

Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi

Sevgi AYHAN

Arş. Gör. Dr., Eskişehir Osmangazi Üniversitesi,
Fen Edebiyat Fakültesi, İstatistik Bölümü
sayhan@ogu.edu.tr

Şenol ERDOĞMUŞ

Prof. Dr., Eskişehir Osmangazi Üniversitesi
Fen Edebiyat Fakültesi, İstatistik Bölümü
senol@ogu.edu.tr

Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi

Özet

Veri madenciliğinin görevlerinden biri olan sınıflandırma probleminin çözümü için geliştirilmiş önemli makine öğrenimi algoritmalarından biri Destek Vektör Makineleri'dir. Literatürde Destek Vektör Makineleri'nin diğer birçok tekniğe göre daha başarılı sonuçlar verdiği kanıtlanmıştır. Destek Vektör Makineleri'nin uygulanması sürecinde çekirdek fonksiyonu seçimi ve parametre optimizasyonu önemli rol oynamaktadır. Bu çalışmada, çekirdek fonksiyonu seçim süreci rassal blok deney tasarımı temeline oturtulmuştur. Çekirdek fonksiyonun seçiminde tek değişkenli varyans analizinden (Univariate ANOVA) yararlanılmıştır. Sonuç olarak en başarılı performansa sahip çekirdek fonksiyonunun radyal tabanlı fonksiyon olduğu kanıtlanmıştır.

Anahtar Kelimeler: Veri madenciliği, Destek vektör makineleri, Çekirdek fonksiyonu seçimi, Rassal blok deney tasarımı, Tek değişkenli varyans analizi

Kernel Function Selection for the Solution of Classification Problems via Support Vector Machines

Abstract

One of the most important machine learning algorithms developed for to accomplish classification task of data mining is Support Vector Machines. In the literature, Support Vector Machines has been shown to outperform many other techniques. Kernel function selection and parameter optimization play important role in implementation of Support Vector Machines. In this study, Kernel function selection process was ground on the randomized block experimental design. Univariate ANOVA was utilized for kernel function selection. As a result, the research proved that radial based Kernel function was the most successful Kernel function was proved.

Keywords: Data Mining, Support Vector Machines, Kernel Function Selection, Randomized Block Experimental Design, Univariate ANOVA

1. Giriş

Makine öğrenimi ve veri madenciliği literatüründe, sınıflandırma probleminin çözümüne ilişkin yapılan çalışmalar önemli yer tutmaktadır. Özellikle, bankacılık ve sigortacılık (riskli gruptaki müşterilerin tahmin edilmesi), tıp (hastalık teşhisi), biyoloji (canlı türlerinin sınıflandırılması), kimya (belirli bir hastalık için ilacın etkilerinin belirlenmesi), sosyal medya (spamlerin saptanması), endüstriyel üretim sistemleri (ortaya çıkan kusurlu ürünlerin belirlenmesi) gibi alanlarda sınıflandırma problemleriyle sıkça karşılaşmaktadır. Dolayısıyla, son yıllarda sınıflandırma problemlerinin çözümü, makine öğreniminin önemli çalışma alanlarından biri olmuştur.

Sınıflandırma problemlerinin çözümü için geliştirilen makine öğrenimi algoritmasının seçiminde dikkat edilecek en önemli kriterlerden biri, algoritmanın genelleme performansısıdır. Genelleme performansı, eğitim verisi, bağımsız niteliklerin sayısı/yapısı, model seçimi ve parametre seçimi gibi faktörlere bağlıdır. Tüm bu faktörler göz önünde bulundurulduğunda, veriden hem gizli hem de anlamlı enformasyonun çıkarılması ve doğru bilgiye ulaşma, algoritmanın genelleme başarısıyla doğru orantılıdır. Diğer bir deyişle, algoritmanın genelleme performansı ne kadar iyiyse elde edilen enformasyon da o kadar gerçekçi olacaktır.

Son yıllarda, sınıflandırma problemlerinin çözümü için geliştirilmiş en başarılı makine öğrenimi algoritmalarından biri Destek Vektör Makineleri'dir. Destek Vektör Makineleri, birçok sınıflandırma probleminin çözümünde başarıyla uygulanmış ve genelleme performansı yüksek ve etkin makine öğrenimi algoritmalarından biri olarak literatürdeki yerini almıştır.

Destek Vektör Makineleri'nin en önemli avantajı, sınıflandırma problemini kareli optimizasyon problemine dönüştürüp çözmesidir. Böylece problemin çözümüne ilişkin öğrenme aşamasında işlem sayısı azalmakta ve diğer teknik/algoritmalara göre daha hızlı çözüme ulaşılmaktadır (Osowski, Siwekand ve Markiewicz, 2004). Teknik bu özelliğinden dolayı, özellikle büyük hacimli veri setlerinde büyük avantaj sağlamaktadır. Ayrıca optimizasyon temelli olduğundan sınıflandırma performansı, hesaplama karmaşıklığı ve kullanılabilirlik açısından diğer tekniklere göre daha başarılıdır (Nitze, Schulthess ve Asche, 2012).

Çeşitli veri setleri için sınıflandırma probleminin çözümüne ilişkin Destek Vektör Makineleri'nin uygulanması sürecinde çekirdek fonksiyonu seçimi ve parametre optimizasyonu önemli rol oynamaktadır. Literatürdeki uygulamalarda genellikle daha iyi sonuçlar verdiği düşüncesi ile radyal tabanlı çekirdek fonksiyonun kullanıldığı görülmüştür. Ancak bu yaklaşımla, uygulamada kullanılan veri seti için daha iyi performansa sahip başka bir çekirdek fonksiyonu yok mu? sorusu ortaya çıkmaktadır. Dolayısıyla tek bir çekirdek fonksiyonu ile Destek Vektör Makineleri'nin performansına ilişkin genel bir yargıya varmak oldukça güçtür.

Birçok çalışmada DVM'nin uygulanması sürecinde çekirdek fonksiyonu seçimi için literatürdeki çalışmalar dikkate alınmış ve radyal tabanlı fonksiyonun kullanıldığı görülmüştür. Ancak radyal tabanlı çekirdek fonksiyonun her alandaki problemin çözümü için uygun olup olmadığı bilgisi bulunmamaktadır. Çekirdek fonksiyonlarının performansı veri setlerine ve probleme göre değişkenlik gösterebilmektedir.

Bu nedenlerle çalışma;

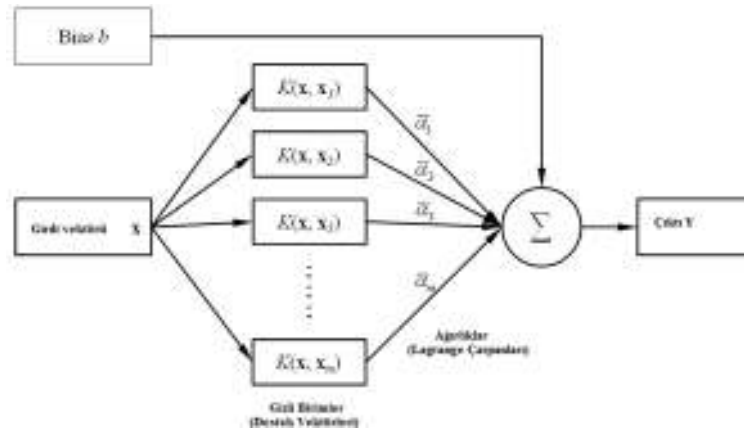
“H1: DVM'nin sınıflandırma performansları üzerinde veri setlerinin istatistiksel olarak anlamlı bir etkisi vardır.”

“H2: DVM'nin sınıflandırma performansları bakımından çekirdek fonksiyonları arasında istatistiksel açıdan anlamlı bir farklılık vardır. “ olmak üzere iki araştırma hipotezi temeline oturtulmuş ve hipotezlerin gerçekleşip gerçekleşmediği araştırılmıştır. Sınıflandırma probleminin çözüm süreci istatistiksel tekniklerle desteklenmiştir.

2. Destek Vektör Makineleri

Destek Vektör Makineleri (DVM), yapısal risk minimizasyonu prensibine göre çalışan dış bükey optimizasyona dayalı makine öğrenmesi algoritmalarıdır. Söz konusu algoritma, veriye ilişkin herhangi bir birleşik dağılım fonksiyonu bilgisine ihtiyaç duymadığı için dağılımdan bağımsız öğrenme algoritmalarıdır (Soman, Loganathan ve Ajay, 2011).

DVM, örüntü tanıma ve sınıflandırma problemlerinin çözümü için Vapnik tarafından geliştirilmiştir (Cortes ve Vapnik, 1995). DVM'in temelleri istatistiksel öğrenme teorisine diğer bir ifadeyle Vapnik-Chervonenkis (VC) teorisine dayanmaktadır (Li, Li, Li, Shyr, Xie ve Li, 2009). Şekil 1'de DVM'nin ağ yapısı verilmiştir.



