Anadolu Üniversitesi Bilim ve Teknoloji Dergisi B-Teorik Bilimler
Anadolu University Journal of Science and Technology B- Theoretical Sciences
Cilt: 3 Sayı: 2 - 2015
Sayfa: 53 - 65
DOI: 10.20290/btdb.67263

ARAŞTIRMA MAKALESİ / **RESEARCH ARTICLE**

Furkan BAŞER [1], Ayşen APAYDIN [2]

SINIFLANDIRMA AMAÇLI DESTEK VEKTÖR MAKİNELERİNİN LOJİSTİK REGRESYON İLE KARŞILAŞTIRILMASI

ÖZ

Gözlemlerin sınıflandırılması, hem veri analizinde hem de karmaşık bir sistemin alt amacı olarak istatistiğin ve makine öğrenmesinin önemli bir bileşenidir. Yeni bir makine öğrenmesi tekniği olan destek vektör makineleri (DVM) de fonksiyon tahmini ve sınıflandırma problemlerinin çözümü için önerilmiş olan bir istatistiksel öğrenme algoritmasıdır. Bu çalışmada, bazı ikili sınıflandırma problemlerinde standart lojistik regresyon modelleri kullanılacak ve bu modeller, doğrusal ve doğrusal olmayan kernel fonksiyonları için DVM modelleri ile karşılaştırılacaktır. Düşük doğum ağırlıklı bebek doğumu ve prostat kanser riskli hastalara ilişkin gerçek veriler ile gerçekleştirilen analizlerden elde edilen bulgular değerlendirildiğinde, destek vektör sınıflandırmasının dikkate değer sonuçlar verdiği belirlenmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, Sınıflandırma, Lojistik regresyon, Destek vektör makineleri.

COMPARISON BETWEEN LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINES FOR CLASSIFICATION PURPOSES

ABSTRACT

The classification of observations is an important constituent of statistics and machine learning, either for analysis of data sets, or as a subgoal of a more complex problem. A novel machine learning technique, Support Vector Machines (SVM), has recently been receiving considerable attention in pattern recognition and regression function estimation problems. This paper uses standard logistic regression models for binary classification problems and compares them with SVM models with linear and non-linear kernel functions. An application with real data associated with giving birth to a low birth weight baby and patients with cancer of prostate are presented as an illustration. Based on the results of the numerical examples, it is determined that Support Vector Classification method produces remarkable results.

Keywords: Machine learning, Classification, Logistic regression, Support vector machines.

[1,] Gazi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Uluslararası Ticaret Bölümü, 06500-Beşevler, Ankara
E-mail: furkan.baser@gmail.com, Tel: 0 312 216 10 00
[2,] Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100-Tandoğan, Ankara.
E-mail: apaydin@science.ankara.edu.tr,Tel: 0 312 212 67 20
Geliş: 14 Nisan 2014 **Düzeltme:** 22 Temmuz 2014 **Kabul:** 21 Aralık 2014

1. GİRİŞ

Destek Vektör Makineleri (DVM), teorik çözümlemeler ile sayısal algoritmaları yapısında birleştiren bir sınıflandırma ve regresyon yöntemidir. İstatistiksel öğrenme teorisinde bu teknik, deneysel risk minimizasyonundan (Empirical Risk Minimization – DRM) ziyade yapısal risk minimizasyonuna (Structural Risk Minimization – YRM) dayandırılan bir öğrenme algoritması olarak geliştirilmiştir. YRM tümevarım prensibi, sonlu örneklemler için Vapnik–Chervonenkis (VC) boyutuna bağlı olarak optimum model karmaşıklığını belirlemek üzere biçimsel bir mekanizma sağlar (Vapnik, 1995). Klasik sinir ağları ile karşılaştırıldığında DVM, bir tek global optimum çözüm elde edebilir ve boyut sorunu ile karşılaşmaz. Bu ilgi çekici özellikleri DVM'yi sıklıkla tercih edilir bir teknik haline getirmektedir.

Son yıllarda yapılan çalışmalarda, DVM'nin farklı yönlerinde önemli gelişmeler kaydedildiği görülmektedir. Bu gelişmeler, DVM'nin kuramsal altyapısını, kullanılan algoritmik stratejileri ve gerçek dünya uygulamalarını da içermektedir. DVM, metin sınıflandırma (Joachims, 1998; Yang et al., 2006), görüntü algılama (Osuna et al., 1997; Min and Cheng, 2009), biyoinformatik (Jaakkola and Haussler, 1998; Brown et al., 2000; Zien et al., 2000) gibi çok çeşitli problemlerde üstün genelleme performansı sağlamaktadır. Bu uygulama alanı için iyi performans, girdi uzayının yüksek boyutlu olduğu problemlerde DVM'nin öğrenme kabiliyetinin, nitelik uzayının boyutundan bağımsız olduğu gerçeği ile ilişkilendirilir. DVM'nin kapsamlı incelemesi, Schölkopf and Smola (2002) tarafından yapılan çalışmada bulunabilir.

Sınıflandırma analizinde, girdilerden sınıf etiketine eşlenen fonksiyonun ve parametrelerinin tahmin edilmesi amaçlanır. Eğitilmiş bir sınıflandırıcı kullanılarak hesaplanan sınıflandırma hatasının minimum yapılması da analizin esasını oluşturmaktadır. Biyolojik verilerin sınıflandırma analizinde uygulanan birçok istatistiksel yöntem mevcuttur. Karar ağaçları, lojistik regresyon (LR), diskriminant analizi, sinir ağları ve destek vektör makineleri bu yöntemler arasındadır (Yang, 2004). Lojistik model, katsayı tahmin yöntemlerinin geliştirilmesi ile birlikte biyoinformatik alanında kullanılan standart bir yöntem haline gelmiştir. Düşük doğum ağırlığının kestirimi (Buescher et al., 1993; Kloiber et al., 1996), kardiyovasküler hastalık riskinin kestirimi (Gardside and Glueck, 1995; Cox et al., 1998) gibi problemlerde LR analizi yoğun bir biçimde kullanılmaktadır.

