S Ü Fen Ed Fak Fen Derg Sayı 22 (2003) 89- 95, KONYA

Diskriminant Analizinde Gerçek Hata Oranına İlişkin Güven Aralığı için Bir Simülasyon Çalışması

Cemal ATAKAN[1]

Özet: Bu çalışmada, gerçek hata oranına ilişkin M tahmin edicisine dayalı mevcut olan analitik güven aralığı ifadelerinden elde edilen sonuçlar, bootstrap tekniği ile elde edilen sonuçlarla karşılaştırılmaktadır. Karşılaştırma sonucunda bootstrap tekniğinin daha dar aralıklar verdiği görülmektedir.

Anahtar kelimeler: Diskriminant analizi, hata oranı, güven aralığı, bootstrap.

A Simulation Study for the Confidence Interval Based on the Actual Error Rate in Discriminant Analysis

Abstract: In this study, the existing analitical confidence intervals based on the actual error rate estimator M compared with the bootstrap by simulation. It has been observed that bootstrap results give norrower intervals.

Keywords : Discriminant analysis, error rate, confidence interval, bootstrap.

Giriş

Diskriminant analizi, üzerinde ölçüm yapılan bir bireyi sonlu sayıda bilinen farklı kitleden birine atanmasını gerçekleştiren istatistiksel bir tekniktir. Her bir kitle ölçümlere karşılık gelen $\underline{X}$ rasgele vektörünün olasılık dağılımı ile karakterize edilir. Olasılık dağılımları tamamen bilindiğinde, problem atama(sınıflandırma) kuralını belirlemektir. Dağılımlar biçimsel bilindiğinde, problem kitlelerden alınan örneklemler yardımıyla parametre tahmini ve atama kuralını oluşturmaktır. Diskriminant analizinde amaç, atama işlemini mümkün olan minumum hatayla yapmaktır. Bu optimizasyon kriterine göre elde edilen diskriminant fonksiyonlarının değerlendirilmesinde hatalı sınıflandırma olasılıklarının (hata oranlarının) bilinmesi önemlidir. Hata oranları; dağılımların ve parametrelerin bilinip bilinmemesine, örneklem hacmi ve atama kuralı gibi değişik kriterlere bağlı olarak hesaplanır.

Π_1 ve Π_2 birbirinden farklı iki kitle olmak üzere, $\underline{X} = (X_1, X_2, ..., X_p)'$ birey üzerinde ölçümlere karşılık gelen p- boyutlu rasgele vektörü Π_i kitlesinden ise $\underline{X}$'in ortak olasılık yoğunluk fonksiyonu $f_i(\underline{x}, \underline{\theta}_i)$ biçiminde gösterilir. Burada, $\underline{\theta}_i$ parametre vektörü ve $\underline{x} \in \Re^p$ dir.

[1] . Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Tandoğan, ANKARA. atakan@science.ankara.edu.tr

$\underline{X}$ 'in aldığı değerler p- boyutlu $\Re^p$ örneklem uzayında olmak üzere, sınıflandırma işlemi bu uzayı $B_1 \cup B_2 = \Re^p$ ve $B_1 \cap B_2 = \varnothing$ olan B_1 ve B_2 bölgelerine ayırır. Eğer $\underline{X}$'in gözlem değeri B_1 bölgesinde ise bu gözlemin yapıldığı gözlem birimi Π_1, aksi halde Π_2 kitlesine atanır.