Şekil 1. DVM'nin Genel Yapısı

Şekil 1’de verilen ağ yapısı incelendiğinde, $K(x_i, x_j)$ çekirdek fonksiyonlarını ve α Lagrange çarpanlarını göstermektedir. Çekirdek fonksiyonları yardımıyla girdilerin iç çarpımları hesaplanmaktadır. Lagrange çarpanları ise ağırlıkları göstermektedir. DVM’de bir örneğe ilişkin çıktı değeri, girdilerin iç çarpımları ile Lagrange çarpanlarının bağımsız kombinasyonlarının toplamına eşittir.

DVM’de amaç, sınıfları birbirinden ayıracak optimal ayırma hiper düzleminin elde edilmesidir. Başka bir ifadeyle, farklı sınıflara ait destek vektörleri arasındaki uzaklığı maksimize etmektir.

DVM iki sınıflı ve çok sınıflı sınıflandırma probleminin çözümü için geliştirilmiş makine öğrenmesi algoritmalarıdır. Çalışmada 2-sınıflı sınıflandırma problemi üzerine odaklanıldığından, bu bölümde iki sınıflı DVM’nin matematiksel yapısı açıklanmıştır.

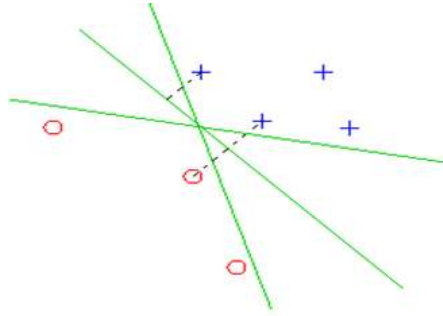
DVM, veri setinin doğrusal olarak ayrılıp ayrılamama durumuna göre temel olarak ikiye ayrılmaktadır. Dolayısıyla çalışmanın bu bölümünde doğrusal ve doğrusal olmayan DVM olmak üzere iki kısımda ele alınmıştır.

2.1. Doğrusal Destek Vektör Makineleri

Her x_i örneği, p adet niteliğe sahip girdi, $y_i \in \{1, -1\}$ örneklerin ait olduğu sınıfı temsil eden çıktı ve $x \in \mathbb{R}^p$ yüksek boyutlu girdi vektörü olmak üzere; (x_i, y_i) ikililerinden oluşan n hacimli bir eğitim kümesi S verildiğinde, farklı sınıflara ait örnekleri birbirinden en iyi şekilde ayıracak,

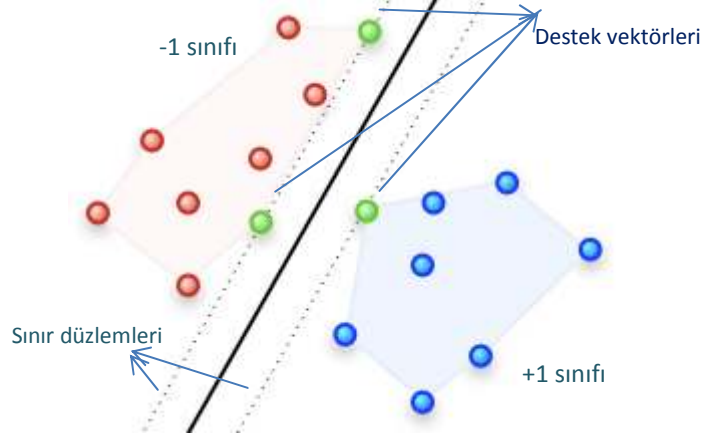
$$w \cdot x + b \quad (1)$$

doğrusal hiper düzleminin bulunmasına yardımcı olan denetimli öğrenme algoritmaları sınıfına ait makine öğrenimi algoritmasıdır (Soman vd., 2011). Burada w , hiper düzlemin normali aynı zamanda ağırlık vektörü ve b sabit olarak tanımlanmıştır.



Şekil 2. 2-Sınıflı Veri Setini Ayıran Farklı Düzlemlere İlişkin Örnekler

Veri setini ayıran düzlemlere ilişkin geometrik gösterim Şekil 2’de verilmiştir. Şekilde görüldüğü gibi farklı sınıflara ait örnekleri birbirinden ayıran birçok doğrusal düzlem bulunabilir. Ancak DVM, farklı sınıflara ait destek vektörleri arasındaki uzaklığı maksimize eden ayırma hiper düzleminin bulunmasını amaçlar



Şekil 3. DVM için 2-Sınıflı Problem Örneği

2-sınıflı ve iki boyutlu bir sınıflandırma problemi için doğrusal DVM’nin geometrik gösterimi Şekil 3’te verilmiştir. Destek vektörleri, ayırma hiper düzlemine en yakın olan her iki sınıfa da ait örnekler olarak ifade edilir ve Şekil 3’te gösterilmiştir. Söz konusu destek vektörleri, ait olduğu sınıfın sınırını belirler ve ayırma hiper düzlemine paralel bir düzlem üzerinde yer alır (Burges, 1998). Destek vektörlerinin üzerinde bulunduğu ve kesikli çizgilerle gösterilmiş düzlemlere sınır düzlemleri denir. Sınır düzlemlerinin tam ortasından geçen ve her iki düzleme de eşit uzaklıkta bulunan düzlem ise hiper düzlem olarak ifade edilir.

Doğrusal DVM veri setinin doğrusal ayrılma ve belirli bir hata ile doğrusal ayrılma durumuna göre ikiye ayrılmaktadır. İzleyen kısımda doğrusal ayrılma ve belirli bir hata ile doğrusal ayrılma durumlarına göre DVM’nin işleyişi açıklanmıştır.

2.1.1. Doğrusal Ayrılma Durumu

Eğitim veri setinin doğrusal olarak ayrılabilme durumunda DVM, en büyük sınıra sahip ayırma hiper düzlemini bulmaya çalışır. Söz konusu ayırma hiper düzleminin bulunabilmesi için veri setindeki tüm örneklerin,

$$f(x_i) = \langle w, x_i \rangle + b \geq +1, \quad y_i = +1 \quad (2)$$

$$f(x_i) = \langle w, x_i \rangle + b \leq -1, \quad y_i = -1 \quad (3)$$

eşitsizliklerini sağlaması gerekir (Soman vd., 2011). Bu eşitsizlikler, Eşitlik 4'te verildiği gibi tek bir eşitsizlikte birleştirilebilir.

$$\forall i \text{ için } y_i (\langle w, x_i \rangle + b) - 1 \geq 0 \quad (4)$$

Bir hiper düzlem, w normali ve orijinden dik uzaklığı $|b|/\|w\|$ olan, $w \cdot x + b = 0$ düzlemdir (Schölkopf ve Smola, 2002).

Eşitsizlik 2 ve 3 dikkate alındığında, sırasıyla, w normali ve orijinden dik uzaklığı $|1 - b|/\|w\|$ olan $H_1 = \langle w, x_i \rangle + b = +1$ düzlemi ile w normali ve orijinden dik uzaklığı $|-1 - b|/\|w\|$ olan $H_2 = \langle w, x_i \rangle + b = -1$ düzlemi paralel düzlemlerdir. Dolayısıyla H_1 ve H_2 sınır düzlemleri hiper düzleme eşit uzaklıkta yer almaktadır. H_1 ve H_2 sınır düzlemleri arasında herhangi bir eğitim örneği yer almamaktadır (Burges, 1998). Ancak, düzlemlerin üzerinde var olan eğitim örnekleri destek vektörleridir ve hiper düzleme en yakın olan eğitim örnekleridir. Ayırma hiper düzlemi, her iki sınıfın destek vektörleri arasındaki uzaklığı diğer bir ifadeyle sınırı maksimum yapan ve sınırın ortasından geçen düzlemdir. Bir eğitim örneğinin hiper düzleme uzaklığı

$$d(w, b; x) = \frac{|\langle w, x_i \rangle + b|}{\|w\|} \quad (5)$$

olmak üzere, sınır değeri (ρ) Eşitlik 6'da verilmiştir (Gunn, 1998).

$$\begin{aligned} \rho(w, b) &= \min_{xy=-1} \left(\frac{|\langle w, x_i \rangle + b|}{\|w\|} \right) + \min_{xy=+1} \left(\frac{|\langle w, x_i \rangle + b|}{\|w\|} \right) \\ &= \frac{1}{\|w\|} (\min_{xy=-1} |\langle w, x_i \rangle + b| + \min_{xy=+1} |\langle w, x_i \rangle + b|) = \frac{2}{\|w\|} \end{aligned} \quad (6)$$

Burada $\|w\|$, ağırlık vektörü olarak adlandırılan w normal düzleminin normudur. Dolayısıyla hiper düzleme en yakın olan örneklerin hiper düzleme olan ters uzaklığı sahip olduğu ağırlık vektörünün normuna eşit olmak zorundadır (Gunn, 1998; Schölkopf ve Smola, 2002). Bu teoremde yola çıkarak, eğitim örneklerini en iyi şekilde ayıran hiper düzlem,

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (7)$$

eşitliğini en küçükleyen düzlemdir. Eşitlik 7'nin en küçüklenmesi, probleme ilişkin VC boyutunun üst sınırının en küçüklenmesi anlamına gelmektedir. Böylece VC boyutunun en küçüklenmesiyle modelin yanlış sınıflandırma olasılığı da düşürülecektir (Soman vd., 2011).

