

KANSER VERİLERİNİN SINIFLANDIRILMASINDA YAPAY SİNİR AĞLARI İLE DESTEK VEKTÖR MAKİNELERİ'NİN KARŞILAŞTIRILMASI

Semra ERPOLAT*
Ersoy ÖZ**

Özet

Sınıflandırma, bilinmeyen herhangi bir verinin belli kurallara göre düzenlenmiş olan veri kümesindeki sınıflardan hangisine ait olduğunu tahmin etme işlemidir. Birçok alanda önemli bir yere sahip olan sınıflandırmanın özellikle insan hayatının söz konusu olduğu tıbbi alandaki yeri daha da önemlidir. Bu çalışmada makine öğrenmesi yöntemlerinden olan Yapay Sinir Ağları (YSA) ve Destek Vektör Makineleri (DVM)'nin sınıflandırma başarıları tıbbi veriler üzerinde sınanmıştır. Bu amaçla UCI Makine Öğrenmesi Ambarları'ndan alınan göğüs kanseri verilerinden yararlanılmıştır. Elde edilen sonuçlardan iki sınıflandırma yönteminin de oldukça iyi sonuçlar verdiği ancak YSA'nın DVM'ye oranla daha iyi sonuçlar verdiği gözlemlenmiştir.

Anahtar Kelimeler: Destek Vektör Makineleri, Yapay Sinir Ağları, Sınıflandırma.

* *Yrd. Doç. Dr., Mimar Sinan Güzel Sanatlar Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü*

** *Dr., Yıldız Teknik Üniversitesi Meslek Yüksekokulu Teknik Programlar Bölümü*

Comparison of Artificial Neural Network and Support Vector Machine for Classifying Cancer Data

Abstract

Classification is the process of estimating to which class inside the data set, that are organized by some specific rules, does any unknown data belong. Classification is very important in many areas, especially in the medical area where the subject is human health. In this study, the classification successes of the two machine learning methods, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), on medical data were tested. To this end, breast cancer data taken from the UCI Machine Learning Repository were benefited from. From the obtained results, it was seen that both methods gave good results but it was observed that ANNs gave better results than SVMs.

Key Words: Support Vector Machine, Artificial Neural Network, Classification.

1.Giriş

Amacı veri gruplarını oluşturan sınıf nitelikleri ve bu sınıfların çıkış karakterleri arasındaki ilişkileri keşfederek, keşfedilen ilişkiler ile yeni bir veri örneğinin sınıf etiketini tahmin etmek olan sınıflandırma, bilinmeyen bir veri örneğinin sonlu sayıda sınıftan hangisine ait olduğunun tahmin edilmesi işlemidir (Erpolat, 2010).

Sınıflandırma problemlerinin çözümünde genellikle istatistiksel yöntemler, makine öğrenmesi yöntemleri ve uzman sistemler kullanılmaktadır. Sınıflandırmada kullanılan istatistiksel yöntemler parametrik ve parametrik olmayan olmak üzere iki gruba ayrılır. Parametrik yöntemler yorumlanması kolay ve güvenilir sonuçlar üretirken, parametrik olmayan yöntemlerin güvenilirlikleri daha düşüktür. İstatistiksel yöntemlerin en büyük dezavantajı gerçek dünya verilerinin genellikle parametrik dağılım varsayımına uygun olmayışıdır. Parametrik yöntemler içinde yaygın kullanılan yöntemler; Fisher Doğrusal Diskriminant Analizi (LDA), Quadratik ve Lojik Diskriminant Analizi iken parametrik olmayan yöntemler içinde yaygın kullanılanları k -en yakın komşu, lojistik regresyon ve lineer programlama yöntemleridir. Sınıflandırma için kullanılan makine öğrenimi yöntemleri ilişkili ve tümevarımsal olmak

üzere iki gruba ayrılır. İlişkili yöntemlerden en çok kullanılanı geri yayılım algoritması iken tüme varımsal yöntemler için CART, ID3, ve CN2 algoritmalarıdır. En çok bilinen makine öğrenmesi yöntemleri olarak YSA, DVM ve karar ağaçları sayılabilir. Sınıflandırma için son yıllarda dikkat çeken bir diğer yöntemler grubu ise genetik algoritmalar ve genetik programlamayı içeren evrimsel hesaplamalardır (Erpolat, 2009).

