Erciyes Üniversitesi

Fen Bilimleri Enstitüsü Dergisi

Cilt 34 Sayı 2, 2018

Erciyes University

Journal of Institue Of Science and Technology

Volume 34, Issue 2, 2018

|  |
| --- |
|  |
| **DAYANIKLI LİNEER DİSKRİMİNANT ANALİZİ İÇİN YENİ BİR YAKLAŞIM****B. Barış ALKAN\*1, Cemal ATAKAN2  Nesrin ALKAN1** 1 Sinop Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 57000, SİNOP 2 Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06100 Tandoğan, ANKARA(Alınış / Received: 28.07.2018, Kabul / Accepted: 30.08.2018, Online Yayınlanma / Published Online: 31.08.2018) |
|  |
|  |  |
| **Anahtar Kelimeler**Minimum kovaryans determinant,Dayanıklı lineer diskriminant analizi, Aykırı gözlemler | **Öz:** Lineer diskriminant analizi, önceden bilinen *p* sayıda özelliğe sahip birimleri, doğadaki gerçek sınıflarına en doğru şekilde atamayı amaçlayan çok değişkenli istatistiksel bir yöntemdir. Burada hedef, birimleri gerçek sınıfına minimum hatayla atamaktır. Lineer Diskriminant Analizi (LDA), veri kümesinde diğer gözlemlerden farklı hareket eden ve aykırı gözlem olarak adlandırılan gözlemlerin varlığında dayanıklı bir yöntem değildir ve güvenilir sonuçlar vermeyebilir. Böyle durumlarda, klasik LDA’nın dayanıklı versiyonlarının kullanımının gerekliliği üzerine literatürde birçok çalışmaya rastlamak mümkündür. Bu çalışmada, jackknife yeniden örnekleme yaklaşımı, minimum kovaryans determinant (MKD) ve LDA yönteminin bir kombinasyonu ile LDA’nın yeni bir dayanıklı versiyonu elde edilmiştir. Önerilen bu yeni yaklaşım ile Croux ve Dehon (2001) tarafından önerilen (Yöntem-1), Hawkins ve McLachlan (1997) tarafından önerilen (Yöntem-2) yaklaşımların aykırı gözlem oranındaki değişimlere göre nasıl etkilendiği yapay veri uygulaması ve benzetim çalışması üzerinden değerlendirilmektedir. Elde edilen bulgular ışığında, önerilen yaklaşımın diğer iki yaklaşıma göre, veri kümesinde aykırı gözlemlerin varlığında performansının bazı durumlarda daha iyi, bazı durumlarda ise en az onlar kadar iyi olduğu görülmektedir. |
|  |  |
|  |  |
| **A NEW APPROACH FOR ROBUST LINEAR DISCRIMINANT ANALYSIS**  |
|  |
|  |
| **Keywords**Minimumcovariance determinant, robust linear discriminant analysis, outliers | **Abstract:** Linear Discriminant Analysis is a multivariate statistical method aiming to alocate individuals with known *p* properties by minimizing the probability of misclassification to real groups (classes) in the natural environment. Linear Discriminant Analysis (LDA) is not a robust method against the presence of outliers in the dataset, and the consequences of using conventional LDA in the presence of outliers in the dataset are far from real. For this reason,it is appropriate to use robust versions of linear discriminant analysis in the presence of outliers.In this study, a new robust version of the LDA is obtained with a combination of the Jackknife resampling approach, minimum covariance determinant (MCD) and LDA method. Proposed new approach is compared with Method-1 proposed by Croux ve Dehon (2001) and Method-2 proposed by Hawkins and McLachlan (1997). These methods will be evaluated through artificial data application and simulation. In the light of the findings the proposed DLDA seemed to give better (or at least as good as them) results in the presence of outliers in the dataset when we compared to the other two approaches based on the classical MCD. |
|  |  |

