920423
Gauss Radyal Çekirdek	1000	10000	1	10	50.14	0.486913
	10000	10000	1	10	48.99	0.490115
	20000	10000	1	10	49.49	0.491263
	50000	10000	1	10	49.67	0.491995
	100000	10000	1	10	50.14	0.498997

## 5.5 Sonuçlar ve Öneriler

Sonuçları değerlendirmeden önce, Destek Vektör Makineleri içinde kullanılan tekniklerin istatistiksel boyutunu da incelememiz gerekmektedir. İstatistik bilim dalı genel anlamda gözlemleri belli bir dağılıma uygunluğu neticesinde en ideal tahmini yapmaya çalışır. Fakat günlük hayatta karşımıza çıkan veriler daima belli bir dağılıma uymak zorunda değildir. Trafik kazaları, borsa verileri, hava durumu tahmini,... gibi veriler bazı durumlarda belli bir dağılıma uygunluk göstermez. Böyle veri setlerinde Destek Vektör Makineleri diğer istatistiksel yöntemlere göre bize daha avantajlı bir yol sunmaktadır. Standart Kümeleme Analizi yöntemlerinde yalnızca metrik fonksiyonlar kul-

lanıldığından bazı veri kümeleri için oluşan eş uzaklık sınıflandırmayı güçleştirmekte ve hata oranının yüksek olmasına sebebiyet vermektedir. Bu sebeple Destek Vektör Makineleriyle oluşan sınıflandırma modeli ile Kümeleme Analizinden daha iyi sonuçlar sunduğu yapılan çalışmalarda belirtilmiştir. Ayrıca istatistiksel tahminleme yöntemleri de gözlem değerlerinden oluşan veri setinin belli bir dağılıma uymasını gerektirdiğinden her veri setinde istenilen sonuçlar ve hata oranı elde edilememiştir.

Destek Vektör Makineleri gözlem değerleri ile eğitim yaparak oluşturduğu modelde olasılık değerleri ile gözlenemeyen veriler için de en büyük marj ile sınıflandırmayı gerçekleştirdiğinden gelecek adına tahminde bulunmak için de kullanılabilir. Özellikle çok sınıflı uygulamalarda gözlemlenen veri setine sonradan oluşacak olan noktasal veri için tahminleme yapılabilir.

Yıllar içerisinde popülerlik kazanan Destek Vektör Makineleri ile ilgili bilimsel çalışmalar regresyon yöntemleri için de farklı bir bakış açısı sunmuştur. Destek Vektör Regresyonu ile regresyon tekniklerinin aykırı değerlerden daha az etkilenebilecek bir modele kavuşmasını sağlamıştır. Genellikle bankacılık işlemlerinde (Hisse senedi tahmini, borsa tahmini, . . . gibi) uluslararası çalışmalarının birçoğunda Destek Vektör Regresyonu kullanılmaktadır. Bu çalışmada da Destek Vektör Makineleri yönteminde son zamanlarda popülerlik kazanan pairwise kernel tekniği ile doküman sınıflandırma çalışması yapılmıştır.

Bu tez çalışmasında elde edilen deneysel sonuçlar ile 20 Newsgroups veri seti sınıflandırılmıştır. Dökümanlardan oluşan bu veri setinde 18774 doküman ve her doküman için 61188 özellik bulunmaktadır. Özellik sayısı probleme matematiksel anlamda boyut kavramına eşdeğer olarak etki etmektedir. Pairwise çekirdek kullanılarak deneyler yapıldığından veri sayısı  $18774 \times 18774 = 352463076$  olmaktadır. Dolayısı ile deneysel çalışmalarda eğitim veya test kümesi yeterince büyük seçilebilmektedir. Bu sayede eğitim veya test kümesindeki verilerin birbirinden farklı olarak seçilmesi sağlanmış olur. En yüksek 250000 veri ile eğitim veya test yapıldığından elde edilen sonuçlarda veri setinin yaklaşık olarak  $250000 / 352463076 \cong 0.0007$  oranında örneklem kullanılmıştır. Tüm deneylerde rasgele örnekleme yöntemi ile veriler seçilmiştir. Veriler seyrek(sparse) olduğundan

deneyler C++ programlama dilinde yapılmıştır. En kısa deney yaklaşık olarak 2dk, en uzun deney ise yaklaşık olarak 73s 15dk sürmüştür.