Tam (1991) finans durumu tehlikede olan şirketleri tahmin etmek için YSA kullanmış ve elde ettiği sonuçları MDA, lojistik regresyon, k -en yakın komşu ve ID3'le kıyaslayarak YSA'nın tahmin konusunda en başarılı yöntem olduğunu rapor etmiştir. Tam ve Kiang (1992) istatistiksel yöntemler ve karar ağaçlarını kullanarak finans durumu tehlikede olan şirketleri tahmin etmede en başarılı yöntemin YSA'nın olduğunu rapor etmişlerdir. Shin vd. (2005) şirketlerin iflas durumunu araştırmak için DVM'yi kullanmışlar ve şirketlerin mali durumuna göre sınıflandırılmasında DVM'nin MLFF-BP'den daha başarılı olduğunu açıklamışlardır.

Bu çalışmada özellikle makine öğrenmesi alanında sınıflandırma amacıyla kullanılan YSA ve DVM yöntemlerinin, göğüs kanseri verilerini sınıflandırmadaki başarısı karşılaştırılmıştır. Bu amaçla UCI Makine Öğrenmesi Ambarları'ndan (http://www.ics.uci.edu/~mlearn/ML_Repository.html, 16 Eylül 2010) alınan kanser verileri kullanılmıştır. Elde edilen sonuçlardan iki sınıflandırma yönteminin de oldukça iyi sonuçlar verdiği ancak YSA'nın DVM'ye oranla daha iyi sonuçlar verdiği gözlemlenmiştir.

İzleyen 2. ve 3. bölümlerde sırasıyla çalışmada kullanılan iki yöntem olan YSA ve DVM ayrıntılı biçimde incelenmiş, 4. bölümde k -katlı çapraz doğrulama yöntemi açıklanmış, 5. bölümde yöntemlerin karşılaştırıldığı uygulama ele alınmış ve son bölümde ise sonuçlar değerlendirilmiştir. Yöntemlerin açıklandığı 2,3 ve 4. bölümler için ana kaynak olarak Erpolat (2010) kullanılmıştır.

2. Yapay Sinir Ağları

YSA, insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip sinir hücrelerinden (nöronlar) oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır (Elmas, 2007). Biyolojik sinir ağlarını taklit eden bilgisayar programları olarak da bilinen YSA'lar; genel olarak matematiksel modelleri oluşturulamayan veya

çok zor tanımlanabilen problemlerin çözümü için kullanılmaktadır (Sağıroğlu, 2003).

Çalışmada, en çok kullanılan aktivasyon fonksiyonlarından birisi olan ve Eşitlik 1 'de verilen sigmoid fonksiyonu kullanılmıştır.

$$y = F(v) = \frac{1}{1 + e^{-v}} = \frac{1}{2} [\tanh(v/2) + 1] \quad (1)$$

Ayrıca; YSA türü olarak bir giriş katmanı, bir veya birden fazla gizli katman ve bir çıkış katmanından oluşan Çok Katmanlı Perseptron ile bu ağı eğitecek algoritma olarak ise hataları çıkıştan girişe geriye doğru azaltmaya çalışmasından dolayı "geri yayılım" ismini alan algoritma tercih edilmiştir. Geri yayılım algoritmasının işleyişi adım adım aşağıdaki şekilde gerçekleşir (Fausett, 1994):

Adım 0: Ağırlık ve eşik değerleri küçük gerçel sayılar olacak şekilde rasgele atanır.

Adım 1: Durdurma kriterleri yanlış olduğu sürece Adım 2-11 tekrarlanır.