bbalkan@sinop.edu.tr

**1. Giriş**

Lineer diskriminant analizi, önceden bilinen p sayıda özelliğe sahip birimleri, doğadaki gerçek sınıflarına en doğru şekilde atamayı amaçlayan çok değişkenli istatistiksel bir yöntemdir [1]. Burada hedef, birimleri gerçek sınıfına minimum hatayla atamaktır. Sınıflar arasındaki ayrımı tanımlayan kurallar, Lineer Diskriminant Analizi (LDA) kullanılarak elde edilir. LDA’da kitlelerin eşit varyans-kovaryans matrisine sahip olma varsayımları vardır [2],[3]. Bu nedenle LDA, veri kümesinde diğer gözlemlerden farklı hareket eden ve aykırı gözlem olarak adlandırılan gözlemlerin varlığında dayanıklı bir yöntem değildir ve bu durumda klasik LDA’nın güvenilir sonuçlar vermesi beklenemez. Böyle durumlarda, klasik LDA’nın dayanıklı versiyonlarının kullanımının önemi üzerine literatürde birçok çalışmaya rastlamak mümkündür. Hawkins ve McLachlan (1997) çalışmalarında aykırı gözlemlerin etkisini azaltmak için yüksek bozulma kriteri geliştirmişlerdir. Croux vd. (2008) klasik LDA ile dayanıklı yaklaşımların sınıflandırma etkinliklerini araştırmışlardır. Rousseeuw (1984) Minimum kovaryans determinant (MKD) yöntemini geliştirmiştir. Rousseeuw ve Driessen (1999) tarafından aykırı gözlemlerin gözlemlenmesinde oldukça iyi ve dayanıklı özelliklere sahip hızlı minimum kovaryans determinant isminde yeni bir tahmin edici geliştirmişlerdir. Todorov (2007), aykırı gözlem problemini önlemek için kontamine olmuş verilerden etkilenmeyen minimum kovaryans determinantını temel alan dayanıklı Wilks lambda istatistiğinin kullanımını önermiştir. Zaman içerisinde farklı yöntemler ve yaklaşımların parametre tahmininin dayanıklılığını artırmak için MKD'ın özelliklerini kullanmaktadır. Bu yöntem çok değişkenli ortalama vektörü ve varyans-kovaryans matrisinin dayanıklı tahminlerini bulmak için kullanılmıştır. MKD algoritması Wiegand vd. (2009) tarafından genetik bir algoritma için kullanılmıştır. Hubert vd. (2012) çalışmalarında deterministik minimum kovaryans determinantı olarak adlandırılan dayanıklı konum ve ölçek parametreleri için yeni bir tahmin edici önermiştir ve önerdikleri bu tahmin edicinin performansını Maronna ve Zamar (2002)'ın önerdiği iki tahmin edici ile karşılaştırmışlardır. Alkan (2016) çalışmasında, jackknife yeniden örnekleme yaklaşımı kullanılarak MKD yönteminin bir uyarlamasını elde etmiştir. Bu uyarlanmış MKD yöntemine dayanan temel bileşenler analizinin sonuçlarının klasik MKD’ya göre aykırı gözlemlere karşı daha dayanıklı bir performans sergilediğini göstermiştir.

Bu çalışmada, jackknife yeniden örnekleme yaklaşımı, minimum kovaryans determinant (MKD) ve LDA yönteminin bir kombinasyonu ile LDA’nın yeni bir dayanıklı versiyonu elde edilmiştir. Önerilen bu yeni yaklaşım ile Croux ve Dehon (2001) tarafından önerilen (Yöntem 1), Hawkins ve McLachlan (1997) tarafından önerilen (Yöntem-2) yaklaşımlarının aykırı gözlem oranındaki değişmelerden nasıl etkilendiği yapay veri uygulaması ve benzetim çalışması üzerinden değerlendirilmektedir. Önerilen yaklaşım ve diğer yöntemler görünüşte hata oranı (APER) tahmin edicisi ve toplam varyans açıklama oranı açısından karşılaştırılmaktadır. Ayrıca çalışmada, ortak varyans-kovaryans matrisi ve grup ortalamaları tahminlerinin diskriminant fonksiyonları ve skorlarının kalitesini etkileyeceğinden hareketle, ortak varyans-kovaryans matrisi tahmininden elde edilecek toplam varyans açıklama oranının, karşılaştırma kriteri olarak kullanılmasının uygun olacağı düşünülmektedir.

Çalışmanın ikinci bölümünde lineer diskriminant analizi için temel teorik kavramlar, üçüncü bölümünde minimum kovaryans determinant yöntemi, dördüncü bölümünde dayanıklı diskriminant analizi ve beşinci bölümünde önerilen yeni yaklaşım için gerekli matematiksel teori kısaca özetlenmiştir. Altıncı bölümde, önceki bölümlerde ele alınan yöntemler gerçek ve yapay veri kümeleri üzerinden değerlendirilmektedir. Yedinci bölümde ise bulgular üzerinden genel bir değerlendirme yapılmaktadır.

**2. Materyal ve Metot**

**2.1. Lineer diskriminant analizi**

Diskriminant analizi, bir bireyin ya da bir nesnenin ölçülen özelliklerine göre sonlu miktarda bilinen farklı kitlelerden birine atanmasını sağlayan çok değişkenli istatistiksel bir yöntemdir (Atakan, 2003). Diskriminant analizinde, ayırma fonksiyonu kitleler üzerine oluşturulan birtakım varsayımlara göre elde edilir. Örneklemlerin alındığı kitlelerin, varyans-kovaryans matrislerinin eşit ya da eşit olmadığına göre farklı diskriminant fonksiyonları bulunur. Kitlelerin varyans kovaryans matrisleri eşit olduğunda lineer diskriminant fonksiyonu elde edilirken, farklı olması durumunda karesel diskriminant fonksiyonu elde edilir.