Sınıflandırma problemlerinde daha doğru, kullanışlı ve genelleme kapasitesi yüksek bir kestirim yaklaşımı geliştirmek üzere klasik yöntemler ile birlikte makine öğrenmesi yöntemlerinin de göz önünde bulundurulması önem arz etmektedir. Bu çalışmada, düşük doğum ağırlıklı bebeklerin kestirimi ve prostat kanser riskli bireylerin kestirimi biçiminde belirlenen iki problem için DVM ve LR çözümlemesi uygulanacak; elde edilen bulgular hatalı sınıflandırma sayılarına göre karşılaştırılacaktır.

Çalışmanın ikinci bölümünde, sınıflandırma amaçlı destek vektör makineleri incelenecek; çekirdek (kernel) gösteriminden yararlanarak doğrusal olmayan destek vektör sınıflandırmasına ilişkin algoritma sunulacaktır. Üçüncü bölümde lojistik regresyon ve ilgili parametre tahmin yöntemi özet bir biçimde ele alınacaktır. Son olarak, çalışmanın dördüncü bölümünde ele alınan uygulamalardan elde edilen bulgular da değerlendirilerek ulaşılan önemli sonuçlar özetlenecektir.

2. DESTEK VEKTÖR MAKİNELERİ İLE SINIFLANDIRMA

Destek vektör makineleri (DVM), sınırlı sayıda öğrenme örüntüsü üzerinden iyi bir genelleme düzeyi elde etmek amacıyla yapısal risk minimizasyonu (YRM) tümevarım prensibini uygulayan bir öğrenme makinesidir. YRM, deneysel riski ve VC (Vapnik–Chervonenkis) boyutunu minimum yapmak üzere eşanlı girişimlerden oluşmaktadır. Teori esasında ayrılabilir ikili sınıflandırma problemi temelinde, Vapnik ve çalışma arkadaşları tarafından AT & T Bell Laboratuarlarında geliştirilmiştir. DVM, karmaşık veri setlerinde, çözümlemesi zor örüntülerin tanımlanmasında kullanışlı bir öğrenme algoritmasını uygulamaktadır. Algoritma, önceden gözlenmemiş verilerin sınıflandırma kestirimi için örneklerden ayırt edebilen bir sınıflandırma öğrenmesini gerçekleştirmektedir.

DVM'nin, destek vektör sınıflandırması (Support Vector Classification – DVS) ve destek vektör regresyonu (Support Vector Regression – DVR) olmak üzere iki temel türü vardır. Yüksek boyutlu bir nitelik uzayını kullanan DVM, destek vektörlerin altkümesi üzerine kurulu fonksiyon kestirimleri verir.

2.1 Doğrusal Sınıflandırıcılar

$\mathcal{X}$ girdi örüntülerinin uzayını (örneğin $\Re^d$) ve $\langle .,. \rangle$, $\mathcal{X}$ uzayında iç çarpımı göstermek üzere; n gözlem içeren eğitim verisi, $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, $\mathbf{x} \in \mathcal{X}$, $y \in \{+1, -1\}$ biçiminde ele alınsın. Sınıflandırma problemleri için hiperdüzlem karar fonksiyonu,

$$D(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b\,, \quad \mathbf{w} \in \mathcal{X}\,, \quad b \in \Re \tag{1}$$

biçiminde tanımlanır (Cristianini and Shawe-Taylor, 2000). Başarılı bir eğitim süreci sonunda, $\mathbf{w}$ ve b katsayı tahminleri kullanılarak, yeni gözlenen $\mathbf{x}$ örüntüleri için sign $D(\mathbf{x})$ işaret fonksiyonuna göre çıktı üretilir.

Destek vektör sınıflandırmasında amaç, deneysel risk (pay sınırının "yanlış" tarafında bulunan veriler için sınırdan sapmalar toplamı) minimum yapılırken; hatasız sınıflandırılmış veriler (pay sınırının "doğru" tarafında bulunan) için ise payın maksimum yapılmasıdır. Teknik olarak bu yorum, girdi uzayını iki bölgeye parçalayan uyarlamalı bir kayıp fonksiyonunun (Δ pay büyüklüğüne bağlı) kullanımını gerektirir. Girdi uzayını, eğitim verisinin bir kısmının model tarafından açıklanabildiği (sıfır kayıp) ve geri kalanının bir miktar belirsizlik ile açıklanabildiği iki bölgeye ayıran kayıp fonksiyonu,

$$L_\Delta\big(y, D(\mathbf{x}, \mathbf{w})\big) = \text{maks}(\Delta - y\, D(\mathbf{x}, \mathbf{w}), 0) \tag{2}$$

biçiminde verilir (Cherkassky and Mulier, 2007). Sınıflandırma problemleri için (2) eşitliğinde verilen DVM kayıp fonksiyonuyla deneysel risk,

$$R_{emp}(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} L_\Delta\big(y_i, D(\mathbf{x}_i, \mathbf{w})\big) \tag{3}$$

ile belirtilir. Pay tabanlı kayıp;
$\xi_i = \text{maks}\ (1 - y_i D(\mathbf{x}_i, \mathbf{w}), 0)\,, i = 1, \dots, n$ gevşek değişkenleri ile ifade edilen pay sınırlarından sapmaların bir göstergesidir.

Optimum ayırma hiperdüzleminin bulunması, doğrusal kısıtlar ile tanımlı bir karesel optimizasyon problemidir. Buna göre, $(\mathbf{x}_i, y_i)$, $i = 1, \dots, n$; $\mathbf{x} \in \Re^d$ eğitim verisi mevcut olmak üzere programlama problemi,

Amaç fonksiyonu:

$$\min_{\mathbf{w} \in \mathcal{X}, b \in \Re} \left[\frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^{n} \xi_i \right]$$

Kısıtlar: (4)

$$y_i[\langle \mathbf{w}, \mathbf{x}_i \rangle + b] \geq 1 - \xi_i, i = 1, \dots, n$$

biçiminde oluşturulur (Cherkassky and Mulier, 2007). Bu modelde kullanıcı tarafından belirlenmesi gereken C katsayısı, hiperdüzlem karar fonksiyonunun karmaşıklığı ile deneysel risk arasındaki arasındaki değişimi kontrol eder.

Bu optimizasyon problemi, yüksek boyutlu girdi uzayları için çözümlenecekse dual formuna dönüştürülmesine ihtiyaç duyulur (Vapnik, 1995). Optimizasyon teorisine göre, eğer amaç fonksiyonu ve kısıtlar kesin konveks ise optimizasyon probleminin dual formunun mevcut olduğu söylenir. Buna göre primal problemin çözümü, dual problemin çözümüne denktir. (4) ile verilen optimizasyon problemi, bu kriterleri sağlar ve bir dual forma sahiptir. Bu durumda, problemi dualine dönüştürmek üzere Kuhn-Tucker teoremi kullanılır (Strang, 1986).