Adım 2: Her eğitim verisi için Adım 3-10 çalıştırılır.

Adım 3: Her x_i ($i=1,2,\dots,n$) giriş sinir hücresine eğitim verisi atanır.

Adım 4: Gizli katmandaki her yt_j ($j=1,2,\dots,p$) sinir hücresi için giriş değerlerinin ağırlıklı toplamı w_{ij} giriş katmanının i . sinir hücresi ile gizli katmanın j . sinir hücresi arasındaki ağırlık, θ_j gizli katmanın j . sinir hücresi için eşik değeri olmak üzere Eşitlik 2' deki gibi hesaplanır.

$$yt_j = \sum_{i=1}^n w_{ij} x_i + \theta_j \quad (2)$$

KANSER VERİLERİNİN SINIFLANDIRILMASINDA YAPAY SİNİR AĞLARI İLE DESTEK VEKTÖR MAKİNELERİNİN KARŞILAŞTIRILMASI

Gizli katmandaki sinir hücrelerinin çıkışını bulmak için aktivasyon fonksiyonunun yt_j 'ye uygulanmış hali Eşitlik 3'teki gibidir. Bu işlem sonucunda bulunan sonuç bir sonraki katmana gönderilir.

$$yt_j = f(yt_j) \quad (3)$$

Adım 5: Çıkış katmanının her z_k ($k=1,2,\dots,m$) sinir hücresi için giriş değerlerinin ağırlıklı toplamı Eşitlik 4'deki gibi hesaplanır. Eşitlikte yer alan v_{jk} değeri gizli katmanın j . sinir hücresi ile çıkış katmanının k . sinir hücresi arasındaki ağırlığı, ξ_k değeri ise çıkış katmanının k . sinir hücresi için eşik değeridir.

$$zt_k = \sum_{j=1}^p v_{jk}y_j + \xi_k \quad (4)$$

Gizli katmandaki sinir hücrelerinin çıkışını bulmak için aktivasyon fonksiyonunun zt_k 'ya uygulanmış hali Eşitlik 5'deki gibidir.

$$z_k = f(zt_k) \quad (5)$$

Adım 6: Çıkış katmanının her z_k ($k=1,2,\dots,m$) sinir hücresi için beklenen t_k değeri ile hesaplanan z_k değeri karşılaştırılıp hata değeri Eşitlik 6'daki hesaplanır.

$$\delta_k = (t_k - z_k) f'(zt_k) \quad (6)$$

Adım 7: Eğitim hızı α olmak üzere ağırlıklardaki ve eşikteki değişim Eşitlik 7'deki gibi hesaplanır.

$$\begin{aligned}\Delta w_{jk} &= \alpha \delta_k y_j \\ \Delta \xi_k &= \alpha \delta_k\end{aligned}\quad (7)$$

Adım 8: Gizli katmanın her y_j ($j=1,2,\dots,p$) sinir hücresi için ağırlıklardaki ve eşikteki değişim Eşitlik 8'deki gibi hesaplanır.

$$\begin{aligned}\delta t_j &= \sum_{k=1}^m \delta_k v_{kj} \\ \delta_j &= \delta t_j f'(y_j) \\ \Delta v_{ij} &= \alpha \delta_j x_i \\ \Delta \theta_j &= \alpha \delta_j\end{aligned}\quad (8)$$

Adım 9: Çıkış katmanının her z_k ($k=1,2,\dots,m$) sinir hücresi için ağırlıklar ve eşik Eşitlik 9'daki gibi güncellenir.

$$\begin{aligned}v_{jk} &= v_{jk} + \Delta v_{jk} \\ \xi_k &= \xi_k + \Delta \xi_k\end{aligned}\quad (9)$$

Adım 10: Gizli katmanın her y_j ($j=1,2,\dots,p$) sinir hücresi için ağırlıklar ve eşik Eşitlik 10'daki gibi güncellenir.

$$\begin{aligned}w_{jk} &= w_{jk} + \Delta w_{jk} \\ \theta_k &= \theta_k + \Delta \theta_k\end{aligned}\quad (10)$$

Adım 11: Durdurma kriterlerine bakılır.