, *g* tane farklı kitle olsun. *p*-boyutlu bir gözlem vektörü yoğunluğu ile kitlesinden gelsin. bireyin sırasıyla kitlelerine ait olması olasılıkları olsun. Burada ’ lara önsel olasılıklar denmektedir ve dir. Genellikle, ortalama vektörü , varyans-kovaryans matrisi olan çok değişkenli bir kitledir. Tüm grupların varyans-kovaryans matrisleri eşit (olduğunda, kitlesine yeni bir gözleminin atanması için ile gösterilen

 , (1)

 sınıflandırma fonksiyonunun değerinin diğer kitlelere göre maksimum olması gerekir [4]. (1) eşitliği ile verilen fonksiyonu, *x*’de doğrusal olduğu için lineer diskriminant fonksiyonu olarak ifade edilir.

**2.2. Minimum kovaryans determinant (MKD) yöntemi**

Hızlı bir hesaplama algoritmasına sahip olan MKD yöntemi, çok değişkenli ortalama vektörü ve varyans-kovaryans matrisinin dayanıklı bir tahmin edicisidir. Varyans-kovaryans matrisini tahmin etmek çok değişkenli istatistiksel yöntemlerin çoğunun temeli olduğundan eğer dayanıklı bir istatistiksel yöntem geliştirilecekse de MKD yönteminin kullanımı oldukça önemlidir. MKD yöntemi ayrıca veri kümesindeki aykırı gözlemlerin tespiti için de kullanılabilmektedir. MKD, temel bileşen analizi, faktör analizi ve diskriminant analizi gibi birçok çok değişkenli istatistiksel yöntemin dayanıklı versiyonlarının geliştirilmesi için kullanılmıştır [5]. MKD yöntemi ile elde edilen tahmin ediciler, veri kümesinde yer alabilecek aykırı gözlemlerden kaynaklanan %50’ye kadar bir bozulmaya karşı dayanıklı duruş sergilemektedirler.

MKD yönteminin amacı, *n* tane gözlem üzerinden varyans-kovaryans matrisinin determinantı minimum olacak şekilde *h* tane gözlemden oluşan alt kümeyi bulmaktır. *h* değeri *n* ve [(*n + p +* 1*)/*2] arasında bir değerdir. Burada *p*, değişken sayısıdır. *h=* [(*n + p +* 1*)/*2] olarak alındığında, MKD mümkün olan en büyük bozulma noktasına sahip olur. Sadece *h>p* olduğunda MKD tahmin edicisi hesaplanabilir. Bu yüzden *n>*2*p* olmalıdır [6]. MKD hesaplaması için gerekli algoritma aşağıda özet bir şekilde verilmiştir.

*MKD Algoritması*

*Adım1. h sayısını elde edilir.*

*Adım2. n’in h’lı kombinasyonunu kadar h boyutlu alt küme elde edilir.*

*Adım3. Bulunan tüm h boyutlu alt kümeler için ayrı ayrı Adım 3.1 uygulanır.*

*Adım 3.1. Her bir h boyutlu alt küme için Örneklem varyans-kovaryans* *matrisini ve ilgili örneklem varyans-kovaryans* *matrisine karşılık gelen örneklem varyans-kovaryans matrisi determinantı bulunur.*

*Adım 4. Adım 3.1’de hesaplanan* *n’in h’lı kombinasyonunu kadar örneklem varyans-kovaryans matrisi determinantından minimum olan determinantı veren alt küme bulunur.*

*Adım 5. Adım 4’de elde edilen örneklem varyans-kovaryans matrisi determinantı minimum olan alt kümenin çok değişkenli ortalama vektörü ve varyans-kovaryans matrisleri örneklemin MKD dayanan dayanıklı tahminleridir.*

**2.3. Dayanıklı lineer diskriminant analizi**

Klasik LDA, veri kümesinde aykırı gözlemlerin olması durumuna karşı dayanıklı olmadığından, bulunan sonuçlar gerçekte olması gerekenden oldukça farklı elde edilir. Bu nedenle, böyle durumlarda, lineer diskriminant analizinin dayanıklı versiyonunun kullanılması güvenilir çıktılar elde edilmesi açısından hayati önem taşımaktadır. LDA’nin dayanıklı versiyonu, kitle ortalama vektörü ve kitle varyans-kovaryans matrisi parametrelerinin, ve dayanıklı tahmin edicileriyle yer değiştirilmesi ile bulunabilir [4]. Çok değişkenli veri kümesi için ortalama vektörü ve varyans-kovaryans matrisi parametrelerinin dayanıklı tahmini yüksek bozulma değeri sağlayan MKD yöntemi ile bulunabilir.