$(\mathbf{x}_i, y_i), i = 1, \dots, n$ eğitim verisi ve C de bir düzenleme sabiti ve α_i, $i = 1, \dots, n$ Lagrange katsayıları için dual problem,

Amaç fonksiyonu:

$$\underset{\boldsymbol{\alpha} \in \Re^n}{\text{maks}} \left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right]$$

Kısıtlar: (5)

$\sum_{i=1}^{n} y_i \alpha_i = 0$

$0 \le \alpha_i \le C/n, \; i = 1, \dots, n$

biçiminde elde edilir (Cristianini and Shawe-Taylor, 2000). $\alpha_i, i = 1, \dots, n$ dual problemin çözümü olmak üzere ve b ise

$$b = y_s - \sum_{i=1}^{n} \alpha_i \, y_i \langle \mathbf{x}_i, \mathbf{x}_s \rangle \tag{6}$$

eşitliği ile belirlendiğinde, hiperdüzlem karar fonksiyonu,

$$D(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b \tag{7}$$

dir. Burada, α_i parametrelerinin sıfırdan farklı olduğu veri örnekleri, destek vektör olarak adlandırılır (Kecman, 2001).

2.2 Doğrusal Olmayan Sınıflandırıcılar

DVM, girdi vektörlerinin, yüksek boyutlu bir nitelik uzayına bazı doğrusal olmayan dönüşümleri aracılığıyla; girdi uzayında doğrusal olmayan sınıf sınırlarını uygulamak üzere nitelik uzayında doğrusal bir model kullanır. Yani nitelik uzayında kurulan doğrusal model, girdi uzayında doğrusal olmayan karar sınırını temsil eder. Böylece, DVM algoritması yardımıyla nitelik uzayında optimum ayırma hiperdüzlemi, karar sınırları arasında maksimum pay oluşacak biçimde seçilir. Karar sınırlarına en yakın eğitim örnekleri destek vektör olarak belirlenir. Diğer eğitim örnekleri ise ikili sınıflandırma karar sınırlarının tanımlanmasında etkisizdir (Gunn, 1998; Vapnik, 1998; Cristianini and Shawe-Taylor, 2000).

Eğitim verisinin başka bir uzayın altkümesi olacak biçimde dönüştürülmesi makine öğrenmesi alanında uzun zamandır bilinen ve uygulanan bir yaklaşımdır. Burada $\mathcal{X}$ girdi uzayı, $\mathcal{F} = \{\Phi(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ ise nitelik uzayı olarak adlandırılır. $\Phi : \mathcal{X} \longrightarrow \mathcal{F}$ doğrusal olmayan bir dönüşümü göstermek üzere, doğrusal olmayan sınıflandırma problemlerinde karar fonksiyonu

$$D(\mathbf{x}) = \sum_{i=1}^{n} w_i \phi_{\mathrm{i}}(\mathbf{x}) + b \tag{8}$$

biçiminde tanımlanır (Cristianini and Shawe-Taylor, 2000). Burada, m nitelik uzayının boyutunu göstermek üzere $\Phi(\mathbf{x}) = [\phi_1(\mathbf{x}) \, \phi_2(\mathbf{x}) \dots \phi_m(\mathbf{x})]$'dir. Destek vektör algoritması sadece, $\mathbf{x}_i$ örneklerinin iç çarpımlarına bağlı olarak tanımlıdır. Buna göre, destek vektör optimizasyon problemini yeniden ifade edilmesine olanak veren, Φ fonksiyonundan ziyade, $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ iç çarpım çekirdeğinin bilinmesi yeterlidir.

Doğrusal olmayan bir DVM (sınıflandırıcısı) için öğrenme algoritması, nitelik uzayında optimum ayırma hiperdüzleminin tasarlanması ile ortaya çıkar. Bu işlem, girdi uzayında, pay hiperdüzleminin oluşturulmasına benzerdir. $(\mathbf{x}_i, y_i),\ i = 1, \ldots, n$ eğitim verisi, k iç çarpım çekirdeği ve C de bir düzenleme sabiti olmak üzere doğrusal olmayan destek vektör sınıflandırması için dual optimizasyon problemi,

Amaç fonksiyonu:

$$\underset{\boldsymbol{\alpha}\in\Re^n}{\text{maks}} \left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i\alpha_j y_i y_j\, k(\mathbf{x}_i, \mathbf{x}_j)\right]$$

Kısıtlar: (9)

$$\sum_{i=1}^{n} y_i\alpha_i = 0$$

$$0 \leq \alpha_i \leq C/n\,,\ i = 1, \ldots, n$$

biçimindedir (Kecman, 2001). Burada $\alpha_i,\ i = 1, 2, \ldots, n$ Lagrange katsayılarını göstermektedir. Bu yaklaşımda esas, etkili bir biçimde kullanılabilecek bir çekirdek fonksiyonunun bulunmasıdır. Böylece, $D(\mathbf{x})$ karar hiperdüzlemi,

$$D(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \tag{10}$$

ve $d > 3$ için aynı zamanda bir hiperdüzlem olan destek vektör sınıflandırıcısı,

$$\operatorname{sign} D(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{11}$$

dir. Burada, $(\mathbf{x}_s, y_s)$ destek vektörlerden biri olmak üzere b parametresi,

$$b = y_s - \sum_{i=1}^{n} \alpha_i\, y_i\, k(\mathbf{x}_i, \mathbf{x}_s) \tag{12}$$

ile elde edilir.

Öğrenme makinesi için kullanılan dönüşüm fonksiyonunun türü, iç çarpımın hesaplanması amacıyla çekirdek fonksiyonlarının farklı seçimlerine göre değişiklik gösterir. Tablo 1 ile makine öğrenmesi ve sinir ağları alanlarında sıklıkla uygulanan iç çarpım çekirdekleri verilmiştir (Kecman, 2001).

