ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:6-Sayı/No: 2 : 175-182 (2005)

ARAŞTIRMA MAKALESİ /**RESEARCH ARTICLE**

TEK KATMANLI YAPAY SİNİR AĞLARI VE DOĞRUSAL DİSKRİMİNANT FONKSİYONUNUN KARŞILAŞTIRMASI

Güvenç ARSLAN[1], İlknur ÖZMEN[2]

ÖZ

Fisher doğrusal diskriminant fonksiyonu (Fisher’s linear discriminant function) ve doğrusal makinelerin (winner-take-all groups / linear machines) incelendiği bu çalışmada, öncelikle Fisher doğrusal diskriminant fonksiyonu ve tek katmanlı yapay sinir ağlarının (single layer neural networks) özel bir durumu olan doğrusal makineler hakkında bazı temel bilgiler verilmiştir. Daha sonra her iki yöntem, aynı veri kümesine uygulanarak güçlü ve zayıf yönleri vurgulanmaya çalışılmıştır.

Anahtar Kelimeler: Yapay sinir ağları, Fisher doğrusal diskriminant fonksiyonu, Doğrusal makineler.

COMPARISON OF SINGLE LAYER NEURAL NETWORKS AND LINEAR DISCRIMINANT FUNCTION

ABSTRACT

In this study Fisher’s linear discriminant function and winner-take-all groups, also known as linear machines, are studied. First, some basic information on Fisher’s linear discriminant function and on linear machines, a special type of single layer neural networks, is given. Both methods are then applied on the same data set in order to stress their strong and weak properties.

Key words: Neural networks, Fisher’s linear discriminant function, Linear machines.

1. GİRİŞ

Yapay sinir ağları basitçe birbirine bağlı hesaplama birimlerinden (hücre) oluşan bir sistem olarak düşünülebilir. Hücreler belli zamanlarda kendisine ulaşan girdilere karşılık bir çıktı hesaplamaktadır. Hesaplanan çıktılar bağlantılar aracılığı ile diğer hücrelere iletilir. Sonunda bazı hücrelerin çıktıları yapay sinir ağının çıktısı olarak düşünülür. Hücreler arasındaki bağıntılara farklı ağırlıklar vererek yapay sinir ağının davranışı değiştirilebilir. Temelde bir algoritma yardımı ile bu ağırlıklar istenen probleme göre ayarlanmaya çalışılır.

Çok değişkenli istatistiksel yöntemler, analize dahil edilen değişkene göre bağımlılık (dependence) analizleri (bir değişken ya da değişkenler kümesinin diğer-

[1] Başkent Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, **e-posta:** guvenca@baskent.edu.tr

[2] Başkent Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, **e-posta:** ilknur@baskent.edu.tr

Geliş: 30 Eylül 2004; **Düzeltme:** 14 Ekim 2004; **Kabul:** 26 Nisan 2005

lerine bağımlı olduğu ve onlarla açıklanabildiği analizler) ve karşılıklı bağımlılık (interdependence) analizleri (değişkenlerin bağımlı ya da bağımsız olarak ayrılmaksızın, değişkenler arasındaki ilişkilerin incelendiği analizler) olmak üzere iki grupta incelenir. Bağımlılık analizlerinden olan diskriminant analizinin (DA) iki temel amacı vardır: Birinci p değişken içeren k grubun çok değişkenli normal dağılım gösteren **X** veri matrisinden yararlanarak değişkenler arasındaki ilişkileri de yansıtan ve grupları birbirinden ayırmayı sağlayan fonksiyonlar bulmaktır. İkincisi ise hesaplanan bu fonksiyonlar aracılığı ile yeni bir gözlemi sınıflama hatası minimum olacak şekilde bu gruplardan birine atamaktır.

2. GENEL BİLGİLER

2.1. Diskriminant Analizi

Veri kümesindeki değişkenler aracılığı ile herhangi bir gözlemi iki ya da daha fazla gruplardan birine atayan DA'de, gözlenen verinin bir değişken açısından gruplara atanması kısmen kolay iken değişken sayısının çok olması durumunda, gözlenen veri kümesinin gruplardan birine atanması kolay değildir. Bazı değişkenler açısından bir gruba ait olabileceği düşünülen bir gözlemin, diğer değişkenler açısından başka bir gruba ait olabileceği düşünülebilir. Veri kümesindeki gözlemlere ilişkin p tane değişken dikkate alındığında, bunların gruplara atanmasında değişkenler arasındaki ilişkilerin de göz önünde bulundurulduğu bir göstergeye ihtiyaç vardır. Bu göstergeye "diskriminant" ya da "ayırma" fonksiyonu adı verilmektedir.

DA'nın uygulanabilmesi için aşağıda belirtilen varsayımların sağlanması gerekir:

- Birbirinden farklı iki ya da daha fazla grup olmalı,
- X veri matrisi çok değişkenli normal dağılım göstermeli,
- Her bir gruba ilişkin değişkenlerin varyans-kovaryans matrisleri eşit olmalı,
- Değişkenler arasında çoklu bağlantı (multicolinearity) olmamalıdır.