Atama Kuralının Oluşturulması

Bu çalışmada Welch tarafından önerilen toplam hatalı sınıflandırma olasılıklarını minimize eden olabilirlik oran kriterine göre elde edilen lineer diskriminant fonksiyonu göz önüne alınmıştır[1]. Π_1 ve Π_2, ortalama vektörleri $\underline{\mu}_1$, $\underline{\mu}_2$ ve aynı Σ varyans-kovaryans matrisine sahip çok değişkenli normal dağılımlı kitleler olduğu bilindiğinde, toplam hatalı sınıflandırma olasılığını minimize eden optimal sınıflandırma kuralı,

$$\xi : \begin{cases} U(\underline{x}) > k \text{ ise } \underline{x} & \Pi_1\text{'e} \\ \text{aksi halde} & \Pi_2\text{'ye atanir.} \end{cases}$$

ile verilir. Burada,

$$U(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2) \right]^{/} \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

iki kitleden birine ait olduğu ancak hangisine ait olduğu bilinmeyen rasgele gözlemin değeri $\underline{x}$ için parametrelerin bilindiği durumda $f_1(\underline{x};(\underline{\mu}_1,\Sigma))$ ve $f_2(\underline{x};(\underline{\mu}_2,\Sigma))$ ortak olasılık yoğunluk fonksiyonlarının birbirine oranlanmasıyla elde edilen kitle lineer diskriminant fonksiyonu, $k = \ln(\frac{q_2}{q_1})$ dır. q_i birimin *i* inci kitleye ait olması olasılığı yani önsel olasılıktır $(i = 1{,}2)$.

Parametreler bilinmediğinde Π_1 ve Π_2 kitlelerinden alınan n_1 ve n_2 hacimli rasgele örneklemlerden hesaplanan tahmin edicileri $\overline{\underline{X}}_1$ ve $\overline{\underline{X}}_2$ örneklem ortalama vektörleri ve S birleştirilmiş örneklem varyans-kovaryans matrisine bağlı optimal sınıflandırma kuralı

$$\hat{\xi} : \begin{cases} W(\underline{x}) > k \text{ ise } \underline{x} & \Pi_1\text{'e} \\ \text{aksi halde} & \Pi_2\text{'ye atanir.} \end{cases}$$

ile verilir. Burada,

$$W(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\overline{\underline{x}}_1 + \overline{\underline{x}}_2) \right]^{/} S^{-1} (\overline{\underline{x}}_1 - \overline{\underline{x}}_2)$$

parametrelerin bilinmediği durumda, kitle lineer diskriminant fonksiyonunun elde edilmesine benzer biçimde elde edilen örneklem lineer diskriminant fonksiyonudur[2,3]. Bu çalışmada önsel olasılıklar eşit alındığından $k = 0$ olacaktır.

Hata Oranları

Atama işlemi yapılırken, bilinen kitlelerin hangisinden geldiği bilenmeyen bir birim ait olduğu kitleye değil de diğer kitleye atanırsa, hata yapılmış olur. Diskriminant analizinde amaç, atama işlemini minumum hatayla yapmaktır. Bu optimizasyon ölçütüne göre elde edilen diskriminant fonksiyonlarının değerlendirilmesinde hata oranlarının ya da hatalı sınıflandırma olasılıklarının bilinmesi önemlidir. Önsel olasılıklar ile ağırlıklandırılmış hata oranına , toplam hata oranı veya toplam hatalı sınıflandırma olasılığı denilmektedir.

Hata oranı genelde diskriminant fonksiyonunun dağılımına bağlı olarak bulunur. Ancak hata oranları dağılımdan bağımsız da hesaplanabilir[4]. Yukarıda verilen atama kurallarına ilişkin optimal, gerçek (koşullu) ve beklenen gerçek (koşulsuz) hata oranı gibi hata oranları tanımlanır. Optimal hata oranı, parametreler bilindiğinde elde edilen diskriminant fonksiyonu göz önüne alınarak bulunan hata oranı, gerçek hata oranı, parametreler bilinmediğinde örneklemlerden elde edilen tahminlere bağlı örneklem diskriminant fonksiyonuna göre bulunan hata oranı ve beklenen gerçek hata oranı, olası tüm örneklemler üzerinden gerçek hata oranının beklenen değeridir[2,3,5,6,7].

Çok değişkenli normal kitleler göz önüne alındığında, $\underline{X} \sim N(\underline{\mu}_i, \Sigma)$ ise $U(\underline{X})$'in dağılımı $(-1)^i(-\frac{\Delta}{2})$ ortalamalı ve Δ^2 varyanslı tek değişkenli normaldir. Burada,

$$\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)^{/} \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

iki kitle arasındaki Mahalanobis uzaklığıdır.