Tablo 1. Klasik çekirdek fonksiyonları

Çekirdek Fonksiyonları	**Sınıflandırma Türü**
$k(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x}^T\mathbf{x}_i) + 1]^q$	q derecesinden polinomiyal
$k(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{1}{2}[(\mathbf{x} - \mathbf{x}_i)^T\Sigma^{-1}(\mathbf{x} - \mathbf{x}_i)]\right\}$	Gauss radyal tabanlı fonksiyon
$k(\mathbf{x}, \mathbf{x}_i) = \tanh[(\mathbf{x}^T\mathbf{x}_i) + b]$*	Çok tabakalı sinir ağı

* Belirli bir b değeri için

3. LOJİSTİK REGRESYON

n gözlem içeren $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, $\mathbf{x} \in \Re^d$, $y_i \in \{+1, -1\}$ eğitim verisi için $\mathbf{x}_i$ gözlemini belirli iki sınıftan birinde sınıflandırılan bir regresyon modeli göz önüne alınsın. Burada, $\mathbf{x}_i$ deneyi, $\mu(\mathbf{x}_i)$ ortalama ile bir Bernoulli denemesi olarak düşünülebilir. Böylece y_i, ortalaması $\mu(\mathbf{x}_i)$ ve varyansı $\mu(\mathbf{x}_i)\big(1 - \mu(\mathbf{x}_i)\big)$ olan bir Bernoulli rasgele değişkenidir. Her bir $\mathbf{x}_i$ deneyi ile karşılık gelen çıktının beklenen değeri arasındaki ilişkiyi modellemek üzere Lojistik Regresyon (LR) fonksiyonu kullanılır. $\boldsymbol{\beta}$ parametre vektörünü göstermek üzere bu fonksiyon,

$$\mu(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})} \tag{13}$$

biçiminde tanımlanır (Menard, 1995; Komarek, 2004). Böylece, ε hata terimi ile regresyon modeli,

$$y = \mu(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon \tag{14}$$

olarak yazılır. Burada hata teriminin, $y = 1$ için $\varepsilon = 1 - \mu(\mathbf{x})$ veya $y = 0$ için $\varepsilon = \mu(\mathbf{x})$ biçiminde sadece iki değer alabileceği görülür. y rasgele değişkeninin, $\mu(\mathbf{x})$ ortalama ve $\mu(\mathbf{x})\big(1 - \mu(\mathbf{x})\big)$ varyans ile Bernoulli dağılımlı olmasından dolayı; ε hatası, 0 ortalama ve $\mu(\mathbf{x})\big(1 - \mu(\mathbf{x})\big)$ varyansına sahiptir. Bu durum, hata ve çıktı değişkeninin, ortalamadan bağımsız σ^2 sabit varyansına sahip olduğu doğrusal regresyondan farklıdır (Komarek, 2004).

LR çözümlemesinde, $\boldsymbol{\beta}$ parametresine göre doğrusal olmayan LR modelinin tahminine, doğrusal regresyonda olduğu gibi hata kareler toplamının (HKT) minimumu ile ulaşılması uygun değildir. Bu durum, LR yönteminde HKT'nin minimumunun, hesaplanmasının zor olmasının yanında; olabilirlik fonksiyonunun maksimumuna eşdeğer olmamasından kaynaklanmaktadır. LR modelini, parametrelerine göre doğrusal bir fonksiyon haline dönüştürmek,

$$\mathrm{g}(\mu) = \ln\left[\frac{\mu}{1 - \mu}\right] \tag{15}$$

lojit fonksiyonunun kullanımıyla mümkündür. (13) eşitliğiyle verilen koşullu ortalama ve çıktı değişkeni için lojit dönüşümünün uygulanması ile

$$\mathrm{g}(y) = \boldsymbol{\beta}^T \mathbf{x} + \tilde{\varepsilon}(\mathbf{x}) \tag{16}$$

biçiminde yeni bir model elde edilir. Parametrelerine göre doğrusal olan bu model bağımsız değişkenin aldığı değerlere bağlı olarak $-\infty$ ile $+\infty$ arasında değişen sürekli bir fonksiyondur. Ancak hata teriminin, normal dağılıma uygunluk göstermemesi ve varyansının da ortalamadan bağımsız olmaması dolayısıyla; bu yeni model için parametre tahmininde klasik en küçük kareler yaklaşımının kullanımı uygun değildir. Bundan dolayı, parametre tahmin yöntemi olarak, en çok olabilirlik veya iteratif yeniden ağırlıklandırılmış en küçük kareler yaklaşımının kullanımı önerilmiştir (Komarek, 2004).

LR, veri bileşenleri ve parametrelerin doğrusal ilişkisinin sonucu olarak bir doğrusal sınıflandırma yöntemidir. Bu da LR yönteminin amacının, pozitif ve negatif veri noktalarını ayırmak üzere bir ayırıcı hiperdüzlemin bulunması biçiminde düşünülebileceğini gösterir.

4. UYGULAMA

4.1 Düşük Doğum Ağırlıklı Bebeklerin Kestirim Problemi

Düşük doğum ağırlığı ve prematürite, yenidoğanın en önemli problemlerinden olup mortalitenin de önde gelen nedenlerindendir (Rosenberg and Battaglia, 1991). Konunun önemi; bebek ölüm oranı ve doğuştan özürlü bebek oranının, düşük doğum ağırlıklı bebekler için yüksek olmasından kaynaklanmaktadır. Hamilelik esnasında kadının durum ve davranışları (diyet, sigara alışkanlığı, doğum öncesi bakım), bebeği doğuma taşıma ve sonucunda normal doğum ağırlıklı bebek olma olasılığını büyük ölçüde etkilemektedir.

Bu uygulamanın amacı, düşük doğum ağırlıklı bebek sorununa neden olan risk faktörlerini ortaya koymak ve model geliştirmektir. Bu amaçla; Hosmer and Lemeshow (2000) çalışmasında da kullanılmış olan, ABD'de 189 kadın üzerinde yapılan bir araştırma neticesinde derlenen veriler kullanılmıştır.

DVM algoritmaları, model karmaşıklığını kontrol eden C parametresi ve çekirdek fonksiyonunun seçimine bağlı bazı parametrelere bağlıdır. Genelleme hatasını minimum tutmak için bu parametrelerin optimumunun belirlenmesi önem arz etmektedir. Bu amaçla, veri seti, yapı tanımlama ve çıkarım aşamasında kullanılmak üzere rasgele iki parçaya ayrılmıştır. Böylece verilerin 126 (2/3)'sı eğitim ve 63 (1/3)'ü ise test verisi olarak değerlendirilmiştir.