Yapılan deneysel çalışmalardan elde edilen sonuçlardan en dikkat çekici durumlar;

- Problemin boyutu yüksek olduğundan doğrusal çekirdek, Gauss Radyal çekirdeğine göre oldukça yüksek bir performans ortaya koymuştur.
- Çalışmada kullanılan problem verileri için MLPK ve TPPK pairwise çekirdekleri benzer performanslar göstermişlerdir.
- Eğitim verilerinin sayısı arttıkça doğruluk oranı daha yüksek bir sınıflandırma gerçekleşmiştir.
- Test verilerinin sayısı eğitim verilerinin sayısından daha yüksek olmaya başladığında sınıflandırmaların doğruluk oranları düşmeye başlamıştır.
- Birbirinden tamamıyla bağımsız eğitim ve test verilerinde sınıflandırmanın doğruluk oranları düşmüştür.  
ve son olarak,
- Çekirdek parametreleri mutlaka deneysel sonuçlarla probleme göre belirlenmelidir.

Daha sonra benzer çalışmalar yapacak araştırmacıların bu tez çalışmasının sonuçlarını iyi irdelemeleri yapacakları çalışmalarda kolaylık sağlayacaktır.

Eğer yapılacak çalışmanın verilerinin boyutu çok yüksekse en iyi sonuçlar doğrusal çekirdek fonksiyonu yardımıyla alınacaktır. Esas araştırmaya geçmeden veri seti için optimal çekirdek parametreleri belirlenmelidir.

## 6 KAYNAKLAR

- Amasyalı, M. Fatih, and Aytunç Beken (2009). Measurement of turkish word semantic similarity and text categorization application. Signal Processing and Communications Applications Conference, *SIU, IEEE* **17:1** 2009.
- Bellman, Richard, et al. (1970). Introduction to matrix analysis. *Nonlinear Analysis and Applications*, **1-2: 1-58**, Vol. **10**. New York: McGraw-Hill.
- Ben-Hur, Asa, and Jason Weston.(2010) A users guide to support vector machines. *Data mining techniques for the life sciences. Humana Press.* 223-239.
- Bierens, Herman J., and Hosin Song. (2012). Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method. *Journal of Econometrics*, **168.1**, 108-119.
- A. Bordes and L. Bottou. (2005). The Huller: a simple and efficient online SVM. In Proceedings of the 16th European Conference on Machine Learning (ECML2005), Lecture Notes in Artificial Intelligence, to appear. *Springer*.
- Bordes, Antoine, et al. (2005). Fast kernel classifiers with online and active learning. *The Journal of Machine Learning Research* **6** : 1579-1619.
- Bousquet, Olivier, Stphane Boucheron, and Gbor Lugosi. Introduction to statistical learning theory. Advanced Lectures on Machine Learning. *Springer Berlin Heidelberg*, 169-207.
- Burges, Christopher JC.(1998) A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* **2.2**, 121-167.
- Buskes, G. J. H. M., and A. Van Rooij.(2000) Almost f-algebras: commutativity and the Cauchy-Schwarz inequality. *Positivity* **4.3** : 227-231.
- Callebaut, D. K. (1965). Generalization of the Cauchy-Schwarz inequality. *Journal of mathematical analysis and applications*, **12.3** : 491-494.

- Cortes, C., and Vapnik, V. (1995). Support-vector networks, *Machine learning*, **35(1-2)**: 3–22.
- Erastö, Panu.(2001) Support Vector Machines-Backgrounds and Practice. University of Helsinki, Rolf Nevanlinna Institute, Faculty of Science.
- Evgeniou, Theodoros, Massimiliano Pontil, and Tomaso Poggio (2000). Regularization networks and support vector machines. *Advances in Computational Mathematics* **13.1**, 1-50.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7(2)**, 179-188.
- Fujii, Masatoshi, et al. Operator inequalities related to Cauchy-Schwarz and Hölder-McCarthy inequalities. *Nihonkai Mathematical Journal* **8.2** (1997): 117-122.
- Gretton, Arthur, et al. (2008). A kernel method for the two-sample problem. arXiv preprint arXiv:0805.2368
- Gunn, Steve R. (1998). Support vector machines for classification and regression. *ISIS technical report* :**14**.
- Herbrich, Ralf. (2002) Learning kernel classifiers: Theory and Algorithms. *Mit Press*.
- Jódar, J. (1996). Hermite matrix polynomials and second order matrix differential equations. *Approximation Theory and its Applications* **12.2**, 20-30.
- Joachims, Thorsten. (1998) Text categorization with support vector machines: Learning with many relevant features. *Springer Berlin Heidelberg*, 1998.
- Kashima, Hisashi, et al. (2009) On pairwise kernels: An efficient alternative and generalization analysis. Advances in Knowledge Discovery and Data Mining. *Springer Berlin Heidelberg*. 1030-1037.
- Kashima, Hisashi, et al. (2010). Cartesian Kernel: An Efficient Alternative to the Pairwise Kernel. *IEICE Transactions on Information and Systems* **93.10**, 2672-2679.
- Lang, Ken. (2004). The 20 newsgroups data set.