DA'de grupların varyans-kovaryans matrislerinin eşit olup olmamasına göre (Box's M testi) Fisher doğrusal (linear) diskriminant fonksiyonu ve karesel (quadratic) diskriminant fonksiyonu kullanılmaktadır. Her iki fonksiyonun temel amacı, ilgilenilen değişkenlere göre belirlenen diskriminant fonksiyonundan yararlanarak gözlemleri iki yada ikiden çok gruba ayırmak ve yeni gözlemleri uygun olan gruplara atamaktır. Fisher doğrusal diskriminant fonksiyonu, verilerin k tane normal dağılımlı ve ortak varyans-kovaryans matrisine sahip kitleden gelmesi durumu için geliştirilmiştir. Karesel diskriminant fonksiyonu ise verilerin normal dağıldığı ancak, grupların varyans-kovaryans matrislerinin farklı olduğu durumlarda kullanılmaktadır. İkiden çok grup $(k \geq 2)$ olması durumunda, iki ve daha çok değişken $(p \geq 2)$ içeren veri kümelerindeki grupları birbirinden ayırmada kullanılan çoklu doğrusal ve çoklu karesel diskriminant fonksiyonları iki grup için kullanılan eşitliklerin çok grup için genellemesidir. İki grup olması durumunda tek bir diskriminant fonksiyonu grupları birbirinden ayırmada yeterli iken çok grup olması durumunda diskriminant fonksiyonu sayısı $r = \min(k-1, p)$ olacak biçimde belirlenir ve doğrusal diskriminant fonksiyonu,

$$y_j = \boldsymbol{a}_j \boldsymbol{x} = a_1 x_1 + a_2 x_2 + \ldots + a_p x_p, \quad j = 1,\ldots,r \qquad \textbf{(1)}$$

biçiminde gösterilir. Burada $x_1, x_2, \ldots, x_p$ orijinal değişkenleri ve $a_1, a_2, \ldots, a_p$ ise bu değişkenlere ilişkin ağırlıkları ya da katsayıları göstermektedir (Tatlıdil, 1996).

Bulunan diskriminant fonksiyonlarından birincisi, gruplar arasındaki en büyük ayrımı sağlayan fonksiyondur. İkinci diskriminant fonksiyonu ise birinci diskriminant fonksiyonu ile ilişkili olmayan ve ilk diskriminant fonksiyonundan sonra gruplar arasında en iyi ayrımı sağlayan fonksiyondur. Öteki fonksiyonlar da benzer biçimde yorumlanır. Bu çalışmada, Fisher'in çoklu doğrusal diskriminant fonksiyonu incelendiği için sadece buna yönelik eşitlikler verilmiştir.

İkiden çok grup olması durumunda bulunacak diskriminant fonksiyonlarının elde edilmesi varyans oranlarının en büyüklenmesi temeline dayanır. Fisher tarafından tanımlanan bu oran,

$$\frac{\boldsymbol{a}'\mathbf{B}\boldsymbol{a}}{\boldsymbol{a}'\mathbf{W}\boldsymbol{a}} = \frac{\boldsymbol{a}'\left(\sum_{i=1}^{k} n_i (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})'\right)\boldsymbol{a}}{\boldsymbol{a}'\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij} - \bar{\boldsymbol{x}}_i)(x_{ij} - \bar{\boldsymbol{x}}_i)'\right)\boldsymbol{a}} \qquad \textbf{(2)}$$

biçimindedir. Burada n_i: her bir gruptaki gözlem sayısını, x_{ij}: gözlem değerlerini, $\bar{\boldsymbol{x}}_i$: p×1 boyutlu her bir gruba ilişkin ortalama vektörünü, $\bar{\boldsymbol{x}}$: p×1 boyutlu genel ortalama vektörünü, **B:** p×p boyutlu gruplar arası kareler ve çarpımlar toplamı matrisini, **W:** p×p boyutlu grup içi kareler ve çarpımlar toplamı matrisini ve ***a*:** p×1 boyutlu katsayılar vektörünü göstermektedir. Eş.2'deki varyans oranlarının en büyüklenmesi için *a*'ya göre türev alınıp gerekli düzenlemeler yapılır. Bu işlemler sonucunda $\left|\mathbf{W}^{-1}\mathbf{B} - \lambda \mathbf{I}\right| = 0$ determinantının çözümünden $\lambda_j \ (j = 1,\ldots,r)$ özdeğerleri ve bu özdeğerlere karşılık gelen özvektörler elde edilir. $\boldsymbol{a}_j$ ile gösterilen bu özvektörler, $y_j = \boldsymbol{a}_j \boldsymbol{x}$ biçimindeki *j*'inci diskriminant fonksiyonunun katsayılarıdır.

$\boldsymbol{x}_0' = (x_{01} \ \ \ldots \ \ x_{0p})$ biçiminde gösterilen yeni bir gözlem vektörünün hangi gruba atanacağına karar verebilmek için

$$\sum_{j=1}^{r}(y_j - \bar{y}_{ij})^2 = \sum_{j=1}^{r}\left[\boldsymbol{a}_j'(x_{ij} - \bar{\boldsymbol{x}}_i)\right]^2$$
$$\bar{y}_{ij} = \boldsymbol{a}_j'\bar{\boldsymbol{x}}_i \quad , \quad i = 1,...,k \qquad \textbf{(3)}$$

eşitliğinden yararlanılabilir. Eş.(3)'teki kareler toplamı en küçük olan gruba atama yapılır (Johnson ve Wichern, 2002).