Girdi ve çıktı değişkenleri aşağıdaki gibidir:

y : Düşük doğum ağırlığı ($0 =$ Doğum ağırlığı≥ 2500 gr.; $1 =$ Doğum ağırlığı< 2500 gr.)
x_1: Annenin yaşı
x_2: Annenin son menstrüal periyoddaki ağırlığı
x_3: Irk ($1 =$ Beyaz; $2 =$ Siyahi; $3 =$ Diğer)
x_4: Hamilelik esnasında sigara kullanma durumu ($0 =$ Hayır; $1 =$ Evet)
x_5: Prematüre doğum sayısı
x_6: Hipertansiyon durumu ($0 =$ Hayır; $1 =$ Evet)
x_7: Rahim hassaslık durumu ($0 =$ Hayır; $1 =$ Evet)
x_8: Hamileliğin ilk üç aylık döneminde doktor muayene sayısı

Birçok uygulama alanında, sadece veri setini sınıflandırabilmek değil, hangi değişkenlerin bu sınıflandırmayı gerçekleştirmek üzere en uygun olduğunun belirleyebilmek de önemlidir. LR modellerinde, katsayıların istatistiksel olarak anlamlı olup olmadığını test etmek mümkündür ve bu testler modellerin aşamalı olarak geliştirilmesinde kullanılabilir (Dreiseitl and Ohno-Machado, 2002).

Ele alınan verilere ilişkin LR çözümlemesinin ilk aşamasında, değişkenlerden çoklu lojistik modele alınacak olanlar belirlenir. Buna göre; her bir aday değişkene ilişkin oluşturulan basit LR modeli için katsayı anlamlılık testinin gerçekleştirilmesi gerekir. Basit LR çözümlemesinde istatistiksel olarak anlamlı değişkenlerin belirlenmesinde kullanılan Wald test istatistiğine ilişkin önem düzeyi (p değeri), 0.25'ten küçük olan değişkenler çoklu LR modeline alınır (Hosmer and Lemeshow, 2000).

Düşük doğum ağırlığı bağımlı değişkeni ile ilişkili olabileceği düşünülen olası değişkenlerin basit LR çözümlemesi sonuçları Tablo 2'de verilmiştir. Buna göre; annenin yaşı (x_1), annenin son menstrüal periyottaki ağırlığı (x_2), prematüre doğum sayısı (x_5), hipertansiyon durumu (x_6), hamileliğin ilk üç aylık döneminde doktor muayene sayısı (x_8) değişkenleri bağımlı değişken ile istatistiksel olarak anlamlı bir ilişki içinde olduklarından dolayı çoklu LR modeline aday değişkenler olarak belirlenmişlerdir.

Tablo 2'de sunulan $\exp(\hat{\beta})$ değerleri, ODDS oranlarını göstermektedir. Bu oran, ilgili değişkenin değerindeki bir birimlik artışın (azalışın), düşük doğum ağırlıklı bebek gözlenme olasılığında yaratacağı değişim miktarını belirtmektedir. Her bir girdi değişkeni ile basit LR çözümlemesi sonucunda oluşturulan modeller için olabilirlik değerleri de tabloda verilmiştir. Uyum iyiliğini belirleyen bu istatistikler, bağımlı değişken için açıklanamayan varyansın önemliliğini yansıtmaktadır. Model uyumu iyileştikçe; "-2(log-olabilirlik)" değeri küçülür. Ayrıca tablodan; x_6 değişkeni ile kurulan basit LR modelinden elde edilen doğru sınıflandırma oranının daha yüksek olduğu anlaşılmaktadır.

Tablo 2. Düşük doğum ağırlıklı bebeklerin kestirim probleminde tek değişkenli lojistik regresyon çözümlemesine ilişkin sonuçlar

Değişken	$\hat{\beta}$	$SE(\hat{\beta})$	**Wald**	sd	p	**exp**$(\hat{\beta})$	**exp**$(\hat{\beta})$ **için %95 Güven Aralığı**	**-2(log-olabilirlik)**	**Doğru Sınıflandırma Oranı**
x_1	-0.058	0.040	2.043	1	0.153	0.944	[0.872, 1.022]	150.377	0.706
x_2	-0.017	0.008	4.294	1	0.038	0.984	[0.968, 0.999]	147.319	0.706
x_3	-	-	5.546	2	0.062	-	-	146.848	0.706
x_3 (1)	0.578	0.622	0.864	1	0.353*	1.783	[0.526, 6.040]		
x_3 (2)	1.015	0.432	5.538	1	0.019	2.761	[1.185, 6.432]		
x_4	0.257	0.398	0.417	1	0.518*	1.293	[0.593, 2.820]	152.143	0.706
x_5	0.866	0.384	5.093	1	0.024	2.378	[1.121, 5.046]	147.053	0.690
x_6	1.499	0.759	3.902	1	0.048	4.479	[1.012, 19.830]	148.512	0.722
x_7	0.425	0.558	0.579	1	0.447*	1.529	[0.512, 4.566]	151.996	0.706
x_8	-0.298	0.212	1.977	1	0.160	0.742	[0.490, 1.125]	150.319	0.706

Çoklu LR modelinin kurulması aşamasında ikinci adım olarak değişkenlerin seçimi; adımsal seçim yöntemlerinden olabilirlik oran test istatistiğine dayalı geriye doğru eleme yöntemi kullanılarak yapılmıştır. Buna göre; annenin son menstrüal periyottaki ağırlığı (x_2), prematüre doğum sayısı (x_5) ve hipertansiyon durumu (x_6) değişkenleri ile oluşturulan lojistik modele ilişkin çözümleme sonuçları Tablo 3'te verilmiştir. Ayrıca, bireyde hipertansiyon olmasının, düşük doğum ağırlıklı bebek gözlenme olasılığını 11.422 kat arttırdığı belirlenmiştir

Tablo 3. Düşük doğum ağırlıklı bebeklerin kestirim probleminde değişken seçiminin ardından gerçekleştirilen çoklu LR çözümlemesine ilişkin sonuçlar

Değişken	$\hat{\beta}$	$SE(\hat{\beta})$	**Wald**	sd	p	**exp**$(\hat{\beta})$	**exp**$(\hat{\beta})$ **için %95 Güven Aralığı**
x_2	-0.022	0.009	6.302	1	0.012	0.978	[0.961, 0.995]
x_5	0.798	0.409	3.816	1	0.051	2.221	[0.997, 4.948]
x_6	2.435	0.966	6.356	1	0.012	11.422	[1.720, 75.856]
Sabit	1.607	1.103	2.120	1	0.145	4.985	-

Modelin eğitim veri setine uyumunu test etmek üzere Hosmer-Lemeshow testi kullanılmıştır. Bu test, her bir gruba ilişkin gözlenen ve beklenen frekansların karşılaştırılarak uyum iyiliği istatistiğinin hesaplanmasına dayanır. Düşük doğum ağırlıklı bebeklerin kestirimi için geliştirilen modele ilişkin gerçekleştirilen Hosmer-Lemeshow testine göre, gözlenen ve beklenen değerler arasında fark olmadığı, modelin veriye uyumunun oldukça iyi olduğu sonucuna ulaşılmıştır ($\mathcal{X}^2 = 8.412, sd = 7, p = 0.298$).