Bu bilgiler ışığında, Eşitlik 7'deki en küçükleme tipindeki optimizasyon problemi Eşitlik 4'teki kısıt altında kareli optimizasyon problemi olarak Eşitlik 8'de verildiği gibi formüle edilmektedir (Cortes ve Vapnik, 1995; Fletcher, 2009).

$$\min \frac{1}{2} \|w\|^2$$

$$\forall i \text{ için } y_i (\langle w, x_i \rangle + b) - 1 \geq 0 \quad (8)$$

Eşitlik 8'deki optimizasyon probleminin çözülmesi sonucunda, sınıflara ait destek vektörleri arasındaki uzaklığı maksimize edecek optimal ayırma hiper düzlemi elde edilir (Gunn, 1998; Cortes and Vapnik, 1995).

DVM'de optimal ayırma hiper düzleminin bulunması için kareli optimizasyon probleminin çözümünde Lagrange çarpanlarından yararlanır. Lagrange çarpanları en küçükleme tipindeki problemi dual probleme dönüştürerek problemin daha kolay çözülmesine imkan verir (Gunn, 1998). Problemin çözümünde kullanılan Lagrange fonksiyonu Eşitlik 9'da verilmiştir.

$$L_p(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (\langle w, x_i \rangle + b)] + \sum_{i=1}^n \alpha_i \quad (9)$$

Bu eşitliklerde $\alpha_i \geq 0$ olmak üzere, her bir α_i Lagrange çarpanı olarak ifade edilir. L_p , w ağırlık vektörü ve b sabitini en küçükleyen ve negatif olmayan dual değişken α_i 'yi en büyükleyen bir fonksiyondur (Burges, 1998).

Lagrange fonksiyonunun w ve b 'ye göre kısmi türevler alınarak Eşitlik 10 ve 11'deki Karush-Kuhn-Tucker (KKT) koşulları elde edilir.

$$\frac{\partial L_p(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^n y_i \alpha_i x_i = 0 \Rightarrow w = \sum_{i=1}^n y_i \alpha_i x_i \quad (10)$$

$$\frac{\partial L_p(w, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^n y_i \alpha_i = 0 \quad (11)$$

Elde edilen eşitlikler Lagrange fonksiyonunda ilgili yerlere koyularak problem en büyükleme tipindeki dual Lagrange problemine ($L_D(\alpha)$) dönüşür. Söz konusu problemle ilişkin model Eşitlik 12’de gösterilmiştir.

$$\begin{aligned} \text{Max } L_D(\alpha) = L_D(w, \alpha, b) &= \sum_{i=1}^n \alpha_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s.t. } \sum_{i=1}^n \alpha_i y_i &= 0 \quad \text{ve} \quad \forall \alpha_i \end{aligned} \quad (12)$$

Optimal hiper düzlemin belirlenmesi için, Eşitlik 12’de verilen model çözülerek dual Lagrange $L_D(\alpha)$ ’yı maksimum yapan α_i değerleri elde edilir. α_i Lagrange çarpanlarından sıfırdan büyük değer alan eğitim örnekler “destek vektörleri” olarak ifade edilir. Optimal ayırma hiper düzlemi, sıfırdan büyük değer alan bu Lagrange çarpanları ile belirlenir (Gunn, 1998; Burges, 1998).

α_i ’nin çözümü ile optimal hiper düzlemin Eşitlik 13 ve 14’te verilen ağırlık vektörü w ve b sabit parametreleri belirlenir.

$$w^* = \sum_{i=1}^n y_i \alpha_i x_i \quad (13)$$

$$b^* = -\frac{1}{2} \langle w^*, x_i \rangle \quad (14)$$

Sonuç olarak elde edilen hiper düzleme bağlı olarak sınıflandırıcı Eşitlik 15’te verilmiştir.

$$f(x) = \text{sgn}(\langle w^*, x_i \rangle + b^*) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_i \langle x_i, x_j \rangle \right) \quad (15)$$

Ayrıca Eşitlik 15’teki sınıflandırıcıya alternatif olarak, algoritmanın gerçek değerler üretmesinden dolayı Eşitlik 16’da verilen sınıflandırıcının kullanımı daha uygun olmaktadır (Gunn, 1998).

$$f(x) = h(\langle w^*, x_i \rangle + b^*) \Rightarrow h(z) = \begin{cases} -1 & z < -1 \\ z & -1 \leq z \leq +1 \\ +1 & z > +1 \end{cases} \quad (16)$$

2.1.2. Belirli Bir Hata İle Doğrusal Ayrılma Durumu

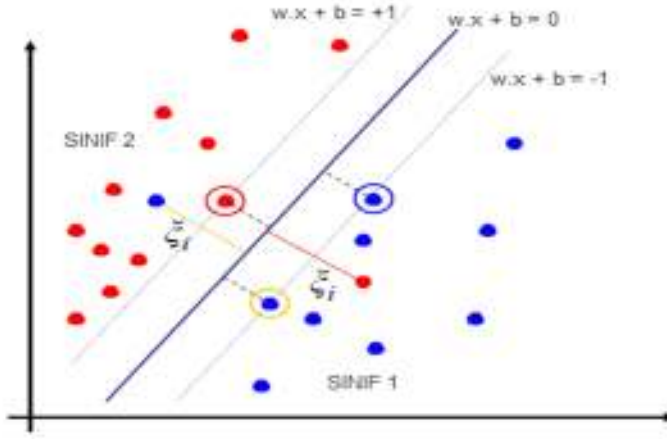
Veri setinin belirli bir hata ile doğrusal olarak ayrılma durumu, veri setinin gürültülü veri içermesi, çok boyutlu olması veya karmaşık yapısından kaynaklanmaktadır (Li vd., 2009). Belirli bir hata ile doğrusal ayrılma durumunda iki sınıflı veri setini ayırmak için gevşek sınır (soft margin) yaklaşımı kullanılmaktadır.

Bu duruma ilişkin DVM'nin geometrik yapısı Şekil 4'te gösterilmiştir. Şekilden de görüldüğü üzere, gevşek sınır yaklaşımında modele, bir örneğin yanlış sınıflandırılması durumunda ait olduğu karar sınırına olan uzaklığının ölçüsü olan ξ_i ayak değişkeni (Cortes and Vapnik, 1995; Burges, 1998) eklenir.

Söz konusu durum için ayırma hiper düzleminin bulunabilmesi için veri setindeki tüm örneklerin 17 ve 18'deki eşitsizlikleri sağlaması gerekir (Cortes and Vapnik, 1995).

$$f(x_i) = \langle w, x_i \rangle + b \geq +1 - \xi_i, \quad y_i = +1 \quad (17)$$

$$f(x_i) = \langle w, x_i \rangle + b \leq -1 + \xi_i, \quad y_i = -1 \quad (18)$$



Şekil 4. 2-Sınıflı Problem İçin Belirli Bir Hata İle Doğrusal Ayrılabilme Durumu

Yanlış sınıflandırma olasılığını düşürmek için doğrusal ayrılma durumundaki dönüşümler yapılarak problem, Eşitlik 19'da verilen kareli optimizasyon problemine dönüşür (Cortes ve Vapnik, 1995; Schölkopf ve Smola, 2002).

$$\text{Min} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (19)$$

$$\text{s.t. } y_i (\langle w, x_i \rangle + b) - 1 + \xi_i \geq 0, \quad \xi_i \geq 0$$

Modeldeki C katsayısı, Lagrange çarpanının alabileceği üst sınır değerini gösteren ceza parametresini ifade etmektedir. Lagrange çarpanı $\alpha_{(i)}$ C ceza parametresine eşit olması durumunda destek vektörleri ayırma hiper düzlemi üzerinde yer almaktadır (Katagiri ve Abe, 2006).