Ayrıca $\boldsymbol{x}_0' = (x_{01} \;\; ... \;\; x_{0p})$ biçimindeki yeni bir gözlem vektörünün hangi gruba atanacağına karar verebilmek için doğrusal diskriminant fonksiyonunun skor değerlerinin elde edilmesinde kullanılan eşitliğinden de yararlanılabilir.

$$d_i(x) = \bar{\boldsymbol{x}}_i' \, \boldsymbol{S}^{-1}\boldsymbol{x} - \frac{1}{2}\bar{\boldsymbol{x}}_i' \, \boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_i + \ln p_i \quad i = 1,...,k \qquad \textbf{(4)}$$

Burada $\mathbf{S} = \mathbf{W}/(n_1 + ... + n_k - k)$, p×p boyutlu ortak varyans-kovaryans matrisini ve $p_i = n_i / \sum_{i=1}^{k} n_i$ önsel olasılıkları göstermektedir.

Eğer $p_1 = p_2 = ... = p_k = 1/k$ ise Eş.(3) ve Eş.(4) eşdeğerdir. Yeni gözlem vektörü, $d_i(x)$ değeri en büyük olan gruba atanır (Özdamar, 1999; Johson ve Wichern, 2002).

İkiden çok grup olması durumunda elde edilen diskriminant fonksiyonlarının önem kontrolleri, Wilk's Λ ya da λ_j özdeğerlerine dayanmaktadır. Burada her bir λ_j için Eş.(5)'te verilen ki-kare değeri hesaplanmakta ve $(p + k - 2j)$ serbestlik dereceli ki-kare tablo değeri ile karşılaştırılmaktadır.

$$\chi_j^2 = \left[n - 1 - \frac{1}{2}(p + k)\right]\ln(1 + \lambda_j) \qquad \textbf{(5)}$$

$\chi_j^2 > \chi^2_{(p+k-2j\,;\,\alpha)}$ ise *j*'inci diskriminant fonksiyonunun grupları ayırmada önemli olduğu belirtilir. Normallik ve ortak varyans-kovaryans varsayımlarının sağlanması durumunda kullanılan bu test, diskriminant fonksiyonunun önemsizliği kabul edilinceye kadar sürdürülür. Ayrıca DA sonuçlarının geçerliliğinin bir göstergesi de doğru sınıflandırma oranının yüksek olmasıdır. Doğru sınıflandırma oranı $(\tilde{n}/n)$, örneklem büyüklüğü ve grup sayısına bağlı olacağından ikisinin de birlikte düşünüldüğü ayırma gücü testi yapılmaktadır. "DA'nin ayırma gücü önemsizdir" biçiminde ifade edilen yokluk hipotezinin testi için

$$Q = \frac{(n - \tilde{n}k)^2}{n(k-1)} \sim \chi^2_{(1\,;\,\alpha)} \;, \; \tilde{n} = \sum_{i=1}^{k} n_{ii} \qquad \textbf{(6)}$$

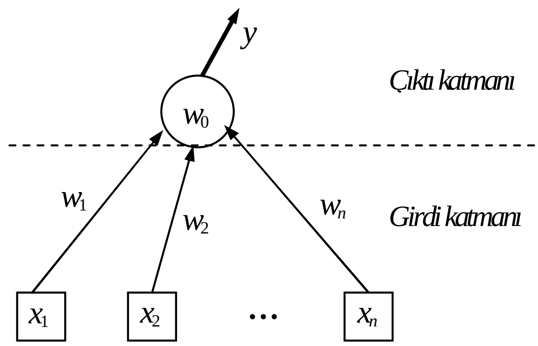
test istatistiği kullanılır (Tatlıdil, 1996).

2.2. Yapay Sinir Ağları

Yapay sinir ağlarının kökleri matematik, istatistik, fizik, bilgisayar bilimleri ve mühendislik gibi değişik bilim alanlarına dayanmaktadır. Sadece eldeki verilerin kullanılması nedeniyle modelleme, zaman serileri analizi, örüntü (pattern) algılama, sinyal işleme gibi çok değişik alanlarda uygulanabilmektedir.

İnsan beyni bugünkü dijital bilgisayarlardan farklı bir şekilde çalışmaktadır. İnsan beyni çok karmaşık ve doğrusal olmayan paralel bir bilgisayar gibidir. Bu paralellik sayesinde bir çok işlemi bugünkü en hızlı bilgisayarlardan bile daha hızlı yapabilmektedir. Bunu fark eden araştırmacılar insan beynindeki sinir hücrelerini taklit eden modeller üzerinde çalışmaya başlamışlardır. Bu konudaki ilk çalışmalarda adı en çok öne çıkan araştırmacılar arasında McCulloch ve Pitts (1943), Hebb (1949) ve Rosenblatt (1958) yer almaktadır.