3. Destek Vektör Makineleri

Vapnik ve Chervonenkis tarafından istatistiksel öğrenme teorisine dayalı olarak türetilmiş yeni bir kontrollü öğrenme yöntemi (Scholkopf ve Smola,

2002) olan DVM, dağılımdan bağımsız olarak çalışan ve eğitici veya yarı eğitici olarak özellikle iki sınıflı sınıflandırma ve regresyon işlemlerini gerçekleştiren sağlam ve etkin bir yöntemdir (Cortes ve Vapnik, 1995). DVM'nin temel amacı iki sınıf arasında en uygun ayrımı yapabilen hiper düzlemi tanımlamaktır (Vapnik, 2000).

Özellikleri belirtilen bir veri kümesine ait sınıf etiketlerini öğrenebilen veya tahmin edebilen bir model oluşturan DVM'nin işleyişi kısaca; eğitim kümesindeki verilerin özellik vektörü olarak ifade edilmesi, bu özellik vektörlerinin çekirdek (kernel) fonksiyonları kullanılarak özellik uzayına eşlenmesi, sınıfları en uygun şekilde ayıran bir hiper düzlemin oluşturulması ve oluşturulan düzlemin geçerliliğinin (genelleme yeteneğinin) test veri kümesi olarak adlandırılan bağımsız veri kümeleri kullanılarak sınanması şeklindedir. Çalışmada çekirdek fonksiyonu olarak Eşitlik 11'de verilen Yarıçap Temelli Fonksiyon-YTF (Radial Basis Function, RBF) kullanılmıştır.

$$K(x, y) = K(x - y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (11)$$

Veri kümelerini sınıflara ayırmak için oluşturulan hiper düzlemler arasında maksimum sınıra sahip sadece bir hiper düzlem vardır. Bu düzleme "ayırıcı hiper düzlem" ve bu düzleme en yakın vektörlere de "destek vektörleri" adı verilir. Hiper düzlem sayısı, sınıflandırılacak verinin boyutuna bağlıdır.

DVM, YRM prensibinin ve Vapnik-Chervonenkis (VC) teorisinin uygulandığı bir yakınsama yöntemidir. YRM prensibi ve VC teorisindeki temel fikir; çok sayıda aday model arasından beklenen risk veya genelleme hatasını minimum yapacak doğru oranda karmaşıklığa sahip modelin seçilmesidir.

DVM ile sınıflandırılacak eğitim veri kümesinin N sayıda örnekten oluştuğu ve $i = 1, \dots, N$ olmak üzere $\{x_i, y_i\}$ ile gösterildiği varsayalım.

Burada $x_i \in R^d$ olmak üzere d -boyutlu bir uzayda özellikler vektörünü (giriş vektörünü), $y_i \in \{-1, +1\}$ olmak üzere sınıf etiketlerini (çıkış vektörünü) göstermektedir. w hiper düzlemin normal vektörü ve b eğim değeri olmak eğitim kümesinin Eşitsizlik 12'deki koşulları yerine getirmesi gerekmektedir.

$$y_i(w \cdot x_i + b) \geq +1, \quad i = 1, \dots, N \quad (12)$$

4. *k*-Katlı Çapraz Doğrulama Yöntemi ve Başarı Ölçümü

Herhangi bir amaç için oluşturulan sistemlerin yürütülmesi sonucunda elde edilen sonuçların doğruluklarının ölçülmesi suretiyle sistemlerin başarı değerlendirilmesi yapılabilmektedir. Bu amaçla kullanılan en yaygın yöntemlerden biri de *k*-katlı çapraz doğrulama yöntemidir. Bu yöntemde *A* olarak temsil edilen veri kümesi her birinde yaklaşık aynı sayıda veri olacak şekilde A_1, A_2, \dots, A_k biçiminde gösterilen rastgele *k* sayıda alt kümeye ayrılır. Yöntemin veri üzerinde sınanmasında *k* sayıda alt kümeden her seferinde bir küme “test kümesi” olarak, geri kalanlar ise “eğitim kümesi” olarak ele alınır.