Doğrusal ayrılabilme durumunda olduğu gibi Eşitlik 19'daki optimizasyon problemine ilişkin modelin çözülmesi sonucunda, sınıflara ait destek vektörleri arasındaki uzaklığı maksimize edecek optimal ayırma hiper düzlemi elde edilir. Bu süreçte kareli optimizasyon probleminin çözümünde Eşitlik 20'de verilen Lagrange fonksiyonundan yararlanılır (Cortes ve Vapnik, 1995).

$$L_D(w, b, \alpha, r, \xi) =$$

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (\langle w, x_i \rangle + b) - 1 + \xi_i] - \sum_{i=1}^n r_i \xi_i \quad (20)$$

Eşitlik 12'deki Lagrange fonksiyonundan farklı olarak burada her bir r_i , ξ_i 'nin pozitif değer almasını garanti eden Lagrange parametreleridir (Demirci, 2007). Lagrange fonksiyonunun w , b ve ξ_i değişkenlerine göre kısmi türevler alınarak Karush-Kuhn-Tucker (KKT) koşulları elde edilir. Elde edilen eşitlikler Lagrange fonksiyonunda ilgili yerlere koyularak problem en büyükleme tipindeki dual Lagrange problemine ($L_D(\alpha)$) dönüşür. Söz konusu probleme ilişkin,

$$\begin{aligned} \text{Max } L_D(\alpha) = L_D(w, b, \alpha, r, \xi) &= \sum_{i=1}^n \alpha_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s.t. } \sum_{i=1}^n \alpha_i y_i &= 0 \quad \text{ve} \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (21)$$

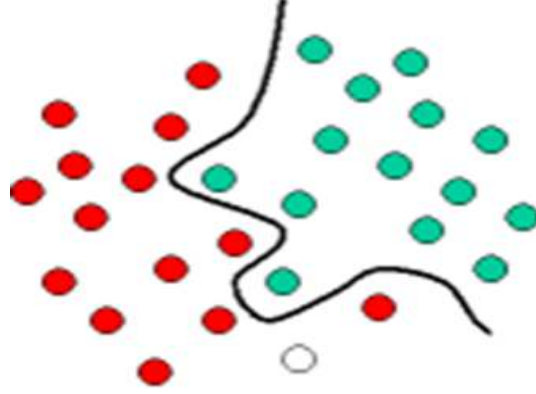
modelinin çözümü sonucunda elde edilen hiper düzleme bağlı olarak elde edilen sınıflandırıcı Eşitlik 22'de verilmiştir (Burges, 1998).

$$f(x) = \text{sign}(w^*, x_i + b^*) = \text{sign}\left(\sum_{i=1}^n y_{(i)} \alpha_{(i)} \langle x_{(i)}, x_{(j)} \rangle\right) \quad (22)$$

2.2. Doğrusal Olmayan Destek Vektör Makineleri

Doğrusal olmayan DVM, veri setinin doğrusal bir fonksiyonla tam veya belirli bir hata ile ayrılamaması durumunda kullanılan algoritmalardır. Gerçek yaşam problemlerinde bir veri setinin hiper düzlem ile doğrusal olarak ayrılması çoğunlukla mümkün değildir. Dolayısıyla sınıfları ayırma işlemi, ayırma eğrisinin tahmin edilmesiyle mümkün olmaktadır. Ancak uygulamada eğrinin tahmin edilmesi oldukça zordur.

Veri setinin doğrusal ayrılamama durumunun geometrik gösterimi Şekil 5'te verilmiştir. Bu durumda p-boyutlu girdi vektörü x 'in P-boyutlu özellik vektörü Φ 'ye dönüştürülmesi gerekmektedir (Cortes ve Vapnik, 1995).



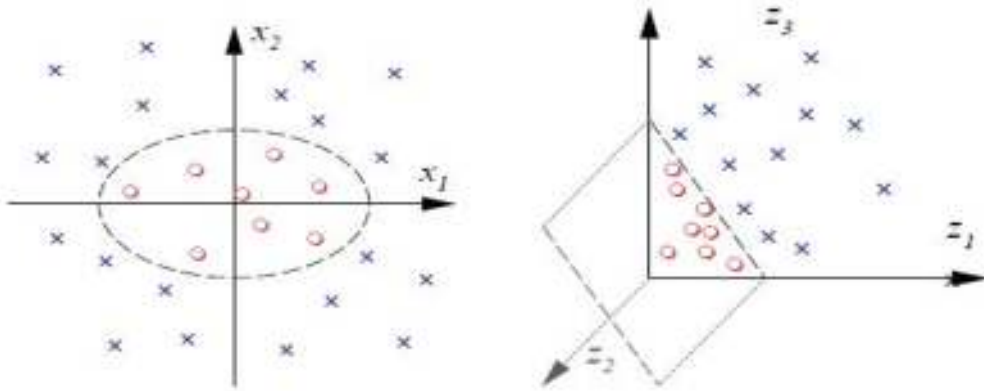
Şekil 5. 2-Sınıflı Problem İçin Doğrusal Ayrılamama Durumu

p-boyutlu girdi vektörü x 'in P-boyutlu özellik vektörü Φ 'ye dönüştürülebilmesi için optimal ayırma düzleminin özellik uzayında tanımlanabilmesi gerekir. Bu amacı gerçekleştirmek için doğrusal olmayan haritalama yaklaşımından yararlanır (Bussat, 2003).

“Doğrusal olmayan haritalama”, orijinal girdi uzayı x 'in bir Hilbert uzayı olan daha yüksek boyutlu F özellik uzayına dönüştürülerek doğrusal ayrımının gerçekleştirilmesi için kullanılan bir yaklaşımdır (Suykens, 2002). “Hilbert uzayı” pozitif skaler çarpıma sahip ve öğeleri fonksiyonlardan oluşan tam iç çarpım uzayları olarak ifade edilmektedir (Çakar, 2007).

Doğrusal olarak ayrılamayan iki boyutlu bir veri seti için doğrusal olmayan haritalama yaklaşımının geometrik olarak açıklaması Şekil 6'da verilmiştir. Doğrusal olmayan haritalama yaklaşımı ile iki boyutlu veri seti üç boyutlu özellik uzayına taşınarak veri setinin doğrusal ayrımı sağlanmıştır.

$$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3$$



Şekil 6. Doğrusal Olmayan Haritalama Yaklaşımı Örneği

Şekil 6'da, iki boyutlu girdi vektörü (x_1, x_2) ve üç boyutlu özellik uzayı (z_1, z_2, z_3) olmak üzere, fonksiyonların özellik uzayı,

$$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \quad (x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1 x_2, x_2^2) \quad (23)$$

eşitliği ile gösterilmektedir. Özellik uzayında haritalanmış girdi vektörlerinin iç çarpımları Eşitlik 24'teki gibi elde edilir (Soman vd., 2011).

$$= (x_1^2 x_1'^2 + 2x_1 x_1' x_2 x_2' + x_2^2 x_2'^2) \quad (24)$$

Böylece veri seti iki boyutlu uzaydan üç boyutlu uzaya taşınarak haritalama işlemi gerçekleştirilmiş olur (Soman vd., 2011). Sonuç olarak, doğrusal olmayan DVM için özellik uzayında tanımlı ayırma hiper düzlemine bağlı olarak sınıflandırıcı karar fonksiyonu,

$$f(x) = \text{sign}((w^0, x_i) + b^0) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_i (\phi(x_i) \cdot \phi(x_j)) \right) \quad (25)$$

eşitliği ile ifade edilir.

3. Uygulama

Sınıflandırma problemlerinin Destek Vektör Makineleriyle çözümü ve en iyi çekirdek fonksiyonunun seçimi için, literatürde en sık kullanılan 11 adet veri seti seçilmiştir. Bu veri setlerine UCI (Machine Learning Repository) makine öğrenimi veri tabanı sisteminden ulaşılmıştır. Söz konusu veri setleri farklı sınıflandırma algoritmalarının uygulanması ve performanslarının karşılaştırılmasında “bir kriter” olarak kabul edilmektedir (Huang ve Wang, 2006). Çalışmamızda 12. veri seti olarak Türkiye’deki özel bir bankanın kredi verileri kullanılmıştır.

Sonuç olarak çalışmada, dördü bankacılık, üçü tıp, birer adet bilgisayar sistemleri, fizik, kimya, biyoloji ve hukuk olmak üzere yedi farklı alandaki sınıflandırma problemlerine ilişkin veri setlerine yer verilmiştir.

Ayrıca literatürde birçok çekirdek fonksiyonu tanımlanmıştır. Ancak her çekirdek fonksiyonu Destek Vektör Makinelerinde kullanımı uygun olmamaktadır. Dolayısıyla, DVM’nin uygulanmasında çekirdek fonksiyonunun seçimi kritik rol oynamaktadır. Uygulamada, DVM için radyal tabanlı, polinomiyal, lineer ve sigmoid çekirdek fonksiyonları kullanılmaktadır.