Rosenblatt (1958) tarafından geliştirilen algılayıcı (perceptron), doğrusal olarak ayrılabilen örüntüleri sınıflamak için kullanılan yapay sinir ağlarının en basit halidir. Bu model tek bir nöron ve uyarlanabilir ağırlıklardan oluşmaktadır (Şekil-1). Bu tür yapay sinir ağlarında hücreler katmanlara yerleştirilir ve katman sayısını girdileri içeren katman dışındaki katmanlar belirlemektedir. Girdi ve çıktı katmanı haricinde katmanlar varsa bunlara gizli katman denir. Burada girdi katmanı hariç sadece çıktı katmanı bulunduğu için algılayıcı tek katmanlı bir yapay sinir ağıdır.



Şekil 1. Algılayıcı

Şekil 1'de $w_0, w_1, w_2, ..., w_n$ ağırlıkları ve $x_1, x_2, ..., x_n$ de girdileri göstermektedir. Bu modelde çıktı

$$y = \varphi\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)$$

olup φ doğrusal olmayan bir fonksiyondur. Örneğin bazı uygulamalarda φ fonksiyonu

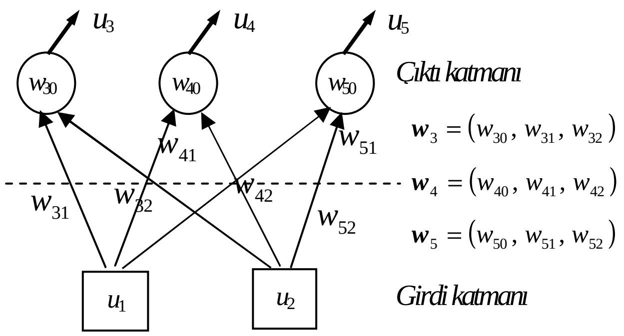
$$\varphi(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

şeklinde alınmaktadır.

Ağırlıklara farklı değerler verildiğinde girdilere göre farklı çıktılar elde edilir. Bir algoritma yardımı ile probleme göre istenen çıktıları verebilecek ağırlıklar bulunmaya çalışılır. Örneğin sınıflandırma problemlerinde aynı gruba ait girdiler için aynı çıktı elde edilmelidir. Ağırlıklar için en az hatalı sınıflandırmayı yapacak değerler bulunarak algılayıcı *x* girdi vektörüne göre sınıflamayı öğrenebilir. Öğrenme için çeşitli algoritmalar geliştirilmiştir. Bunlardan bir tanesi de algılayıcı için geliştirilen Öğrenme Algoritmasıdır. Eğer sonlu sayıda veriden oluşan ayrılabilir (bir doğru veya düzlem ile) bir veri seti için doğru sınıflamayı sağlayan ağırlıklar varsa, söz konusu algoritma bu ağırlıkları bulmaktadır. Bu sonuç Perceptron Yakınsama Teoremi olarak bilinir (Gallant 1994).

2.2.1. Doğrusal Makineler

Bu çalışmada kullanılan model doğrusal makineler (winner-take-all groups / linear machines) olarak da bilinmektedir. Söz konusu modelde tek katman vardır. Çıktı hücrelerinden sadece bir tanesinin +1 değerli ve diğerlerinin hep -1 değerli olması kısıtlaması vardır, yani her zaman tek bir hücre kazanmaktadır.



Şekil 2. Doğrusal Makine (p=2 ve c=3)

Bu modelde genel olarak birden çok (*c* adet) çıktı hücresi vardır. Bunlara ait ağırlık vektörleri sırasıyla w_{p+1}, w_{p+2}, …, w_{p+c} ile gösterilecektir. Herhangi bir çıktı hücresine ait u_i sonucu aşağıdaki gibi hesaplanmaktadır.

$$S_i = \sum_{j \geq 0} w_{ij} u_j \ , \ i = p+1, \ldots, p+c \text{ olmak üzere}$$

$$u_i = \begin{cases} +1, & \text{her } k \neq i \text{ için } S_i > S_k \\ -1, & \text{bazı } k \neq i \text{ için } S_i < S_k \ . \\ 0, & \text{diğer durumlarda} \end{cases}$$

Yapay sinir ağının öğrenme aşaması için kullanılacak veri örnekleri ve doğru çıktılar $\{E^k, C^k\}$ ile gösterilecektir. Burada $\boldsymbol{E^k} = (E_0^k, E_1^k, \ldots, E_p^k)$, p+1 boyutlu bir vektör olup girdileri göstermek için kullanılacaktır. İlk bileşen $E_0^k \equiv 1$ ve i=1,...,p olmak üzere $E_i^k \in \{+1, -1, 0\}$. $\boldsymbol{C^k} = (C_{p+1}^k, \ldots, C_{p+c}^k)$ sadece bir tanesi +1 ve geri kalan bileşenleri -1 olan *c* boyutlu bir vektör olup doğru sınıfı göstermek için kullanılacaktır.