Deneysel çalışmalar *k*-katlı çapraz doğrulama yönteminde *k* için optimum değer 10 olduğunu göstermiştir (Breiman vd., 1984). Fakat veri kümesinin az sayıda olduğu bazı durumlarda bu sayı 2 veya 5 olabilmektedir. Çalışmada 10-katlı çapraz doğrulama yöntemi kullanılmıştır.

k tane farklı eğitim ve test kümeleriyle eğitilen ve sınanan sistemin yine *k* tane başarı ölçüsü elde edilmiş olur. Bu nedenle sistemin genel başarısını belirlemek üzere, elde edilen bu ölçümlerin ortalaması alınır. Sistemin Sınıflandırma Başarı Oranı (*SBO*), doğru sınıflandırılan veri sayısının (*dsvs*) toplam veri sayısına (*tvs*) oranlanmasıyla elde edilir. Eşitlik 13'te sırasıyla sistemin *SBO* ve ortalama *SBO* yer almaktadır.

$$SBO = \frac{dsvs}{tvs}, \quad \text{ortalama } SBO = \frac{\sum_{i=1}^k SBO_i}{k} \quad (13)$$

5. Uygulama

5.1. Veri Kümesinin Açıklanması

Yukarıda açıklanan sınıflandırma yöntemlerinin kanser verileri üzerindeki sınıflandırma başarılarını karşılaştırmak amacıyla, araştırmacılara makina öğrenmesi ve bilgi sistemleri için uluslararası düzeyde kaynak veri sağlayan bir merkez olan UCI Machine Learning Repository'den (<http://www.ics.uci.edu/~mllearn/MLRepository.html>, 16 Eylül 2010) alınan göğüs kanseri verileri kullanılmıştır.

Söz konusu göğüs kanseri verisi Dr. William H. Wolberg tarafından (Wisconsin Üniversitesi, Madison) düzenlenmiş olup toplam 699 gözlemi

KANSER VERİLERİNİN SINIFLANDIRILMASINDA YAPAY SİNİR AĞLARI İLE DESTEK
VEKTÖR MAKİNELERİNİN KARŞILAŞTIRILMASI

içermektedir. Veri kümesinin nitelik ve özelliklerine ilişkin bilgiler Tablo 1'deki gibidir.

Tablo 1: Veri Kümesinin Açıklaması.

Nitelik Adı	Açıklama	İçeriği	Türü
ID	Hasta No	1-10	Girdi
Clump	Kitle kalınlığı	1-10	Girdi
UnifSize	Hücre boyutunun değişkenliği	1-10	Girdi
UnifShape	Hücre şeklinin değişkenliği	1-10	Girdi
MargAdh	Majinal yapışma	1-10	Girdi
SingEpiSize	Tek epitel hücre boyutu	1-10	Girdi
BareNuc	Sade çekirdek	1-10	Girdi
BlandChrom	Yumuşak kromatin	1-10	Girdi
NormNucl	Normal nukleoli	1-10	Girdi
Mit	Mitoses	1-10	Girdi
Class	İyi huylu veya kötü huylu	2: iyi huylu 4: kötü huylu	Çıktı

Tablodan, çalışmada kullanılan veri kümesinin 1'den 10'a kadar değer alabilen 10 tane girdi verisinin ve iki sınıftan oluşan bir adet çıktı verisinin olduğu görülmektedir.

Çalışmada göğüs kanseri hastalarını iyi huylu veya kötü huylu olarak sınıflandırmak amacıyla kullanılan YSA ve DVM yöntemleri, SPSS Clementin 12.0 paket programı ile gerçekleştirilmiştir.