Parametrelerin bilinmediği durumda elde edilen $W(\underline{X})$ örneklem lineer diskriminant fonksiyonunun dağılımı kolaylıkla elde edilememektedir, ancak bazı şartlar altında elde edilebilmektedir[2,8]. ($W(\underline{X})$ ifadesindeki tahmin ediciler yerine örneklemden elde edilen tahmin değerleri alındığında, $W(\underline{X})$'in bu tahmin değerlerine koşullandırılmış koşullu dağılımı elde edilebilir. Böylece $\underline{X} \sim N(\underline{\mu}_i, \Sigma)$ ise $W(\underline{X})$ 'in koşullu dağılımı, ortalaması $\left[\underline{\mu}_i - \frac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)\right]^{/} s^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$ ve varyansı $(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)^{/} s^{-1} \Sigma s^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$ olan tek değişkenli normaldir[3]. Burada *s*, örneklem varyans-kovaryans matrisi *S* tahmin edicisinin tahmin değeridir.

Diskriminant fonksiyonlarının dağılımları göz önüne alındığında, Π_1'den bir birimin hatalı atanması durumunda ξ kuralına göre optimal hata oranı,

$$\alpha_1(\xi) = P(U(\underline{X}) \le 0 / \underline{X} \in \Pi_1)$$
$$= \Phi(-\Delta/2)$$

$\hat{\xi}$ kuralına göre gerçek (veya koşullu) hata oranı,

$$\alpha_1(\hat{\xi}) = P(W(\underline{X}) \le 0 / \underline{X} \in \Pi_1, \underline{\bar{x}}_1, \underline{\bar{x}}_2, s)$$

$$= \Phi\left(- \frac{\left[\underline{\mu}_1 - \frac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)\right]^{/} s^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)}{\left[(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)^{/} s^{-1} \Sigma s^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\right]} \right)$$

ve beklenen gerçek hata oranı

$$E(\alpha_1(\hat{\xi})) = E[P(W(\underline{X}) \le 0 / \underline{X} \in \Pi_1)]$$

dır. Burada $\Phi(.)$ standart normal dağılım fonksiyonudur. Π_2'den bir birimin hatalı atanması durumunda da benzer ifadeler elde edilebilir. Buradan önsel olasılıkların eşit olması durumunda toplam hatalı sınıflandırma olasılığı optimal ve gerçek hata oranları için sırasıyla

$$\alpha(\xi) = \frac{1}{2}(\alpha_1(\xi) + \alpha_2(\xi))$$

ve

$$\alpha(\hat{\xi}) = \frac{1}{2}(\alpha_1(\hat{\xi}) + \alpha_2(\hat{\xi}))$$

biçiminde elde edilir[2].

M Tahmin Edicisi ve Güven Aralıkları

Gerçek hata oranı için literatürde bir çok hata oranı tahmin edicisi tanımlanmıştır. Bunların bazıları parametrik, bazıları da parametrik olmayan tahmin edicilerdir[4,6,9,10]. Bu tahmin edicilerin bir çoğunun istatistiki özellikleri için analitik ifadeleri elde etmek zor olduğundan, simülasyon çalışmalarıyla tahmin edicilerin istatistiki özellikleri verilmeye çalışılmıştır[11,12]. Bununla birlikte gerçek hata oranı için asimptotik bir tahmin edici olan M tahmin edicisinin istatistiki özellikleri için analitik ifadeler mevcuttur[13,14]. Bu tahmin edici, Π_1 'den bir birimin hatalı atanması durumunda