Doğrusal makinelere de Perceptron Öğrenme Algoritması uygulanabilir (Gallant, 1994). Bu algoritmayı açıklamadan önce bazı tanımlamalar yapılacaktır. *c* adet çıktı hücresine ait ağırlık vektörleri sırasıyla $w_{p+1}, w_{p+2}, \ldots, w_{p+c}$ olsun. Öğrenme esnasında geçici ağırlık vektörlerine ihtiyaç olacaktır; bunlar da $\pi_{p+1}, \pi_{p+2}, \ldots, \pi_{p+c}$ ile gösterilecektir. İlk aşamada geçici ağırlık vektörleri $\boldsymbol{\pi} = (\pi_{p+1}, \pi_{p+2}, \ldots, \pi_{p+c})$ hep 0 vektörü olarak seçilir. Daha sonraki aşamalarda veri örnekleri içerisinde doğru sınıflandırma sayısı kullanılarak, gerekirse o ana kadar bulunan en iyi ağırlıklar olan $\boldsymbol{w} = (w_{p+1}, w_{p+2}, \ldots, w_{p+c})$ ile değiştirme yapılır. Bu şekilde belli bir iterasyon sayısına kadar en çok doğru sınıflandırma sayısına sahip ağırlıklar bulunur. Ağırlık vektörlerinin performansını değerlendirmek için *run* ve *num_ok* değişkenleri kullanılmaktadır. *Run* ilgili ağırlık vektörlerinin ard arda doğru sınıflandırdığı örnek veri sayısı iken *num_ok* de tüm örnek verileri içerisindeki doğru sınıflandırılmış örnek sayısıdır. Öğrenme için kullanılan algoritmanın adımları aşağıda özetlenmiştir:

1. i =1, p+1, …, p+c için $\pi_i = (0,0,\ldots,0)$; run_π = run_w = num_ok_π = num_ok_w = 0.

2. Veri örnekleri içerisinden rasgele bir tane seç; (E^k, C^k).

3. Eğer π, E^k yı doğru sınıflandırıyorsa (yani $i \neq j$ için $\pi_i \cdot E^k > \pi_j \cdot E^k$)

3.1. $run_\pi = run_\pi + 1$.

3.2. Eğer

3.2.-A) $run_\pi > run_w$ ise

3.2.1. Örnek verilerin tamamını inceleyerek num_ok_π değerini hesapla.

3.2.2. Eğer $num\_ok_\pi > num\_ok_w$ ise

RATCHET: İyileşme.
R1) $w = \pi$.

R2) $run_w = run_\pi$.

R3) $num\_ok_w = num\_ok_\pi$.

R4) Eğer veri örneklerinin tamamı doğru sınıflandırıldıysa dur (örnekler ayrılabilir)
3.2.-B) aksi takdirde ($run_\pi \leq run_w$)

CHANGE: Yeni ağırlık vektörleri oluştur.

C1) $j \neq i$ için $\pi_i \cdot E^k \leq \pi_j \cdot E^k$ olacak şekilde bir çıktı hücresi seçerek *i* ve *j* hücrelerinin ağırlıklarını aşağıdaki gibi değiştir:

$$\pi_i = \pi_i + E^k \quad \pi_j = \pi_j - E^k.$$

C2) $run_\pi = 0$.

4. İterasyon sonu. İterasyon sınırına ulaşılmadıysa ikinci adıma dön.

3. UYGULAMA

Tek katmanlı yapay sinir ağları ile doğrusal diskriminant fonksiyonu sonuçlarının karşılaştırıldığı çalışmada iki farklı veri kümesi kullanılmıştır. Her iki veri kümesi de üç grup (k=3) ve iki değişkenden (p=2) oluşmaktadır. Her bir grupta 15'er gözlem (n_i=15, i=1, 2, 3) olmak üzere toplam 45 (n) gözlem ile çalışılmıştır. I. veri kümesi Ek-1 ve II. veri kümesi Ek-2 de verilmiştir. II. veri kümesi, tek katmanlı yapay sinir ağları ile doğrusal diskriminant fonksiyonu sonuçlarının karşılaştırılmasını pekiştirmek amacıyla I. veri kümesinden öteleme yoluyla elde edilmiştir. Bu amaçla I. veri kümesindeki x_2 değişkeninde sırasıyla 1.grup için (-3), 2.grup için (+3) ve 3.grup için (-3) öteleme yapılmıştır.

Fisher doğrusal diskriminant fonksiyonunun kullanıldığı çalışmada, normallik ve ortak varyans-kovaryans (Box's M=11.079, p=0.112, α=0.05) varsayımları sağlanmıştır. I. ve II. veri kümesi için r=2 olduğundan ayırma amacıyla kullanılabilecek doğrusal diskriminant fonksiyonları

I. Veri Kümesi

$$y_1 = 0.543x_1 + 0.865x_2 \rightarrow \chi_1^2 = 131.955^*$$
$$y_2 = 0.868x_1 - 0.787x_2 \rightarrow \chi_2^2 = 97.854^+$$

II. Veri Kümesi

$$y_1 = 1.011x_1 - 0.027x_2 \rightarrow \chi_1^2 = 116.783^*$$
$$y_2 = -0.159x_1 + 1.169x_2 \rightarrow \chi_2^2 = 21.574^+$$

$^*\ \chi^2_{3;\,0.05} = 7.815$, $^+\ \chi^2_{1;\,0.05} = 3.841$

olarak elde edilmiştir. Bu diskriminant fonksiyonlarının önem kontrollerini yapmak amacıyla Eş.(5)'ten yararlanılmıştır. Buna göre her iki veri kümesi için de iki diskriminant fonksiyonu da önemlidir. I. ve II. veri kümesi için doğru sınıflandırma oranları ve DA'nin ayırma gücünün önemliliğine ilişkin Eş.(6)'dan hesaplanan Q test istatistiği sonuçları Tablo-1'de özetlenmiştir. Bu sonuçlara göre her iki veri kümesi için de DA ayırma gücü sonuçlarının tatmin edici olduğu ifade edilir.