5.2. Analizlerin Gerçekleştirilmesi

Çalışmada yöntemleri karşılaştırmak üzere öncelikle veri kümesi 10-katlı çapraz doğrulama yöntemine göre eğitim ve test kümelerine bölünmektedir. Söz konusu bölünme Tablo 2'de gösterildiği gibidir.

Tablo 2: Eğitim ve Test Kümelerinin Düzeni.

Alt Küme No	1	2	3	4	5	6	7	8	9	10
İyi Huylu Gözlem Sayısı	46	46	46	46	46	46	46	46	45	45
Kötü Huylu Gözlem Sayısı	25	24	24	24	24	24	24	24	24	24
Toplam	71	70	70	70	70	70	70	70	69	69

Tablo 2’den 1-inci çapraz doğrulamada iyi huylu gözlemden 46, kötü huylu gözlemden 25 olmak üzere toplam 71 gözlem içeren birinci alt küme “test kümesi”, geri kalan gözlemlerin tümü ise “eğitim kümesi” olarak ele alınmıştır. Bu durumda eğitim kümesinde 412 iyi huylu ve 216 kötü huylu olmak üzere toplam 628 gözlem yer almıştır. Diğer çapraz doğrulamalar için kullanılan test ve eğitim kümeleri de benzer biçimlerde oluşturulmuştur. Buna göre 2’den 8’e kadarki (2 ve 8 dahil) çapraz doğrulamaların tümü için 46 iyi huylu 24 kötü huylu olmak üzere toplam 70’er gözlemden oluşan “test kümeleri” ve 412’si iyi huylu ve 215’i kötü huylu olmak üzere toplam 627 gözlemden oluşan “eğitim kümeleri” oluşturulmuşken, 9 ve 10-uncu çapraz doğrulamalar için ise 45 iyi huylu 24 kötü huylu olmak üzere toplam 69’ar gözlemden oluşan “test kümeleri” ve 411’i iyi huylu ve 215’i kötü huylu olmak üzere toplam 626 gözlemden oluşan “eğitim kümeleri” oluşturulmuştur.

Yapılan denemeler sonucunda, 10-katlı çapraz doğrulama yöntemi sonucunda yöntemlerden elde edilen doğru sınıflandırılmış örnek sayıları ve SBO değerleri Tablo 3’te gösterildiği gibi elde edilmiştir.

KANSER VERİLERİNİN SINIFLANDIRILMASINDA YAPAY SİNİR AĞLARI İLE DESTEK
VEKTÖR MAKİNELERİ'NİN KARŞILAŞTIRILMASI

Tablo 3: Yöntemlerden Elde Edilen Bulgular.

Yöntem	k-ıncı Çapraz Doğrulama	Test Kümesi Örnek Sayısı	Doğru Sınıflandırılmış Örnek Sayısı	SBO (%)
YSA	1	71	68	95,77
	2	70	70	100,00
	3	70	68	97,14
	4	70	63	90,00
	5	70	69	98,57
	6	70	68	97,14
	7	70	69	98,57
	8	70	70	100,00
	9	69	67	97,10
	10	69	68	98,55
		ortalama	69,9	68
DVM	1	71	68	95,77
	2	70	67	95,71
	3	70	68	97,14
	4	70	61	87,14
	5	70	70	100,00
	6	70	67	95,71
	7	70	70	100,00
	8	70	70	100,00
	9	69	67	97,10
	10	69	69	100,00
		ortalama	69,9	67,7

Tablo 3'ten, göğüs kanseri verisinin YSA yöntemi ile 10-katlı çapraz doğrulamaya göre sınıflandırılması sonucunda elde edilen ortalama doğru sınıflandırılmış gözlem sayısının 68 ve ortalama SBO değerinin ise %97,284 olduğu buna karşılık DVM yöntemi ile 10-katlı çapraz doğrulamaya göre sınıflandırılması sonucunda ise elde edilen ortalama doğru sınıflandırılmış gözlem sayısının 67,7 ve ortalama SBO değerinin ise %96,857 olduğu görülmüştür. Buna göre, göğüs kanseri verisini sınıflandırmada YSA yönteminin DVM'den daha iyi sonuçlar verdiği söylenebilir.