$$\begin{aligned}\hat{\alpha}_1^M &= \Phi(-\frac{D}{2})+\phi(-\frac{D}{2})[(p-1)/(Dn_1)+D\{4(4p-1)-D^2\}/(32(n_1+n_2)\\ &+\{(p-1)(p-2)\}/(4Dn_1^2)+(p-1)\{-D^3+8(2p+1)D+(16/D)\}/(64n_1D)\\ &+(D/12288)\{3D^6-4(24p+7)D^4+16(48p^2-48p-53)D^2\\ &+192(-8p+15)\}/(n_1+n_2)^2]\end{aligned}$$

biçiminde tanımlanmıştır. Burada $\phi(.)$ standart normal olasılık yoğunluk fonksiyonu ve $D^2=(\underline{X}_1-\underline{X}_2)^{/}S^{-1}(\underline{X}_1-\underline{X}_2)$ örneklem Mahalanobis uzaklığıdır. Π_2'den bir birimin hatalı atanması durumunda, $\hat{\alpha}_2^M$, $\hat{\alpha}_1^M$ ifadesinde n_1 ve n_2 değerlerinin yer değişmesiyle elde edilir. Buradan M tahmin edicisine göre gerçek hata oranı için tahmini toplam hatalı sınıflandırma olasılığı

$$\hat{\alpha}^M=\frac{1}{2}(\hat{\alpha}_1^M+\hat{\alpha}_2^M)$$

ifadesinden elde edilir.

McLachlan, önsel olasılıkların ve varyans kovaryansların eşitliği altında çok değişkenli normal kitlelerden alınan örneklemleri göz önüne alarak, gerçek hata oranın M tahmin edicine bağlı yaklaşık %$(1-\lambda)$ 'lık güven aralığı ifadesini

$$\alpha_1(\hat{\xi}):\hat{\alpha}_1^M \pm \{\sqrt{v_1(DS)}\}\ \Phi^{-1}(1-\frac{\lambda}{2})$$

biçiminde elde etmiştir[15]. Bu güven aralığı ifadesi örneklem hacimlerinin büyük olması durumunda $\hat{\alpha}_1^M-\alpha_1(\hat{\xi})$ ifadesinin dağılımının yaklaşık olarak $N(0,v_1(\Delta))$ olmasından elde edilmiştir. Burada,

$$\begin{aligned}v_1(\Delta)=\{\phi(\frac{\Delta}{2})\}^2[\frac{1}{n_1}+\frac{(\Delta^2/8)}{n_1+n_2}+\{\frac{\Delta^2+4(3p-4)+(p^2-4p+5)(16/\Delta^2)}{(4n_1)^2}\}\\ +\{\frac{(\Delta^2-2p)/8}{n_1n_2}\}+\{\frac{\Delta^4+2(11p-16)\Delta^2+8(5p-4)}{(64n_1(n_1+n_2))}\}\\ +\{\frac{2\Delta^6+16(2p-5)\Delta^4-32(4p-13)\Delta^2}{32(n_1+n_2)^2}\}]\end{aligned},$$

$DS=\frac{n_1+n_2-p-3}{n_1+n_2-2}D^2$ ve $\Phi^{-1}(1-\frac{\lambda}{2})$, %$(1-\frac{\lambda}{2})$ olasılığına karşılık gelen standart normal tablo değeridir.

Π_2'den bir birimin hatalı atanması durumunda, $\alpha_2(\hat{\xi})$ için güven aralığı, $\alpha_1(\hat{\xi})$ ifadesinde n_1 ve n_2 değerlerinin yer değişmesiyle elde edilir. Buradan M tahmin edicisine göre gerçek hata oranı için tahmini güven sınırları

$$\alpha(\hat{\xi}):\hat{\alpha}^M \pm \{\sqrt{v(DS)}\}\ \Phi^{-1}(1-\frac{\lambda}{2})$$

ifadesinden elde edilir.