Tablo 1-Doğru Sınıflandırma Oranları ve Q Test İstatistiği Sonuçları

	$\tilde{n}/n$	Q^*
I. veri kümesi	%100	722.5
II. veri kümesi	%89	547.6

$^*\ \chi^2_{1;\,0.05} = 3.841$

Eş.(4) ile verilen doğrusal diskriminant skor değerlerinin hesaplanmasında kullanılan fonksiyonlar, eşit önsel olasılıklar ile (p_i=0.333, i=1, 2, 3) her iki veri kümesi için aşağıdaki gibi bulunmuştur.

I. Veri Kümesi

$$d_1(x) = -16.595 + 3.446x_1 + 4.398x_2$$
$$d_2(x) = -17.620 - 3.410x_1 - 4.694x_2$$
$$d_3(x) = -20.953 - 4.827x_1 + 5.793x_2$$

II. Veri Kümesi

$$d_1(x) = -9.555 + 4.087x_1 + 0.295x_2$$
$$d_2(x) = -9.693 - 4.051x_1 - 0.590x_2$$
$$d_3(x) = -9.730 - 4.186x_1 + 1.689x_2$$

Yeni bir gözlem vektörünün hangi gruba atanacağına bu fonksiyonlar yardımı ile karar verilebileceği gibi Eş.(3) ile verilen kareler toplamından da yararlanılabilir. Buna göre $\mathbf{x}_0' = (x_{01}', x_{02}')$ olarak gösterilen üç farklı yeni gözlem vektörü için her iki veri kümesine ilişkin Eş.(3) ve Eş.(4)'ten elde edilen sonuçlar Tablo 2 ve Tablo 3'te verilmiştir:

Aynı veri kümeleri için kullanılan doğrusal makine Şekil-2' deki gibidir. Üç farklı grup olduğu için k=3 ve girdi vektörleri iki boyutlu olduğu için de p=2 alınmıştır.

Kullanılan algoritma deterministik olmadığı için elde edilen sonuçlar değişebilmektedir. Bu nedenle C++ dilinde yazılan program her bir sınıflama için 10 defa çalıştırılmıştır. Tablo 2 ve Tablo 3'te verilen üç farklı yeni gözlem vektörü için bu algoritmanın kullanıldığı sınıflandırma sonuçları Tablo 4'te verilmiştir.

Tablo 2. I. Veri Kümesi İçin Yeni Gözlem Vektörlerinin Sınıflandığı Gruplar

x'_{01}	x'_{02}	$\sum_{j=1}^{2}(y_j-\bar{y}_{1j})^2$	$\sum_{j=1}^{2}(y_j-\bar{y}_{2j})^2$	$\sum_{j=1}^{2}(y_j-\bar{y}_{3j})^2$	$d_1(x)$	$d_2(x)$	$d_3(x)$	Sınıflanan Grup
4.7583	1.432	9.202$^+$	102.641	92.735	6.100*	-40.568	-35.626	1
-4.020	-2.817	106.512	2.114$^+$	56.467	-42.837	9.313*	-17.867	2
-1.103	3.562	27.604	79.270	8.109$^+$	-4.730	-30.579	5.006*	3

$^+$ Her bir gözlem vektörü en küçük kareler toplamını veren gruba atanır.
* Her bir gözlem vektörü en büyük değeri veren gruba atanır.

Tablo 3. II. Veri Kümesi İçin Yeni Gözlem Vektörlerinin Sınıflandığı Gruplar

x''_{01}	x''_{02}	$\sum_{j=1}^{2}(y_j-\bar{y}_{1j})^2$	$\sum_{j=1}^{2}(y_j-\bar{y}_{2j})^2$	$\sum_{j=1}^{2}(y_j-\bar{y}_{3j})^2$	$d_1(x)$	$d_2(x)$	$d_3(x)$	Sınıflanan Grup
4.7583	-1.568	9.200$^+$	84.166	92.663	9.430*	-28.044	-32.297	1
-4.020	0.183	66.971	2.115	0.269$^+$	-25.931	6.484	7.407*	3
-1.103	0.562	27.577	11.053	8.105$^+$	-13.897	-5.556	-4.164*	3

$^+$ Her bir gözlem vektörü en küçük kareler toplamını veren gruba atanır.
* Her bir gözlem vektörü en büyük değeri veren gruba atanır.