Tablo 1. Çalışmada Kullanılan Veri Setleri Ve Özellikleri

No	Veri Seti	Örnek Sayısı	Nitelik Sayısı		Sınıf Sayısı
			Kategorik	Sayısal	
1	Australian Credit Approval-Statlog	690	8	6	2
2	Bank	4521	9	7	2
3	German Credit-Statlog	1000	20		2
4	Hearth Disease-Statlog	270	6	7	2
5	Ionosphere	351		34	2
6	Pima Indian Diabets (PIM)	768	-	8	2
7	Spambase (Spam)	4601	-	57	2
8	Wisconsin Breast Cancer-WBC-orijinal	699	-	10	2
9	Glass	214	-	9	6
10	Iris Plant (Iris)	150	-	4	3
11	Wine	178	-	13	3
12	Türkiye kredi verisi (Türkiye)	167	13	4	2

DVM algoritmasının uygulanmasında, çekirdek fonksiyonlarının optimal hiper parametre değerlerinin belirlenmesinde grid arama yöntemi kullanılmıştır. Yönte-

min uygulanmasında her bir parametre için alt sınır, üst sınır ve belirli bir aralık değeri belirlenir. Parametre değerleri sınır değerleri içinde belirlenen aralık kadar atlayarak her bir değer noktası için algoritmaya ilişkin bir sınıflandırma performansı belirler. En iyi sınıflandırma performansını veren parametre değerleri optimal hiper parametre değerleri olarak belirlenir. Lineer, radyal tabanlı (C,γ), polinomiyal (C,γ,α,d) ve sigmoid (C,γ,α) olmak üzere dört çekirdek fonksiyonu için çalışmamızda belirlenen parametre değer aralıkları Tablo 2’de verilmiştir. Söz konusu çekirdek fonksiyonlarına ilişkin C ceza parametresinin alt sınırı 0,0001 ve üst sınırı 5000 olarak belirlenmiştir. Parametrelere ilişkin artış aralıkları 1 olarak alınmıştır. Ancak bu aralık logaritmik artışı göstermektedir. Örneğin, alt sınır $2^{(-13)}$ (0,0001) değerinden başlarsa bir sonraki parametre değeri $2^{(-12)}$ (0,0002441406) olarak alınmaktadır. Diğer parametre değerleri için benzer şekilde yorumlanabilir.

Tablo 2. Grid Arama İçin Belirlenen Parametre Değer Aralıkları

<i>Parametreler</i>	<i>Alt Sınır</i>	<i>Üst Sınır</i>	<i>Aralık</i>
<i>C</i>	0.0001	5000	$[2^{-13}, 1, 2^{13}]$
<i>γ</i>	0.001	500	$[2^{-9}, 1, 2^9]$
<i>α</i>	0.0001	50	$[2^{-13}, 1, 2^6]$
<i>d</i>	1	3	1

“DVM’nin sınıflandırma performansları üzerinde veri setlerinin istatistiksel olarak anlamlı bir etkisi var mı?” ve “çekirdek fonksiyonları arasında DVM’nin performansları açısından farklılık var mı?” sorularına yanıt bulmak ve sonucunda, en iyi çekirdek fonksiyonun seçimini gerçekleştirmek amacıyla rassal blok deney tasarımı düzenlenmiştir. Tablo 3’te verilen deney tasarımı tablosundan da görüleceği gibi, 12 farklı veri seti bloklar olarak ele alınmış ve dört farklı çekirdek fonksiyonu için DVM’nin sınıflandırma performansları elde edilmiştir. Çizelgede sınıflandırma performansları “SP” ile gösterilmiştir. Böylece performanslar arasında anlamlı bir fark olup olmadığı ve fark anlamlıysa bu farkın, çekirdek fonksiyonlarından mı yoksa veri setlerinden mi kaynaklandığı araştırılmıştır.

Tablo 3. Çekirdek Fonksiyonlarının Performanslarının Karşılaştırılması İçin Oluşturulmuş Rassal Blok Deney Tasarımı

Rassal Bloklar Deney Tasarımı	Çekirdek Fonksiyonları			
	Linear	Radyal Tabanlı	Polinomiyal	Sigmoid
1	SP ₁₁	SP ₁₂	SP ₁₃	SP ₁₄
2	SP ₂₁	SP ₂₂	SP ₂₃	SP ₂₄
.
.
.
12	SP ₁₂₁	SP ₁₂₂	SP ₁₂₃	SP ₁₂₄

Rassal blok deney tasarımı çalışmaları için analizlerde parametrik testlerden tek değişkenli varyans analizi (Univariate ANOVA) kullanılmıştır. Ancak parametrik testlerin uygulanabilmesi için verinin normal dağılıma uyması gerekmektedir. Algoritmaların sınıflandırma performansları bir olasılık değerini gösterdiğinden, bu değerlerin normallik varsayımını sağlaması beklenemez. Bu sıkıntıyı gidermek amacıyla literatürde, olasılık veya yüzde ile temsil edilen veriler için ArcSin (açı) dönüşümünün kullanılması önerilmektedir (Fernandez, 1992). ArcSin dönüşüm formülü Eşitlik 26'da verilmiştir.

$$\text{Dönüşüm değeri} = \theta = \sin^{-1}(\sqrt{sp}) \quad (26)$$

Dönüşüm değeri, performans değerinin karekökünün ArcSinüsüne eşittir. Sonuç olarak bu aşamada, elde edilen performans değerlerine ArcSin dönüşümü uygulanmış, böylece veriler analize hazır hale getirilmiştir. Ayrıca, en iyi çekirdek fonksiyonun seçimi ve en iyi indirgeme algoritmasının seçimi aşamalarının gerçekleştirilmesi için Tukey ikili karşılaştırma testinden yararlanılmıştır (Conagin, Barbin ve Demétrio, 2008). Tukey testi ile genelleme performansı en yüksek çekirdek fonksiyonunun belirlenmesi amaçlanmıştır.

4. Sonular

alıřmada DVM performanslarının elde edilmesinde, DVM iin zel olarak geliřtirilmiř ve parametre optimizasyonunda grid arama yntemini kullanan DTREG programı kullanılmıřtır. Ayrıca, alıřmada yer verilen istatistiksel analizler iin IBM SPSS 20 paket programı kullanılmıřtır. İstatistiksel testler % 95 gven dzeyine gre yapılmıřtır.

4.1. DVM Sınıflandırma Performansları

Bu kısımda, grid arama yntemi kullanılarak drt ekirdek fonksiyonu iin elde edilen optimal hiper parametre deęerleri ve DVM algoritmasının sınıflandırma performansları ayrıntılı olarak aıklanmıřtır. 12 veri seti iin drt ekirdek fonksiyonuna iliřkin DVM algoritmasının sınıflandırma performansları Tablo 4'te gsterilmiřtir.

Tablo 4'teki DVM performansları deęerlendirildięinde, Australian credit (% 88,52), German credit (% 79,40), Pima Indian diabetes (% 78,13), ionosphere (% 94,32), bank (% 89,60), Trkiye kredi (% 100) ve glass (% 95,46) eęitim verileri iin en iyi sınıflandırma performansına sahip ekirdek fonksiyonunun radyal tabanlı ekirdek fonksiyonu olduęu grlmektedir.

Lineer ekirdek fonksiyonu, hearth disease (% 87,40) ve spambase (% 94,60) eęitim veri setleri iin daha bařarılı sınıflandırma performansına sahiptir.

Iris eęitim veri seti iin lineer ve polinomiyal ekirdek fonksiyonları % 98 sınıflandırma doęruluyla en yksek sınıflandırma performansına sahip fonksiyonlar olarak belirlenmiřtir. Polinomiyal ve sigmoid ekirdek fonksiyonlarının Wisconsin breast ve wine eęitim veri seti iin sırasıyla % 97,66 ve % 100 sınıflandırma doęruluęu ile en iyi sınıflandırma performansına sahip ekirdek fonksiyonları olduęu grlmektedir.

Elde edilen sonulardan yola ıkararak, eęitim veri setleri iin radyal tabanlı ekirdek fonksiyonun genel olarak daha bařarılı sınıflandırma performansına sahip olduęu sylenebilir. Ancak ekirdek fonksiyonları iin DVM'nin sınıflandırma performansı algoritmanın genelleme bařarisına baęlıdır. Dolayısıyla hangi ekirdek fonksiyonun daha bařarılı sınıflandırma performansına sahip olduęunu syleyebilmek iin test verisi iin elde edilen sınıflandırma performanslarına bakmak daha doęrudur.