Bootstrap Güven Aralığı

Bootstrap, bir yeniden örneklem tekniğidir. Bazı parametre tahmin edicileri için analitik ifadeler mevcut olmadığında, bootstrap tekniği kullanılarak bu parametre tahmin edicileri için değerler elde etmek mümkündür. Bu yöntemde, kitleden alınan n birimlik örneklemden, iadeli olarak n birimlik yeni örneklemler oluşturulur. Oluşturulan bu örneklemler göz önüne alınarak, ilgilenilen parametre tahmin edicileri için tekrarlı örneklem sayısı kadar tahmin değeri elde edilir. Bu tahmin değerleri, parametre tahmin edicileri için bir örneklem oluşturur. Oluşturulan bu örneklemden, tahmin edicilerin bilinmeyen tahmin değerleri elde edilir[16,17,18].

Bootstrap yöntemi diskriminant analizinde uygulandığında, Π_1 ve Π_2 kitlelerinden alınan n_1 ve n_2 hacimli rasgele örneklemlerden iadeli çekilişler yapılarak oluşturulan yeni n_1 ve n_2 hacimli örneklemler göz önüne alınarak hata oranı tahmin edicileri için tahmin değeri elde edilir. Bu işlem bootsrap tekrar sayısı olan B defa tekrarlanırsa, B tane hata oranı tahmin değeri yani hata oranı için bir örneklem elde edilmiş olur. Bu B tane hata oranı tahmin değerinin toplamının B'ye oranlanmasıyla, hata oranı tahmin edicisinin bootstrap tahmin değeri elde edilmiş olur. Böylece $\hat{\alpha}$ hata oranı tahmin edicisinin bootstrap tahmin değeri

$$\hat{\alpha}^B = \frac{1}{B}\sum_{b=1}^{B}\hat{\alpha}^b$$

eşitliği elde edilir. Burada $\hat{\alpha}^b$, her bir bootstrap tekrarı için $\hat{\alpha}$ 'nın tahmin değeridir. Elde edilen bu $\hat{\alpha}^b$ bootsatrap tahmin değerleri küçükten büyüğe doğru sıralandığında, sıra istatistiklerine dayalı dengeli bir bootstrap güven aralığı oluşturulabilir. $\hat{\alpha}^b$ bootstarap değerleri küçükten büyüğe doğru sıralandığında, baştan $(\lambda/2)$ inci ve sondan $(1-\lambda/2)$ inci yüzdeliklere karşı gelen tahmin değerleri gerçek hata oranı için oluşturulan $\%(1-\lambda)$'lık bootstrap güven aralığının sırasıyla alt ve üst sınırlarını verir.

Simülasyon Çalışması ve Sonuçlar

Bu bölümde simülasyon ile gerçek hata oranı için M tahmin edicisine bağlı hem analitik ifadeye hem de bootstrap yöntemine göre elde edilen güven sınırları değerlendirilmektedir. Simülasyon çalışmasında, değişken sayısı $p = 2, 3, 5$ olan çok değişkenli aynı(ortak) varyans-kovaryanslı normal dağılımlı kitleler için $\Delta = 1, 2, 3$, $n_1 = n_2 = n = 25, 50, 100$, bootsrap tekrar sayı $B = 1000$ ve $(1-\lambda) = 0.95$ alınmıştır. Sonuçlar Tablo 1 - 3 ' de verilmiştir. Tablolarda; $\alpha(\xi)$ optimal hata oranını, $\alpha(\hat{\xi})$ gerçek hata oranını, $\hat{\alpha}^M$ gerçek hata oranının M tahmin edicisine ilişkin tahmin değerini, $\hat{\alpha}^M_{alt}$ ve $\hat{\alpha}^M_{üst}$ gerçek hata oranının M tahmin edicisine bağlı güven aralığının alt ve üst sınırlarını, $\hat{\alpha}^{MB}$ gerçek hata oranının M tahmin edicisine bootstrap yönteminin uygulanmasıyla elde edilen tahmin değerini, $\hat{\alpha}^{MB}_{alt}$ ve $\hat{\alpha}^{MB}_{üst}$ gerçek hata oranının M tahmin edicisine bootstrap yönteminin uygulanmasıyla elde edilen tahmin değerlerine göre bulunan güven aralığının alt ve üst sınırlarını göstermektedir.