Tablo 4. Doğrusal Makineye Ait Sonuçlar

I. Veri Kümesi	$\tilde{n}/n$ %	**Sınıflanan Grup**			**İterasyon**
Deneme 1	100.00	1	2	3	25
Deneme 2	100.00	1	2	3	109
Deneme 3	100.00	1	2	3	154
Deneme 4	100.00	1	2	3	18
Deneme 5	100.00	1	2	3	5
Deneme 6	100.00	1	3	3	61
Deneme 7	100.00	1	2	3	13
Deneme 8	100.00	1	2	3	31
Deneme 9	100.00	1	2	1	13
Deneme 10	66.67	1	3	3	10000
II. Veri Kümesi	$\tilde{n}/n$ %	**Sınıflanan Grup**			**İterasyon**
Deneme 1	66.67	1	3	3	10000
Deneme 2	75.56	3	3	3	10000
Deneme 3	73.33	2	1	3	10000
Deneme 4	75.56	3	2	3	10000
Deneme 5	66.67	1	3	3	10000
Deneme 6	64.44	3	2	3	10000
Deneme 7	68.89	3	3	3	10000
Deneme 8	75.56	3	3	3	10000
Deneme 9	66.67	1	3	3	10000
Deneme 10	66.67	3	3	3	10000

4. SONUÇ VE TARTIŞMA

Tablo 4'te görüldüğü gibi I. veri kümesi için doğrusal makineler ile elde edilen sonuçlar çoğunlukla Fisher'in doğrusal diskriminant fonksiyonu ile elde edilen sonuçlara çok yakındır. Deneme 6, 9 ve 10 da daha düşük bir performans tespit edilmiştir. Örneğin bunlardan Deneme 10 da, verilerin sınıflandırılmasında ancak %66.67 ($\tilde{n}/n$) lik bir doğru sınıflandırma oranı elde edilmiştir. Ayrıca söz konusu bu üç denemede üç yeni gözlemden bir tanesi yanlış sınıflandırılmıştır. Burada dikkati çeken bir nokta son deneme hariç %100 doğru sınıflandırma oranının elde edilmesi için gerekli olan iterasyon sayısının genelde küçük olmasıdır. Diğer yandan Fisher'in doğrusal diskriminant fonksiyonu ile elde edilen sonuçlara bakıldığında doğru sınıflandırma oranı I. veri kümesi için %100 olarak bulunurken (Tablo 1) yeni gözlemlerin tamamı doğru sınıflara atanmıştır (Tablo 2).

II. veri kümesine bakıldığında ise doğrusal makineler ile elde edilen sonuçların Fisher'in doğrusal diskriminant fonksiyonuna göre o kadar iyi olmadığı görülmektedir. Yapılan denemede verilerin sınıflandırılmasında doğrusal makineler için doğru sınıflandırma oranı ($\tilde{n}/n$) %64.44 ile %75.56 arasında değişirken (Tablo 4) bu oran Fisher'in doğrusal diskriminant fonksiyonu için %89 (Tablo 1) olarak belirlenmiştir. Fisher'in doğrusal diskriminant fonksiyonu için üç yeni gözlemden sadece ikinci gözlem yanlış sınıfa atanırken (Tablo 3) doğrusal makinelerde genelde iki gözlem yanlış gruba atanmıştır (Tablo 4).

Özellikle II. veri kümesindeki bu performans düşüklüğünün başlıca nedeni bu tür yapay sinir ağı modellerinin doğrusal bir fonksiyon (genel olarak düzlem) ile ayrılabilen veri kümeleri için daha uygun olmasından kaynaklanmaktadır. Genelde verilerin doğrusal olarak ayrılıp ayrılamayacağını baştan tespit etmek zor olduğu için bu tür modeller sınıflandırmada minimum hata oranı elde etmek amacıyla uygulanmaktadırlar. Her şeye rağmen, bu algoritmada sadece doğru sınıflandırılan gözlemlerin sayısı kullanıldığı için sonuçların kabul edilebilir düzeyde olduğu söylenebilir.

Burada vurgulanması gereken önemli bir nokta doğrusal makinelerde verilerin dağılımı ile ilgili herhangi bir varsayımın olmamasıdır. Diğer yandan Fisher'in doğrusal diskriminant fonksiyonu için normallik ve ortak varyans-kovaryans varsayımlarının sağlanması gerekir. Verilere ait bu ek bilgi sayesinde Fisher'in doğrusal diskriminant fonksiyonu ile sonuçların daha iyi olduğu görülmektedir.

İki veri kümesinin kullanıldığı bu çalışmada, genel bir değerlendirmenin yapılması mümkün olmasa da elde edilen sonuçlar her iki yöntemin de güçlü ve zayıf yönleri hakkında bir fikir vermektedir. Eğer varsayımların sağlanması konusunda herhangi bir sorun yoksa Fisher'in doğrusal diskriminant fonksiyonunun tercih edilmesi tavsiye edilebilir. Diğer yandan varsayımların sağlanmasında problemler varsa doğrusal makineler gibi alternatif yöntemlerin araştırılması faydalı olacaktır.

Doğrusal makineler için yazılan program çalıştırıldığında II. veri kümesi için %80 lik doğru sınıflandırma oranına ulaşılabileceği görülmüştür. Sonuç olarak normallik varsayımının sağlanmadığı durumlarda çok karmaşık olmayan bu tür modellerin kullanılabileceği düşünülmektedir. Kullanılan algoritmada bazı değişiklikler yapılarak performansın arttırılması da mümkün olabilir. Örneğin çıktı hücreleri bağımsız olarak eğitilerek performansta bir iyileşme olup olmayacağı incelenebilir.