- Liu, Qingshan, Hanqing Lu, and Songde Ma. (2004). Improving kernel Fisher discriminant analysis for face recognition. *Circuits and Systems for Video Technology. IEEE Transactions on Information and Systems* **14.1**, 42-49.
- Park, Sang-Hyeun, and Johannes Fürnkranz. (2007). Efficient pairwise classification. *Machine Learning. Springer Berlin Heidelberg*, 658-665.
- Rapaport, F., Zinovyev, A., Dutreix, M., Barillot, E., and Vert, J. P. (2007). Classification of microarray data using gene networks. *BMC Bioinformatics* **8.1** : **35**.
- Sabuncuoğlu, A. (2004). Lineer Cebir. *Nobel Yayın Dağıtım*, **2. Baskı**, 192:37 Yenışehir, Ankara.
- Schölkopf, Bernhard, and Alexander J. Smola. (2002). Learning with kernels: support vector machines, regularization, optimization and beyond. *the MIT Press*, 2002.
- Scholkopf, Bernhard. (2001). The kernel trick for distances. *Advances in neural information processing systems*. 301-307.
- Trevor, Hastie and Robert Tibshirani. (1998). Classification by pairwise coupling. *The annals of statistics* **26.2**, 451-471.
- Tong, Simon, and Daphne Koller (2001). Active learning for structure in Bayesian networks. International joint conference on artificial intelligence. *Lawrence Erlbaum Associates LTD. Vol. 17. No. 1*.
- Vapnik Vladimir, N. (1998). Statistical learning theory. New York: *Wiley*.
- Vapnik, V. (2002). The Nature of Statistical Learning Theory. *Data mining and knowledge discovery* **6**: 1-47.
- Vapnik, Vladimir, Esther Levin, and Yann Le Cun. (1994). Measuring the VC-dimension of a learning machine. *Neural Computation* **6.5**: 851-876.
- Vapnik, Vladimir N., and A. Ya Chervonenkis. (1982). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory of Probability and Its Applications* **26.3** : 532-553.

Vert, Jean-Philippe, Jian Qiu, and William Noble.(2007) A new pairwise kernel for biological network inference with support vector machines. *BMC bioinformatics* **8.Suppl 10, S8**.

Villard, Gilles. (1996). Computing Popov and Hermite forms of polynomial matrices. Proceedings of the 1996 international symposium on Symbolic and algebraic computation. ACM.

Walgampaya, C. K., and M. Kantardzic. (2006). Cost-Sensitive Analysis in Multiple Time Series Prediction. *Conference on Data Mining DMIN*. **Vol. 6**.

Zhou, Dengyong, et al. (2004). Learning with local and global consistency. *Advances in neural information processing systems* **16**.753760, 284.

### **İnternet Kaynakları:**

- 1) <http://qwone.com/~jason/20Newsgroups> (2008). Home Page.
- 2) <http://leon.bottou.org/projects/lasvm> (2005).



## ÖZGEÇMİŞ

Adı SOYADI : Üzeyir FİDAN  
Doğum Yeri ve Tarihi : Denizli, Türkiye, 10.02.1986  
Yabancı Dili : İngilizce  
İletişim : uzeyirfidan@hotmail.com

### **Eğitim Durumu**

Lise : Manisa Fatih Anadolu Lisesi, 2004  
Lisans : Afyon Kocatepe Üniversitesi, Matematik ABD, 2011  
Yüksek Lisans : Afyon Kocatepe Üniversitesi, İstatistik ABD, 2013