Tablo 4. Çekirdek Fonksiyonlarının DVM Performansları Ve Parametre Değerleri

Veri Setleri	Çekirdek fonksiyonları	Çekirdek fonksiyonlarına ilişkin parametre değerleri				Eğitim (%)	Test (%)	Ortalama (%)
		C	γ	Degree	Coef			
Austaralian Credit	Linear	0,1	-	-	-	86,20	86,20	86,2
	Radial	0,1	0,435603	-	-	88,52	86,47	87,495
	Polinomial	0,25763	0,03684	3	0,215444	87,77	87,09	87,43
	Sigmoid	0,39789	0,16681	-	0,02154	86,20	86,17	86,185
German Credit	Linear	0,39789	-	-	-	78,60	76,50	77,55
	Radial	41,01832	0,003327	-	-	79,40	76,90	78,15
	Polinomial	0,199474	0,003327	3	12,91549	78,80	76,70	77,75
	Sigmoid	25,06597	0,011081	-	0,46416	76,90	77,00	76,95
Glass	Linear	50	-	-	-	72,90	61,22	67,06
	Radial	12,56605	15,02665	-	-	95,46	73,37	84,415
	Polinomial	1,34876	4,516	3	4,64159	91,6	72,00	81,8
	Sigmoid	50	0,122583	-	0,21544	64,00	62,6	63,3
Hearth disease	Linear	0,103702	-	-	-	87,40	83,7	85,55
	Radial	0,79370	0,03684	-	-	86,67	85,19	85,93
	Polinomial	25,06595	0,01107	3	0,00464	85,93	84,07	85
	Sigmoid	0,19947	0,12258	-	0,02154	84,81	84,07	84,44
Pima Indian Diabet	Linear	0,383166	-	-	-	77,99	77,21	77,6
	Radial	1,64183	0,12258	-	-	78,13	77,73	77,93
	Polinomial	0,1	0,011072	3	10	77,99	76,95	77,47
	Sigmoid	25,06595	0,03684	-	1	77,99	77,08	77,535
Ionosphere	Linear	1	-	-	-	87,59	81,15	84,37
	Radial	1	0,1	-	-	94,32	91,30	92,81
	Polinomial	1	0,1	1	0,01	87,23	84,06	85,645
	Sigmoid	1	0,1	-	0,01	80,85	79,71	80,28
Iris Plant	Linear	1502,6652	-	-	-	98	96,7	97,35
	Radial	1,1071732	1,3572	-	-	97,3	97,3	97,3
	Polinomial	3,684032	50	1	0	98	98	98
	sigmoid	40,7886	0,12258	-	0,07743	97,3	97,3	97,3
Spambase	Linear	10	-	-	-	94,60	92,82	93,71
	Radial	10	0,01	-	-	93,45	94,00	93,725
	Polinomial	10	0,01	1	0,01	92,50	92,30	92,4
	Sigmoid	10	0,01	-	0,001	88,33	87,54	87,935
	Linear	0,199473	-	-	-	99,99	96,01	98

Wine	Radial	0,22901	0,4	-	-	99,99	97,20	98,595
	Polinomial	0,1	1,3572	3	0,0000	100	97,20	98,6
	Sigmoid	12,5660	0,12258	-	0,0000	100	98,3	99,15
Wisconsin	Linear	0,09629	-	-	-	97,22	96,93	97,075
Breast	Radial	0,397897	0,4078	-	-	97,22	97,07	97,145
Cancer (WBC)	Polinomial	25,06596	0,12258	1	1	97,66	97,07	97,365
	Sigmoid	0,1	0,40788	-	0,21544	97,66	97,66	97,66
	Linear	100	-	-	-	89,33	85,50	87,415
Bank	Radial	100	0,01	-	-	89,60	89,85	89,725
	Polinomial	100	0,01	1	0,001	89,33	85,50	87,415
	Sigmoid	100	0,01	-	0,001	85,82	82,60	84,21
Türkiye Kredi verisi	Linear	0,09321	-	-	-	97,01	86,83	91,92
	Radial	0,806234	0,15590	-	-	100	89,82	94,91
	Polinomial	0,06787	1,35721	1	0,0000	97,01	86,23	91,62
	Sigmoid	3,684	0,03684	-	4,64159	91,02	89,82	90,42

Test verileri için çekirdek fonksiyonlarının sınıflandırma performansları incelendiğinde, glass (% 73, 37), hearth disease (% 85, 19), Pima Indian diabet (% 77, 73), ionosphere (% 91,30), spambase (% 94,00), bank (% 89,85) ve Türkiye kredi (% 89,82) veri setleri için radyal tabanlı çekirdek fonksiyonunun en iyi performansa sahip çekirdek fonksiyonu olduğu gözlemlenmiştir.

Polinomiyal çekirdek fonksiyonun Australian credit (% 87,09) ve Iris (% 98) veri setinde daha başarılı sonuçlar verdiği görülmektedir. German credit (% 77), wine (% 98,3) ve WBC (% 97,66) veri setleri için sigmoid çekirdek fonksiyonunun daha başarılı sınıflandırma performansına sahip olduğu belirlenmiştir. Sonuç olarak hem eğitim hem test verisi için 12 veri setinin yedisinde radyal tabanlı çekirdek fonksiyonu daha başarılı sınıflandırma performansına sahiptir.

Eğitim ve test verileri için DVM'nin sınıflandırma performanslarının ortalaması dikkate alındığında wine ve WBC veri setleri için sigmoid çekirdek fonksiyonu, diğer 10 veri seti için radyal tabanlı çekirdek fonksiyonu en iyi sınıflandırma performansına sahip çekirdek fonksiyonları olduğu görülmüştür.

Sonuç olarak elde edilen sınıflandırma performansları incelendiğinde radyal tabanlı çekirdek fonksiyonun genel olarak daha başarılı sonuçlar verdiği kanısına varılabilir. Ancak sadece elde edilen sonuçlardan bu kanıya varmak sakıncalı olabilir. Dolayısıyla algoritmanın sınıflandırma performansları açısından çekirdek fonksiyonları arasında farklılık olup olmadığı konusunda kesin bir yargıya varabilmek için istatistiksel olarak test edilmesi gerekmektedir.

4.2. En İyi Çekirdek Fonksiyonunun Belirlenmesi

Çekirdek fonksiyonları arasında algoritmanın sınıflandırma performansı bakımından istatistiksel açıdan bir farklılık olup olmadığını belirlemek için,

“H₀: DVM'nin sınıflandırma performansları bakımından çekirdek fonksiyonları arasında istatistiksel açıdan anlamlı bir farklılık yoktur.” hipotezi test edilmiştir.

Ayrıca veri setlerinin, dört çekirdek fonksiyonu için DVM'nin sınıflandırma performansı üzerinde etkisi olup olmadığını belirlemek için,

“H₀: DVM'nin sınıflandırma performansları üzerinde veri setlerinin istatistiksel olarak anlamlı bir etkisi yoktur.” hipotezi kurulmuş ve test edilmiştir.

Her iki hipotezin test edilmesi amacıyla Tek Değişkenli Varyans Analizi ve parametrik olmayan karşılığı Friedman Sıralamalı İki Yönlü Varyans Analizi kullanılmıştır. Böylece, veri setlerinin sınıflandırma performansı üzerinde etkileri söz konusuysa bu etkiler yok edilerek, çekirdek fonksiyonlarının karşılaştırılması açısından daha sağlıklı sonuçlara ulaşılmıştır. Çekirdek fonksiyonlarının ve veri setlerinin algoritmanın sınıflandırma performansı üzerindeki etkilerinin belirlenmesine ilişkin tek değişkenli varyans analizi ve Friedman iki yönlü varyans analizi sonuçları Tablo 5'te verilmiştir.

Tablo 5. Sınıflandırma Performansları Üzerinde Çekirdek Fonksiyonlarının ve Veri Setlerinin Etkilerinin Belirlenmesine İlişkin Elde Edilen Analiz Sonuçları

Tek Değişkenli Varyans Analizi				
<i>Kaynak</i>	<i>Tip 3 K.T.</i>	<i>sd</i>	<i>F</i>	<i>p</i>
Model	1,678 ^a	14	59,689	0,000
Etkileşim etkisi	54,934	1	27355,085	0,000
Veri setleri	1,654	11	74,867	0,000
Çekirdek fonksiyonu	,024	3	4,036	0,015
Hata	,066	33		
Toplam	56,679	48		

^a R Squared = ,962 (Adjusted R Squared = ,946)

^{*} $p \leq 0,05$

Tek değişkenli varyans analizi sonuçları incelendiğinde modelin istatistiksel olarak anlamlı olduğu ($p=0,000 \leq 0,05$) bulunmuştur. Diğer bir deyişle, çekirdek fonksi-

yonları ve veri setlerinin DVM'nin sınıflandırma performansları üzerinde istatistiksel olarak anlamlı etkisi olduğu kanıtlanmıştır. Veri setleri ve çekirdek fonksiyonlarının ana etkilerine bakıldığında, çekirdek fonksiyonları ve veri setleri için p anlamlılık değerlerinin 0,05'ten küçük olduğu görülmektedir. Dolayısıyla, sınıflandırma performansları bakımından her bir veri seti ve çekirdek fonksiyonları kendi içinde istatistiksel olarak farklılık göstermektedir.