SEMRA ERPOLAT
ERSOY ÖZ

6. SONUÇ

Özellikle makine öğrenmesi alanında sınıflandırma amacıyla kullanılan YSA ve DVM yöntemlerinin, göğüs kanseri verilerini sınıflandırmadaki başarısının karşılaştırıldığı çalışmada, UCI Makine Öğrenmesi Ambarları'ndan (<http://www.ics.uci.edu/~mllearn/MLRepository.html>, 16 Eylül 2010) alınan kanser verileri kullanılmıştır.

Yöntemler, 10-katlı çapraz doğrulamaya göre işletilmiş ve elde edilen sonuçların ortalaması alınmıştır. Sonuçlardan göğüs kanseri veri kümesi için her iki sınıflandırma yönteminin de oldukça iyi sonuçlar verdiği ancak YSA'nın DVM'ye oranla daha iyi olduğu gözlemlenmiştir.

KAYNAKÇA

BREIMAN L., FRIEDMAN J.H., OLSHEN R.A. ve STONE C.J. (1984), **Classification and Regression Trees**, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.

CORTES Corinna. ve VAPNIK Vladimir (1995), "Support Vector Networks", *Machine Learning*, Vol. 20, No. 3. s.1-31.

ELMAS Ç. (2007), **Yapay Zeka Uygulamaları**, Seçkin Yayıncılık San. ve Tic. A.Ş., Ankara.

ERPOLAT Semra (2009), "Hibrit Sınıflandırma Sistemi ve Kural Çıkarmada Farklı Bir Yaklaşım", *Marmara Üniv. Sosyal Bilimler Enst. İktisadi ve İdari Bilimler Fak., Ekonometri Bölümü, Yöneylem Araştırması Anabilim Dalı*, Doktora Tezi.

ERPOLAT Semra (2010), "Verileri Sınıflandırmada Farklı Bir Tasarım: Hibrit Sınıflandırma Sistemi", *19. İstatistik Araştırma Sempozyumu İAS2010*, Bildiriler Kitabı.

FAUSETT L. (1994), **Fundamentals of Neural Networks: Architectures, Algorithms and Applications**, Prentice-Hall, Upper Saddle River, NJ.

KANSER VERİLERİNİN SINIFLANDIRILMASINDA YAPAY SİNİR AĞLARI İLE DESTEK
VEKTÖR MAKİNELERİNİN KARŞILAŞTIRILMASI

SAĞIROĞLU S., BESDOK E. ve ERLER M. (2003), **Mühendislikte Yapay Zeka Uygulamaları-I: Yapay Sinir Ağları**, Ufuk Yayıncılık, Kayseri.

SCHOLKOPF B. ve SMOLA A.J. (2002), **Learning With Kernels: Support Vector Machines. Regularization and Beyond**, The MIT Press, Cambridge.

SHIN K.S., LEE T.S., ve KIM H.J. (2005), “An application of Support Vector Machines in Bankruptcy Prediction Model”, *Expert System Application*, Vol. 28, s.127–135.

TAM K.Y. (1991), “Neural Network Models and The Prediction of Bank Bankruptcy”, *Omega*, Vol. 19, No. 5, s.429-445.

TAM K.Y. ve KIANG M. (1992), “Predicting Bank Failures: A Neural Network Approach”, *Decision Science Applied Artificial Intelligence*, Vol. 4, s.265-282.

VAPNIK V.N. (2000), **The Nature of Statistical Learning Theory**, Second edition, Springer-Verlag, New York.

UCI Machine Learning Repository,
<http://www.ics.uci.edu/~mllearn/MLRepository.html>. (16 Eylül, 2010).