Tablo 1 : $\Delta = 1$ ($\alpha(\xi) = 0.3085$) için Hata oranları ve Güven Sınırları

p	2			3			5		
n	25	50	100	25	50	100	25	50	100
$\alpha(\hat{\xi})$	0.3160	0.3412	0.3090	0.3256	0.3111	0.3088	0.3572	0.3216	0.3139
$\hat{\alpha}^M$	0.3696	0.3559	0.2799	0.3049	0.2713	0.3032	0.2883	0.3561	0.3037
$\hat{\alpha}^M_{alt}$	0.2153	0.2495	0.2364	0.1592	0.1741	0.2319	0.1370	0.2402	0.2303
$\hat{\alpha}^M_{üst}$	0.5239	0.4624	0.3234	0.4507	0.3685	0.3745	0.4395	0.4721	0.3771
$\hat{\alpha}^{MB}$	0.3626	0.3501	0.2793	0.2806	0.2650	0.2984	0.2512	0.3350	0.2922
$\hat{\alpha}^{MB}_{alt}$	0.2524	0.2758	0.2454	0.1753	0.1816	0.2521	0.1192	0.2672	0.2455
$\hat{\alpha}^{MB}_{üst}$	0.4873	0.4228	0.3121	0.3879	0.3463	0.3467	0.3958	0.4078	0.3413

Tablo 2: $\Delta = 2$ ($\alpha(\xi) = 0.1587$) için Hata oranları ve Güven Sınırları

p	2			3			5		
n	25	50	100	25	50	100	25	50	100
$\alpha(\hat{\xi})$	0.1700	0.1665	0.1606	0.1753	0.1763	0.1610	0.1670	0.1630	0.1627
$\hat{\alpha}^{M}$	0.1550	0.1717	0.1577	0.1659	0.1542	0.1911	0.1189	0.1657	0.1493
$\hat{\alpha}^{M}_{alt}$	0.0479	0.0929	0.1046	0.0547	0.0794	0.1325	0.0244	0.0870	0.0973
$\hat{\alpha}^{M}_{üst}$	0.2621	0.2505	0.2109	0.2770	0.2290	0.2497	0.2134	0.2443	0.2012
$\hat{\alpha}^{MB}$	0.1415	0.1676	0.1557	0.1543	0.1444	0.1864	0.1009	0.1520	0.1429
$\hat{\alpha}^{MB}_{alt}$	0.0736	0.1186	0.1166	0.0744	0.0913	0.1408	0.0421	0.0794	0.1001
$\hat{\alpha}^{MB}_{üst}$	0.2168	0.2195	0.1960	0.2412	0.1981	0.2344	0.1648	0.2298	0.1866

Tablo 3: $\Delta = 3$ ($\alpha(\xi) = 0.0668$) için Hata oranları ve Güven Sınırları

p	2			3			5		
n	25	50	100	25	50	100	25	50	100
$\alpha(\hat{\xi})$	0.0774	0.0686	0.0 677	0.0683	0.0696	0.0689	0.0731	0.0779	0.0689
$\hat{\alpha}^{M}$	0.0816	0.0745	0.0 797	0.1034	0.0959	0.0670	0.1002	0.0416	0.0737
$\hat{\alpha}^{M}_{alt}$	0.0060	0.0251	0.0 437	0.0169	0.0385	0.0347	0.0145	0.0076	0.0522
$\hat{\alpha}^{M}_{üst}$	0.1571	0.1238	0.1 156	0.1900	0.1534	0.0992	0.1859	0.0756	0.0952
$\hat{\alpha}^{MB}$	0.0739	0.0714	0.0 777	0.0895	0.0920	0.0647	0.0819	0.0367	0.0720
$\hat{\alpha}^{MB}_{alt}$	0.0204	0.0335	0.0 528	0.0422	0.0470	0.0402	0.0366	0.0118	0.0556
$\hat{\alpha}^{MB}_{üst}$	0.1346	0.1162	0.1 046	0.1381	0.1408	0.0929	0.1304	0.0688	0.0901