DVM'nin sınıflandırma performansları bakımından söz konusu farklılıkların hangi çekirdek fonksiyonları arasında olduğunu belirlemek için Tukey testinden yararlanılmıştır. Yapılan analizler sonucunda elde edilen sonuçlar Tablo 6'da özetlenmiştir.

DVM'nin sınıflandırma performansı bakımından radyal tabanlı çekirdek fonksiyonunun ($p \leq 0,05$) lineer ve sigmoid çekirdek fonksiyonlarından istatistiksel olarak farklılık gösterdiği kanıtlanmıştır. Ortalama fark sonuçlarına göre radyal tabanlı çekirdek fonksiyonunun sınıflandırma performansının diğer iki çekirdek fonksiyonuna göre daha başarılı olduğu söylenebilir.

Tablo 6. Çekirdek Fonksiyonları Arasındaki Farklılıkların Belirlenmesi İçin Elde Edilen Analiz Sonuçları

Çekirdek fonksiyonlarının İkili Karşılaştırmaları	Tukey Testi		
	Ort. Farklar	Std. Hata	p
Lineer-radyal tabanlı	- 0,057	0,018	0,004
Lineer-polinomiyal	- 0,025	0,018	0,176
Lineer-sigmoid	- 0,04	0,018	0,830
Radyal tabanlı-polinomial	0,031	0,018	0,095
Radyal tabanlı-sigmoid	0,053	0,018	0,007
Sigmoid-polinomiyal	- 0,021	0,018	0,251

Ek olarak, polinomiyal tabanlı çekirdek fonksiyonu ile radyal tabanlı çekirdek fonksiyonu arasında istatistiksel bir farklılık bulunmamıştır. Dolayısıyla her iki çekirdek fonksiyonunun sınıflandırma performanslarının birbirine benzer olduğu söylenebilir. Ayrıca, dört çekirdek fonksiyonu performanslarına göre homojen gruplara ayrılmıştır. Tablo 7'de çekirdek fonksiyonlarının yer aldığı alt homojen gruplar gösterilmektedir.

Tablo 7. Çekirdek Fonksiyonlarının Yer Aldığı Alt Gruplara İlişkin Sonuçlar

	Çekirdek fonksiyonları	N	Alt gruplar	
			1	2
Tukey HSD ^{a,b}	lineer	12	1,0483	
	sigmoid	12	1,0522	
	polinomiyal	12	1,0736	1,0736
	radyal	12		1,1050
	p		,518	,331

P* 0,05.

Çekirdek fonksiyonlarının yer aldığı alt grupların $p \geq 0,05$ olduğu için homojen gruplar olduğu görülmektedir. Bu sonuçlara göre lineer ve sigmoid çekirdek fonksiyonlarının sınıflandırma performansları bakımından benzerlik gösterdiği ve radyal tabanlı çekirdek fonksiyonun bu gruptan tamamen farklı sonuçlar verdiği yargısına varılmaktadır. Polinomiyal çekirdek fonksiyonu her iki grupta da yer almaktadır. Tüm analizler ve algoritmanın sınıflandırma performansları dikkate alındığında, polinomiyal çekirdek fonksiyonu, radyal tabanlı çekirdek fonksiyonuna göre daha kötü diğer iki çekirdek fonksiyonuna göre daha iyi sınıflandırma performansına sahiptir. Sonuç olarak, DVM'nin performansı için en iyi çekirdek fonksiyonun radyal tabanlı çekirdek fonksiyonu olduğu görülmektedir.

Bu çalışmada, grid arama yöntemi ile her bir veri setine ilişkin en iyi çekirdek fonksiyonu olarak belirlenen radyal tabanlı çekirdek fonksiyonu ve polinomiyal çekirdek fonksiyonu için elde edilen optimal hiper parametre değerleri ve sınıflandırma performansları Tablo 8'de özetlenmiştir.

Tablo 8'deki sonuçlar dikkate alındığında, radyal tabanlı çekirdek fonksiyonu için C ceza parametresinin değerinin veri setlerine göre 100 ile 0,1 arasında değiştiği gözlenmiştir. Polinomiyal tabanlı çekirdek fonksiyonu için ise C parametresi 0,06 ile 25,06 arasında değer almaktadır. Gamma parametresinin değer aralığı radyal tabanlı çekirdek fonksiyonu için 0,003 ile 15,02 iken polinomial fonksiyon için 0,003 ile 100 arasında değişmektedir.

Bu sonuçlardan yola çıkarak, polinomial çekirdek fonksiyonu için gamma parametresinin değerleri radyal tabanlı fonksiyona göre farklı veri setleri için daha fazla değişkenlik gösterdiği söylenebilir. Radyal tabanlı çekirdek fonksiyonu için de C ceza parametresinin değerleri daha fazla değişkenlik göstermektedir. Ayrıca polinomial çekirdek fonksiyonu için α katsayısı değerlerinin söz konusu veri setleri için 0 ile 12,91 arasında değiştiği gözlenmiştir.

Tablo 8. Radyal Tabanlı ve Polinomiyal Çekirdek Fonksiyonlarının Optimal Hiper Parametre Değerleri ve Genelme Performansları

Veri Setleri	Radyal Tabanlı			Polinomiyal			Test (%)
	C	γ (gamma)	Test (%)	C	γ (gamma)	α (katsayı)	
Australian credit	0,1	0,435603	86,47	0,25763	0,03684	0,215444	87,09
German	41,01832	0,003327	76,90	0,199474	0,003327	12,91549	76,70
Glass	12,56605	15,02665	73,37	1,34876	4,516	4,64159	72,00
Hearth disease	0,79370	0,03684	85,19	25,06595	0,01107	0,00464	84,07
Pima Indian Diabet	1,64183	0,12258	77,73	0,1	0,011072	10	76,95
Ionosphere	1	0,1	91,30	1	0,1	0,01	84,06
Iris	1,1071732	1,3572	97,30	3,684032	50	0	98,00
Spambase	10	0,01	94,00	10	0,01	0,01	92,30
Wine	0,22901	0,4	97,20	0,1	1,3572	0,0000	97,20
WBC	0,397897	0,4078	97,07	25,06596	0,12258	1	97,07
Bank	100	0,01	89,85	1,35721	0,01	0,001	85,50
Türkiye kredi	0,806234	0,15590	89,82	0,06787	100	0,0000	86,23

5. Tartışma ve Öneriler

Bu çalışmada literatürden farklı olarak, çekirdek fonksiyonu seçim süreci rassal blok deney tasarımı temeline oturtulmuştur. Rassal blok deney tasarımıyla veri setlerinin farklılıklarından doğacak DVM performansları üzerindeki etkilerinin yok edilmesi amaçlanmıştır. Böylece çekirdek fonksiyonlarının performansları arasında farklılık olup olmadığı veri setlerinin etkileri arındırılarak belirlenmiştir. Sonuç olarak en başarılı performansa sahip çekirdek fonksiyonunun radyal tabanlı fonksiyon olduğu ve polinomiyal çekirdek fonksiyonu ile performansları arasında bir farklılık olmadığı kanıtlanmıştır. Ancak polinomiyal çekirdek fonksiyonunun algoritmanın uygulanmasında çözüm zamanı açısından maliyetli olduğu gözlemlenmiştir. Örneğin, radyal tabanlı çekirdek fonksiyonu ile 690 örneklili bir veri seti için 1 dk gibi bir zamanda çözüme ulaşılabılırken, polinomiyal çekirdek fonksiyonu ile 5 saat 30 dk gibi bir zaman diliminde çözüme ulaşılmaktadır. Dolayısıyla, DVM’de radyal tabanlı çekirdek fonksiyonun kullanılması hem çözüm zamanı açısından hem de performans başarısı açısından daha avantajlı olduğu kanıtlanmıştır.

Sonuç olarak bu çalışma, araştırmacılara sınıflandırma süreçlerinde yararlanabilecekleri istatistiksel bakış açısıyla desteklenmiş önemli ipuçları sağlamaktadır.