LR çözümlemesine göre; annenin son menstrüal periyottaki ağırlığı (x_2), prematüre doğum sayısı (x_5) ve hipertansiyon durumu (x_6) değişkenleri ile doğum ağırlığı (y) arasındaki bağıntı,

$$\mu(\mathbf{x}) = 1/[1 + \exp(1.607 - 0.022x_2 + 0.798x_5 + 2.435x_6)]$$

biçimindedir. Model yardımıyla elde edilen kestirim değerleri ile denekler, $\mu(\mathbf{x}) < 0.50$ ise düşük doğum ağırlıklı bebek ($y = 0$); $\mu(\mathbf{x}) \geq 0.50$ ise normal+ doğum ağırlıklı bebek ($y = 1$) biçiminde sınıflandırılmıştır.

DVM yöntemine göre; x_2, x_5, x_6 girdileri ve doğum ağırlığı çıktısı için doğrusal model tahmini,

$$D(\mathbf{x}) = 0.706 - 0.014x_2 + 1.587x_5 + 1.786x_6$$

biçiminde elde edilmiştir. Bu uygulamada iteratif arama sonucunda $C = 100$ olarak seçilmiştir. DVM algoritmasının avantajı, çekirdek fonksiyonunun seçimine bağlı olarak doğrusal ve doğrusal olmayan fonksiyonlar için farklı öğrenme makinelerinin oluşturulabilmesidir.

Tablo 4 ile farklı çekirdek fonksiyonlarının kullanımına bağlı olarak doğrusal ve doğrusal olmayan DVM çözümlemesine ilişkin bulgular sunulmuştur. Tabloda ayrıca LR çözümlemesine ilişkin doğru sınıflandırma oranları da verilmiştir. İç çarpım çekirdeği olarak özellikle Gauss Radyal Tabanlı Fonksiyon seçildiğinde; DVM'nin, LR çözümlemesine göre daha iyi sonuçlar verdiği gözlenmiştir. Test veri seti için lojistik modelden elde edilen doğru sınıflandırma oranının, DVM yönteminde doğrusal ve doğrusal olmayan modellerden elde edilen doğru sınıflandırma oranlarından daha küçük kalması, LR'nin genelleme başarısının daha düşük olduğu sonucunu ortaya çıkartmaktadır.

Tablo 4. Düşük doğum ağırlıklı bebeklerin kestirim probleminde LR ve DVM yaklaşımının, doğru sınıflandırma oranlarına göre karşılaştırması

Yöntem		**Eğitim**	**Test**
Lojistik Regresyon		73.8	65.1
Destek Vektör Makineleri			
Çekirdek Fonksiyonu	Doğrusal ($C = 100$)	78.6	66.7
	Polinomiyal ($q = 2$, $C = 0.32$)	78.6	68.2
	Polinomiyal ($q = 3$, $C = 700$)	79.4	65.1
	Gauss Radyal Tabanlı Fonksiyon ($C = 2100$)	81.7	69.8

4.2 Prostat Kanser Riskli Bireylerin Kestirim Problemi

Çalışmanın bu kesiminde, prostat kanser riskli bireylerin kestirimine ilişkin bir sınıflandırma modeli geliştirmek üzere; LR ve DVM çözümlemesi kullanılacak ve elde edilen bulgular karşılaştırılacaktır. Bu amaçla; Hosmer and Lemeshow (2000) çalışmasında da kullanılmış olan, 376 birey üzerinde yapılan bir araştırma neticesinde derlenen veriler kullanılmıştır. Veri seti, 251 (2/3)'i eğitim ve 125 (1/3)'i ise test verisi olacak biçimde rasgele iki parçaya ayrılmıştır.

Girdi ve çıktı değişkenleri aşağıdaki gibidir:
y : Prostat kanserinin tümör penetrasyonu ($0 =$ Penetrasyon yok; $1 =$ Penetrasyon var)
x_1: Yaş
x_2: Irk (1 = Beyaz; 2 = Siyahi)
x_3: Dijital rektal muayene sonucu ($1 =$ Nodül yok; $2 =$ Tek loplu nodül (Sol); $3 =$ Tek loplu nodül (Sağ); $4 =$ İki loplu nodül)
x_4: Rektal muayenede kapsül tutulumu tespiti ($0 =$ Hayır; $1 =$ Evet)
x_5: PSA düzeyi (mg/ml)
x_6: Ultrasonla elde edilen antijen hacmi (cm^3)
x_7: Gleason skoru (0 – 10)

Bağımlı değişken ile ilişkili olabileceği düşünülen olası değişkenlerin basit LR çözümlemesi sonuçları Tablo 5'te verilmiştir. Yaş (x_1) değişkeni ile prostat kanserinin tümör penetrasyonu arasında doğrusal bir ilişki olmadığının belirlenmesinden dolayı, x_1 değişkeninin çoklu LR modeline aday değişkenler arasından çıkarılmasına karar verilmiştir. Ayrıca, gleason skoru (x_7 değişkeni) ile kurulan basit LR modelinde model uyumunun diğerlerine göre daha iyi ("-2(log-olabilirlik)" değeri küçük) ve doğru sınıflandırma oranının yüksek olması dikkat çekmektedir.