Tablolar incelendiğinde, bütün durumlarda bootstrap güven aralıklarının daha dar olduğu görülmektedir. Bununla birlikte örneklem hacmi artırıldığında, hem analitik ifadeden, hem de bootstrap yönteminden elde edilen aralıklar daha daralmaktadır. Δ değeri artırıldığında yani kitleler birbirinden uzaklaştırıldığında hata oranı değeri küçüldüğünden, buna orantılı olarak güven aralıkları da daralmaktadır. Değişken sayısının, aralıklar üzerinde çok etkili olmadıkları da söylenebilir.

Bu sonuçlardan, gerçek hata oranı için literatürde bilinen diğer tahmin edicilere bağlı analitik olarak elde edilemeyen güven aralıkları, bootstrap tekniği kullanılarak elde edilebilir.

Kaynaklar

[1] Welch. B. L. Note on the Discriminant Functions. *Biometrica*, 31, 218-220, (1939).

[2] Anderson, T. W. *An Introduction to Multivariate Satistical Analysis*, Second edition, John Wiley& Sons. Inc., New York, (1984).

[3] Lachenbruch, P. A. *Discriminant Analysis*. Hafner Press., New York, (1975).

[4] Smith, C.A.B. Some Examples of Discrimination, Annals of Eegenics, 18, 272-282, (1947).

[5] Hills, M. Allocation Rules and their Error Rates", Journal pf the Royal Statistical Society, Ser. B, 28, 1-20, (1966).

[6] Lachenbruch, P. A. & Mickey, M. R. Estimation of Error Rates in Discriminant Analysis. *Technometrics*, 10, 1-11, (1968).

[7] Seber, G. A. F. *Multivariate Observations*. John Wiley & Sons. Inc., (1984).

[8] Wald, A. On Statistical Problems Arising in the Clasification of an Individual into One of Two Groups, Annals of Mathematical Statistics, 15, 145-162, (1944).

[9] Snapinn, S. M. An Evaluation of Smoothed Error Rate Estimators in Discriminant Analysis, *Institute of Statistics Mimeo Series* No:1483, University North Carolina at Chapel Hill, (1983).

[10] Atakan, C. Ve Öztürk, F. Comparisons of Some Smoothed Error Rate Estimators in Discriminant Analysis, Hacettepe Bulletin of Natural Sciences and Engineering Series B, 27, 51-64, (1998a).

[11] Atakan, C., Öztürk, F. Diskriminant Analizinde Hata Oranı Tahmin Edicilerinin Yan ve Varyanslarının Jackknife ve Bootstrap Tahminleri, Celal Bayar Üniversitesi Fen-Edebiyat Fakültesi Dergisi, Fen Bilimleri Serisi(Matematik),4, 14-20, (1998b).

[12] Meaux, L. M., Young, D. M., Seaman, J. W. A Comparision of Parametric Conditional Error Rate Estimators for the Two-Group Linear Discriminant Function, J. Statist. Comput. Simul., Vol.69, 277-291, (2001).

[13] McLachlan, G. J. (1974). An Asimptotic Unbiased Technique for Estimation the Error Rates in Discriminant Analysis. *Biometrics*, 30, 239-249, (1974).

[14] McLachlan, G. J. *Discriminant Analysis and Statistical Pattren Recognition*, John Wiley & Sonsc Inc., New York, (1992).

[15] McLachlan, G. J. (1975). Confidence Intervals for the Conditional Probability of Misallocation in Discriminant Analysis, Biometrics, 32, 161-167, (1975).

[16] Efron, B. Bootstrap Methods: Another look at the Jackknife, *Annals of Stat*., 7, 1-26, (1979).

[17] Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. J. W. Arrowsmith ltd., Bristol, England, (1982).

[18] Efron, B. & Gong. A Leisurely Look at the Bootstrap, the Jackknife and Cross Validation. *The American Statistician*, 37, 36-48, (1983).