Kaynaklar

- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition, data mining and knowledge discovery. Kluwer Academic Publishers, 2 (2), 121-167.
- Busuttil, S. (2003). Support vector machines. In Proceedings of the Computer Science Annual Research Workshop, Villa Bighi, Kalkara, University of Malta.
- Conagin, A., Barbin, D., Demétrio, C.G.B. (2008). Modifications for the Tukey test procedure and evaluation of the power and efficiency of multiple comparison procedures. *Scientia Agricola*, 65, 428-432.
- Cortes, C., Vapnik, V. (1995). Support vector networks, *Machine Learning*, 20,1-25.
- Çakar, Ö. (2007). Fonksiyonel analize giriş I. A.Ü. Fen Fakültesi Döner Sermaye İşletmesi Yayınları, no:13, (Erwin KREYSZİG'den Uyarlama).
- Demirci, D. A. (2007). Destek vektör makineleri ile karakter tanıma, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Fernandez, G.C.J. (1992). Residual analysis and data transformations: Important tools in statistical analysis. *HortScience*, 27, 297–300.
- Fletcher, T. (2009). Support vector machines explained. www.cs.ucl.ac.uk/sta/T.Fletcher/
- Gunn, S. R. (1998). Support vector machines for classification and regression. Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science. <http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf>
- Huang, C. L. and Wang, C. J. (2006). A GA-based feature selection and parameter optimization for support vector machines. *Expert Systems with Applications*, 31:231-240.
- Katagiri, S. and Abe, S. (2006). Incremental training of support vector machines using hyperspheres. *Pattern Recognition Letters*, 27 (13), 1495-1507
- Li, S., Li, H., Li, M., Shyr, Y., Xie, L. and Li, Y. (2009). Improved prediction of lysine acetylation by support vector machines. *Protein and peptide letters*, 16, 977-983.
- Nitze, I., Schulthess, U. And Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proceedings of the 4th GEOBIA, Janeiro - Brazil., 35-40.

Oowski, S., Siwekand, K., and Markiewicz, T. (2004). MLP and SVM Networks – a Comparative Study. Proceedings of the 6th Nordic Signal Processing Symposium – NORSIG.

Schölkopf, B., and Smola A.J. (2002). Learning with Kernels. MIT Press,626 s.

Soman, K.P., Loganathan, R. and Ajay, V. (2011). Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd., 486 s.

Suykens, J. A. K. (2002). Least squares support vector machines. River Edge, NJ : World Scientific, xiv, 294 s.

UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, CA.

ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ DERGİSİ

YAYIN ve YAZIM KURALLARI

1. Eskişehir Osmangazi Üniversitesi İİBF Dergisi, İktisadi ve İdari Bilimler alanında özgün makaleleri yayınlamayı amaçlayan hakemli bir dergidir. Yılda iki kez yayınlanan dergi, alanında kuramsal ve uygulamalı çalışmalara yer verir.
2. Dergiye gönderilecek makaleler Türkçe veya İngilizce olabilir.
3. Yayına gönderilecek makalelerin aynı anda başka bir derginin değerlendirme sürecinde bulunmaması, hiçbir yerde yayına kabul edilmemiş ve yayınlanmamış olması gerekmektedir.
4. Yayılanmak üzere dergiye gönderilen makaleler ile birlikte yazar/ların adı-soyadı, ünvanı, kurum, ve elektronik posta adresleri ile açık iletişim adreslerini içeren bilgiler ayrı bir sayfada gönderilmelidir.
5. Yazım kurallarına uygun olarak gönderilen makaleler dergi editörü tarafından incelenir. Hakeme gönderilmesi uygun görülmeyen makaleler yazar(lar)ına bildirilir.
6. Hakeme gönderilmesi uygun görülen makaleler, konusunda uzman iki hakeme gönderilir. Hakem raporları doğrultusunda editör gerekli gördüğü durumda üçüncü bir hakem belirleyebilir.
7. Makale metninde makalenin Türkçe ve İngilizce başlıkları, 120 kelimeyi aşmayacak şekilde Türkçe ve İngilizce özetler ile en fazla beşer adet Türkçe ve İngilizce anahtar kelimeler yer almalıdır. Makale metninde yazar/ların kimlik bilgileri yer almamalıdır.
8. Dergiye gönderilecek yazılar A4 ebadında kağıda, Times New Roman, 12 punto, 1,5 aralıkla, metin, tablo ve şekiller, kaynakça ve ekler dahil 25 sayfayı aşmayacak şekilde yazılmış olmalıdır. Sayfalar numaralandırılmalıdır.
9. Tüm metin iki yana yaslı, paragraflar arasında 12nk boşluk verilmiş, başlıklar ve metin dahil olmak üzere soldan girinti yapılmamış olmalıdır. Gönderilecek çalışmaların sayfa kenar boşlukları her taraftan 2,5 cm olacak şekilde ayarlanmalıdır.

10. Tüm başlıklar kalın (bold), sola yaslı (girintisiz) ve yalnızca kelimelerin ilk harfleri büyük olacak şekilde yazılmalıdır. Alt başlıklar 1., 1.1, 1.1.1. şeklinde numaralandırılmalıdır.

11. Metin içi atıflarda Harvard metodu olarak adlandırılan ve yazar soyadı, tarih ve sayfa numaralarının verildiği sistem tercih edilmelidir (Örn: Clegg, 1997: 53). İki-den fazla yazarı olan kaynaklara atıflarda ilk yazarın soyadı ve "vd." ibaresi kullanılmalıdır (Örn: Morgan vd., 1994). Aynı parantez içerisinde birden fazla kaynak noktalı virgül (;) işareti ile ayrılmalıdır (Örn: Hassard ve Parker, 1994; Boje, 1996).

12. Metin içinde yer alacak tablo, şekil, grafik, harita vb.'lerinin de bu ölçüleri aşmayacak şekilde metin içine ortalanarak yerleştirilmiş olması ya da gerekiyorsa ekler bölümünde -metin sonunda- kaynakçadan hemen önce yer alması gereklidir.

13. Metin içindeki tüm şekiller ve grafikler sıra numarası ile (Şekil 1) kendi içinde ve şekil ya da grafiğin altında; tablolar ise yine kendi içinde numaralanmak üzere (Tablo 1) tablonun üzerinde numaralandırılmış ve isimlendirilmiş olmalıdır. Tablo, grafik ve şekil başlıkları sayfaya ortalanmış, kalın (bold) ve yalnızca kelimelerin baş harfleri büyük olacak şekilde yazılmalıdır.

14. Tablo, şekil ve grafiklerin varsa kaynakları; tablo, şekil ve grafiklerin hemen altında metin içi atıf kurallarına uygun olarak verilmelidir. Matematiksel ve istatistiksel simgeler Microsoft Office denklem düzenleyicisi ile hazırlanmalıdır.

15. Makalenin sonunda yazar soyadlarına göre alfabetik olarak düzenlenecek kaynakça kısmı bulunmalıdır. Kaynakçada sadece makalede kullanılan eserler yer almalıdır ve kaynakça aşağıda belirtilen örneklere uygun olarak hazırlanmalıdır.

KİTAPLAR

Kazgan, G. (1989), İktisadi Düşünce veya Politik İktisadın Evrimi, İstanbul: Remzi Kitabevi.

Wood, R. ve T. Payne (1998), Competency Based Recruitment and Selection, London: Wiley.

Mondy, R. W., R. M. Noe, ve S. R. Premeaux (2002), Human Resource Management, NJ: Prentice Hall.

DERLEME KİTAPTAN BÖLÜM

Toynbee, A. (2000), "Osmanlı İmparatorluğu'nun Dünya Tarihindeki Yeri", Ed. Kemal Karpaz, Osmanlı ve Dünya, İstanbul: Ufuk Kitapları, 49-67.

MAKALELER

Paskaleva, V. (1967), "Osmanlı Balkan Eyaletlerinin Avrupalı Devletlerle Ticaretleri Tarihine Katkı 1700-1850", İÜ. İktisat Fakültesi Dergisi, 27(1-2), 48-59.

Li, T. ve R. J. Calantone (1998), "The Impact of Market Knowledge Competence on New Product Advantage: Conceptualization and Empirical Examination", Journal of Marketing, 61(2), 13-29.

İNTERNET KAYNAKLARI

Yazarı Belli Olan İnternet Kaynakları:

Salmon, P. (2003), "Decentralization and Supranationality: The Case of the European Union", <http://www.imf.org/external/pubs/fiscal/salmon.pdf>, (Erişim: 02.10.2003).

Yazarı Belli Olmayan İnternet Kaynakları:

"Special Topic: Corporate Income Taxation and FDI in the EU-8", <http://siteresources.worldbank.org/INTLATVIA/Resources/QER3spec.doc>, (Erişim: 28.10.2004).

<http://www.tcmb.gov.tr>, (Erişim: 28.10.2004).

Belirtilen formatta hazırlanan çalışmalar elektronik posta aracılığıyla iibfdergi@ogu.edu.tr adresine ekli Microsoft Word belgesi olarak gönderilmelidir. Yazarlara, yazının ulaştığına dair bilgi ve değerlendirme sürecini dergi internet sitesinden izlemede kullanabilecekleri makale takip numarası yollanacaktır. Yazarlar gerekirse editöre, derginin diğer iletişim kanalları yanında aşağıdaki adresten doğrudan posta yoluyla ulaşabilirler:

Prof. Dr. Sami Taban

ESOGÜ İİBF Dergi Editörü

Eskişehir Osmangazi Üniversitesi

İktisadi ve İdari Bilimler Fakültesi

Meşelik Kampusu 26480

ESKİŞEHİR