Çoklu LR modelinin kurulması aşamasında ikinci adım olarak değişkenlerin seçimi; adımsal seçim yöntemlerinden olabilirlik oran test istatistiğine dayalı geriye doğru eleme yöntemi kullanılarak yapılmıştır. Böylece ırk (x_2), dijital rektal muayene sonucu (x_3), rektal muayenede kapsül tutulumu tespiti (x_4), PSA düzeyi (x_5) ve gleason skoru (x_7) değişkenleri ile oluşturulan lojistik modele ilişkin çözümleme sonuçları Tablo 6'da verilmiştir. ODDS oranları incelendiğinde; dijital rektal muayene sonucunun (x_3), rektal muayenede kapsül tutulumu tespitinin (x_4) ve gleason skorunun (x_7), prostat kanserinin tümör penetrasyonunun ortaya çıkması olasılığını önemli düzeyde arttırdığı belirlenmiştir.

Modelin eğitim veri setine uyumunu test etmek üzere Hosmer-Lemeshow testi kullanılmıştır. Buna göre, gözlenen ve beklenen değerler arasında fark olmadığı, modelin veriye uyumunun oldukça iyi olduğu sonucuna ulaşılmıştır ($\mathcal{X}^2 = 1.883, sd = 8, p = 0.984$

Tablo 5. Prostat kanser riskli bireylerin kestirim probleminde tek değişkenli lojistik regresyon çözümlemesine ilişkin sonuçlar

Değişken	$\hat{\beta}$	$SE(\hat{\beta})$	**Wald**	sd	p	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ **için %95 Güven Aralığı**	**-2(log-olabilirlik)**	**Doğru Sınıflandırma Oranı**
x_1	-0.022	0.020	1.246	1	0.264*	0.978	[0.940, 1.017]	333.653	0.614
x_2	-0.507	0.440	1.328	1	0.249	0.602	[0.254, 1.427]	333.503	0.614
x_3	-	-	18.138	3	0.000	-	-	314.802	0.641
x_3 (1)	0.869	0.383	5.140	1	0.023	2.385	[1.125, 5.055]		
x_3 (2)	1.341	0.400	11.210	1	0.001	3.822	[1.744, 8.379]		
x_3 (3)	1.768	0.449	15.496	1	0.000	5.861	[2.430, 14.135]		
x_4	1.876	0.455	17.030	1	0.000	6.527	[2.678, 15.910]	314.395	0.677
x_5	0.042	0.010	18.035	1	0.000	1.043	[1.023, 1.064]	308.249	0.669
x_6	-0.016	0.008	3.443	1	0.064	0.985	[0.969, 1.001]	331.268	0.614
x_7	1.239	0.191	41.925	1	0.000	3.452	[2.373, 5.023]	273.970	0.713

$* \, p > 0.25$

Tablo 6. Prostat kanser riskli bireylerin kestirim probleminde değişken seçiminin ardından gerçekleştirilen çoklu LR çözümlemesine ilişkin sonuçlar

Değişken	$\hat{\beta}$	$SE(\hat{\beta})$	**Wald**	sd	p	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ **için %95 Güven Aralığı**
x_2	-1.045	0.552	3.586	1	0.058	0.352	[0.119, 1.037]
x_3	-	-	7.765	3	0.051	-	-
x_3 (1)	0.527	0.442	1.425	1	0.233	1.695	[0.713, 4.028]
x_3 (2)	1.215	0.459	7.014	1	0.008	3.371	[1.371, 8.287]
x_3 (3)	0.918	0.544	2.845	1	0.092	2.503	[0.862, 7.272]
x_4	1.059	0.535	3.926	1	0.048	2.884	[1.011, 8.224]
x_5	0.026	0.011	5.381	1	0.020	1.026	[1.004, 1.049]
x_7	0.956	0.203	22.077	1	0.000	2.602	[1.746, 3.877]
Sabit	-7.725	1.312	34.649	1	0.000	0.000	-

LR çözümlemesine göre; ırk (x_2), dijital rektal muayene sonucu (x_3), rektal muayenede kapsül tutulumu tespiti (x_4), PSA düzeyi (x_5) ve gleason skoru (x_7) değişkenleri ile prostat kanserinin tümör penetrasyonu (y) arasındaki bağıntı,

$$\mu(\mathbf{x}) = 1/[1 + \exp(-7.725 - 1.045x_2 + 0.527x_3(1) + 1.215x_3(2) + 0.918x_3(3) \\ +1.059x_4 + 0.026x_5 + 0.956x_7)]$$

biçimindedir.

Tablo 7. Prostat kanser riskli bireylerin kestirim probleminde LR ve DVM yaklaşımının, doğru sınıflandırma oranlarına göre karşılaştırması

Yöntem		**Eğitim**	**Test**
Lojistik Regresyon		76.1	72.8
Destek Vektör Makineleri			
Çekirdek Fonksiyonu	Doğrusal $(C = 2500)$	76.5	74.4
	Polinomiyal $(q = 2, C = 17300)$	77.3	75.2
	Polinomiyal $(q = 3, C = 9700)$	76.9	74.4
	Gauss Radyal Tabanlı Fonksiyon $(C = 13200)$	83.7	72.0

DVM yöntemine göre ise x_2, x_3, x_4, x_5, x_7 girdileri ve doğum ağırlığı çıktısı için doğrusal model tahmini,

$$D(\mathbf{x}) = -9.477 - 2.817x_2 + 0.665x_3(1) + 1.513x_3(2) + 1.310x_3(3) \\ +3.751x_4 + 0.051x_5 + 1.147x_7$$

biçiminde elde edilmiştir. Bu uygulamada iteratif arama sonucunda $C = 2500$ olarak seçilmiştir.

LR çözümlemesine ile farklı çekirdek fonksiyonlarının kullanımına bağlı olarak doğrusal ve doğrusal olmayan DVM çözümlemesi sonucunda elde edilen bulgular Tablo 7'de sunulmuştur. Buna göre, eğitim ve test veri setleri için doğru sınıflandırma oranları incelendiğinde DVM yaklaşımının daha iyi sonuçlar ürettiği belirlenmiştir. Ayrıca test veri setinde görülen sınıflandırma hata sayısının az olması, DVM yönteminin genelleme kapasitesinin daha yüksek olduğu biçiminde yorumlanır.