Bu çalışmada özellikle tek katmanlı modeller seçilmiştir. Bunun başlıca nedeni bu modellerin daha basit olmasıdır. Özellikle algoritmaya ilişkin programın kodlanması düşünüldüğünde bu önemli bir etkendir. Katman sayısı arttırılırsa daha etkili yapay sinir ağı modelleri kullanılabilecektir. Bu doğrultuda özellikle "Support Vector Machines" adlı modeller ileride uygulanması düşünülen yapay sinir ağları arasında yer almaktadır (Boser vd., 1992; Haykin, 1999).

KAYNAKÇA

Boser, B., Guyon, I. ve Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. *Fifth Annual Workshop on Computational Learning Theory*, ss. 144-152. Morgan Kaufmann, San Mateo. CA.

Gallant, S. I. (1994). *Neural Network Learning and Expert Systems*. The MIT Press. Massachusetts.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey.

Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York.

Johnson, R. A. ve Wichern, D. W. (2002). *Applied Multivariate Statistical Analyses*. Prentice Hall, New Jersey.

McCulloch, W.S. ve Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 114-133.

Özdamar, K. (1999). *Paket Programlar İle İstatistiksel Veri Analizi II*. Kaan Kitabevi, Eskişehir.

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65, 386-408.

Schalkoff, R. J. (1992). *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons.

Tatlıdil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Cem Web Ofset Ltd. Şti, Ankara.

EK-1 : I.VERİ KÜMESİ

Grup 1		Grup 2		Grup 3	
x_1	x_2	x_1	x_2	x_1	x_2
3.406	4.439	-4.256	-3.556	-4.306	4.967
3.811	4.893	-2.921	-2.543	-4.581	3.527
4.395	3.351	-2.962	-3.071	-3.606	3.334
5.340	2.770	-4.886	-5.544	-3.640	4.041
4.238	4.093	-2.452	-3.472	-2.899	4.127
4.249	3.562	-4.154	-3.769	-5.044	3.889
2.917	4.121	-5.018	-3.542	-3.501	4.047
4.992	4.851	-1.488	-3.488	-2.967	2.224
3.556	3.414	-4.680	-5.309	-3.466	4.475
4.437	3.178	-4.493	-5.967	-3.131	2.235
4.034	4.173	-4.162	-3.635	-4.114	2.561
4.326	5.138	-6.386	-3.763	-5.422	4.341
4.797	3.955	-3.136	-3.849	-4.450	4.243
1.962	2.517	-4.632	-4.351	-3.310	3.130
4.698	3.328	-5.684	-5.190	-3.607	3.318

Kaynak: Schalkoff (1992), s. 84

EK-2 : II.VERİ KÜMESİ

Grup 1		Grup 2		Grup 3	
x_1	x_2	x_1	x_2	x_1	x_2
3.406	1.439	-4.256	-0.556	-4.306	1.967
3.811	1.893	-2.921	0.457	-4.581	0.527
4.395	0.351	-2.962	-0.710	-3.606	0.334
5.340	-0.230	-4.886	-2.544	-3.640	1.041
4.238	1.093	-2.452	-0.472	-2.899	1.127
4.249	0.562	-4.154	-0.769	-5.044	0.889
2.917	1.121	-5.018	-0.542	-3.501	1.047
4.992	1.851	-1.488	-0.488	-2.967	-0.776
3.556	0.414	-4.680	-2.309	-3.466	1.475
4.437	0.178	-4.493	-2.967	-3.131	-0.765
4.034	1.173	-4.162	-0.635	-4.114	-0.439
4.326	2.138	-6.386	-0.763	-5.422	1.341
4.797	0.955	-3.136	-0.849	-4.450	1.243
1.962	-0.483	-4.632	-1.351	-3.310	0.130
4.698	.0328	-5.684	-2.190	-3.607	0.318

Güvenç Arslan, 1969 yılında Çorum'da doğdu. Lisans eğitimini İstanbul Teknik Üniversitesi, Matematik Mühendisliği Bölümü'nde, Yüksek Lisans eğitimini Akdeniz Üniversitesi, Matematik Bölümü'nde, Doktora eğitimini Anadolu Üniversitesi, İstatistik Bölümü'nde tamamladı. Halen Başkent Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü'nde Yrd. Doç. olarak görev yapmaktadır. Evli ve bir çocuk sahibidir. İlgi alanları; olasılık dağılımlarının karakterizasyonu, uygulamalı matematik, yöneylem araştırması, yapay zeka, programlama dilleridir.

İlknur Özmen, 1967 yılında Kütahya'da doğdu. Lisans, Yüksek Lisans ve Doktora eğitimini Hacettepe Üniversitesi, İstatistik Bölümü'nde tamamladı. Halen Başkent Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü'nde Yrd. Doç. olarak görev yapmaktadır. Evli ve bir çocuk sahibidir. İlgi alanları; çok değişkenli istatistiksel yöntemler, parametrik olmayan istatistiksel yöntemler, matematiksel istatistik, uygulamalı istatistik, Poisson regresyon analizidir.