5. SONUÇ

Uygulamalı istatistikte, gözlemlerin ayrık iki kategoride sınıflandırılması biçimindeki problemler ile sıklıkla karşılaşılmaktadır. Varsayımlarının doğrulanması durumunda sınıflandırma yöntemleri, yeni gözlemleri ait olduğu gruplara, eğitim örnekleri yardımıyla bir ayırıcı fonksiyona bağlı olarak atanmasını sağlar. Bu çalışmada, düşük doğum ağırlıklı bebeklerin ve prostat kanser riskli hastaların kestirimi biçimindeki iki sınıflandırma probleminde LR ve DVM yöntemi karşılaştırılmıştır. Doğrusal model durumunda, her iki uygulama için DVM algoritmasının, LR çözümlemesine göre eğitim ve test veri setinde daha küçük sınıflandırma hata oranına sahip olduğu belirlenmiştir. Doğrusal olmayan model için ise çekirdek fonksiyonu olarak polinomiyal veya Gauss Radyal Tabanlı Fonksiyon seçildiğinde; DVM'nin sınıflandırma başarısının yükseldiği gözlemlenmiştir.

İstatistiksel öğrenme teorisine dayalı olan DVM, mevcut veri setini kullanarak model geliştirmek için genelleme düzeyi iyi olacak biçimde öğrenme yöntemini uygular. Böylece DVM, eğitim veri seti üzerinde model tahminini, önceden gözlenmemiş veriler için kestirim performans kaybına yol açmaksızın gerçekleştirir. DVM'nin dezavantajı, sınıflandırma sonucunun bütünüyle ikili sınıflandırma biçiminde olması ve sınıf üyeliklerine ilişkin olasılıkların verilmemesidir.

LR çözümlemesi gibi klasik istatistiksel yöntemler, eğitim veri setine uygunluk optimum olacak biçimde bir sınıflandırma modeli geliştirmeyi amaçlar. Bu strateji, eğitim veri setine uygun model tahminini kolaylıkla verse de model tahmini, önceden gözlenmemiş veriler (test veri seti) için iyi kestirimler üretmekte yetersiz kalmaktadır. Bu problem, genellikle eğitim veri setine aşırı uyum (overfitting) olarak değerlendirilir ve önceden bilinmeyen örnekler için zayıf genelleme düzeyine neden olur. LR'nin başka bir dezavantajı ise veri seti üzerinde, muhtemel doğrusal olmayan yapıları belirleyebilen bir teknik olmamasıdır. LR'nin aksine DVM, daha karmaşık karar sınırları üretmek üzere tasarlanmıştır.

KAYNAKLAR

Brown, M.P., Grudy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M. and Haussler, D. (2000). Knowledge-Based Analysis of Microarray Gene Expression Data by using Support Vector Machines. *Proceedings of National Academy of Sciences*, 97(1): 262–267.

Buescher, P.A., Larson, L.C., Nelson, M.D. and Lenihan, A.J. (1993). Prenatal WIC Participation Can Reduce Low Birth Weight and Newborn Medical Costs: A Cost Benefit Analysis of Wic Participation in Nort Carolina. *Journal of the American Dietetic Association*, 93:163-166.

Cherkassky, V. and Mulier, F. (2007). *Learning From Data: Concepts, Theory, and Methods*. New Jersey: John Wiley & Sons.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press.

Cox, B.D., Whichelow, H.J. and Prevost, A.T. (1998). The Development of Cardiovascular Disease in Relation to Anthropometric Indices and Hypertension in British Adults. *International Journal of Obesity*, 22(10): 966–973.

Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics*, 35: 352–359.

Gardside, P.S. and Glueck, C.J. (1995). The Important Role of Modifiable Dietary and Behaviour Characteristic in The Causation and Prevention of Coronary Heart Disease Hospitalization and Mortality. *Journal of American College of Nutrition*, 14: 71-79.

Gunn, S.R. (1998). Support Vector Machines for Classification and Regression, Technical Report, University of Southampton.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression*. Newyork: John Wiley& Sons.

Jaakkola, T.S. and Haussler, D. (1998). Exploiting Generative Models in Discriminative Classifiers, in: Kearns, M.S., Solla, S.A. and Cohn, D.A. (Eds.), *Advances in Neural Information Processing Systems*, Cambridge: MIT Press.

Joachims, T. (1998). Text Categorization with Support Vector Machines. Proceedings *of The European Conference on Machine Learning* (ECML).

Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. Cambridge, Massachusetts: MIT Press.

Kleinbaum, D.G. (1994). *Logistic Regression: A Self Learning Text*. Newyork: Springer-Verlag.

Kloiber, L.L., Winn, N.J., Shaffer, S.G. and Hassanein, R.S. (1996). Late Hyponatremia in Very Low Birth Weight Infants: Incidence and Associated Risk Factors. *Journal of the American Dietetic Association*, 96: 880-884.

Komarek, P. (2004). *Logistic Regression for Data Mining and High-Dimensional Classification*, Robotics Institute.

Menard, S. (1995). *Applied Logistic Regression Analysis*. California: Sage Publications.

Min, R. and Cheng, H.D. (2009). Effective Image Retrieval using Dominant Color Descriptor and Fuzzy Support Vector Machine. *Pattern Recognition*, 42: 147–157.

Minka, T.P. (2001). Algorithms for Maximum-Likelihood Logistic Regression. Technical Report Stats 758, Carnegie Mellon University.

Osuna, E., Freund, R. and Girosi, F. (1997). Training Support Vector Machines: An Application to Face Detection. *Proceedings of Computer Vision and Pattern Recognition*, 130–136.

Rosenberg, A.A. and Battaglia, F.C. (1991). The Newborn Infant, Current Pediatric Diagnosis and Treatment, in: Hathaway, W.E., Groothuis, J.R., Hay, W.W. and Paisley, J.W. (Eds.), 50–103.

Schölkopf, B. and Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press.

Strang, G. (1986). *Introduction to Applied Mathematics*. Wellesley: Wellesley-Cambridge Press.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Newyork: Springer.

Vapnik, V. (1998). *Statistical Learning Theory*. Newyork: John Wiley & Sons.

Yang, Z.R. (2004). Biological Applications of Support Vector Machines. *Briefings in Bioinformatics*, 5(4): 328–338.

Yang, C.H., Jin, L.C. and Chuang, L.Y. (2006). Fuzzy Support Vector Machines for Adaptive Morse Code Recognition. *Medical Engineering and Physics*, 28: 925–931.

Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T. and Müller, K.-R. (2000). Engineering Support Vector Machine Kernels that Recognize Translation Initiation Sites. *Bioinformatics*, 16(9): 799–807.