Ortak varyans-kovaryans matrisini tahmin etmek için ise en kolay yol, bireysel gruplardan elde edilen varyans-kovaryans matrisilerinin tahmin edicilerinin birleştirilmesiyle

 (2)

elde edilmesidir [7],[8].

Yüksek bozulma değeri tahmin edicisine dayanan ortak varyans-kovaryans matrisini tahmin etmek için birçok yöntem vardır. Bu çalışmada yüksek bozulma nokta tahmin edicisine dayanan ortak varyans-kovaryans matrisinin tahmini için aşağıda açıklamaları verilen yöntemler ve sonraki bölümde tanıtılacak önerilen yeni bir yaklaşım kullanılmıştır.

*Yöntem-1:* Croux ve Dehon (2001) tarafından önerilen bu yöntem S tahminlerine dayanan lineer diskriminant analizini sağlamlaştırmak için uygulanmıştır.

*Yöntem-2:* Hawkins ve McLachlan (1997) tarafından önerilen bu yöntem FSA’ya (Feasible Solution Algorithm) dayanan en küçük grup içi kovaryans determinantı tahmin edicisidir.

Yukarıda verilen yöntemlerden biri ile elde edilen ve tahminleri kullanılarak başlangıç dayanıklı uzaklıkları,

 (3)

ile hesaplanabilir. Burada , *k* ıncı kitlenin *i* inci birimidir [9].

Bu başlangıç dayanıklı uzaklıklar ile,, i=1,2,.., ; k=1,2,…,*g*, gözlemi için bir ağırlık tanımlanabilir.

Eğer, ise , aksi takdirde alınır.

Bu ağırlıklar ile dayanıklı sınıflandırma kurallarını yapılandırmak için gerekli olan en son yeniden ağırlıklandırılmış ortak varyans-kovaryans matrisi, ve grup ortalamaları vektörü nün ağırlıklandırılmış tahminleri

olarak hesaplanır. Burada , *k*. grup için ağılıkların toplamı ve ağırlıkların genel toplamı olup,

biçimindedir.

**2.4. LDA’nın Yeni Bir Dayanıklı Yaklaşımı**

MKD yönteminin Jackknife yeniden örnekleme yaklaşımına göre modifiye edilmiş versiyonuna dayanan önerilen dayanıklı LDA algoritması aşağıda sunulmuştur.

*Önerilen yeni bir dayanıklı LDA algoritması*

*Adım1. Her bir sınıfta yer alan j-inci gözlem dışarda bırakılır.*

*Adım2. Her bir sınıfta yer alan için* (*n-*1) *gözlem için sayıları elde edilir.*

*Adım3. Her bir sınıf için* (*n-*1)*’in h’lı kombinasyonu kadar h boyutlu alt küme elde edilir.*

*Adım3.1. Her bir sınıf için elde edilen h gözlemi bulunan alt kümeler için,*

*Adım 3.1.1. Her bir h boyutlu alt küme için Örneklem varyans-kovaryans matrisini ve ilgili örneklem varyans-kovaryans matrisine karşılık gelen örneklem varyans-kovaryans matrisi determinantını bulunur.*

*Adım 4. Her bir sınıf için minimum determinanta sahip olan alt küme seçilir. Bu alt kümelerin her sınıf için örneklem ortalama vektörü ve örneklem varyans-kovaryans matrisi bulunur.*

*Adım 5. Adım* 1-4*, j=*1,2*…,n ve k=*1,2*,..,g için tekrarlanır ve buradan elde edilen her bir sınıf için örneklem ortalama vektörlerinin ortalamasından ve örneklem varyans-kovaryans matrislerinin ortalamasından sırasıyla, çok değişkenli ortama vektörü ve varyans-kovaryans matrisi tahminleri elde edilir.*

*Adım 6. Adım 5’den elde edilen tahminler kullanılarak ortak varyans-kovaryans matrisi tahmini elde edilir ve sınıflandırma fonksiyonları bulunur.*

Bu çalışmada, veri kümesinde aykırı gözlemlerin varlığında, önerilen dayanıklı LDA’nın kullanılmasının, klasik yönteme göre nasıl sonuçlar verdiği gözlenecektir. Bu çalışmada yapılan tüm hesaplamalar ve analizler için R istatistik paketi içerisinde yer alan *robustbase, rrcov* ve *MASS* kütüphanelerinden faydalanarak yazılan R programı kullanılmıştır

**3. Bulgular**

**3.1. Yapay veri uygulaması için elde edilen bulgular**

Yapay veri kümesi olarak üç grup ve her bir grupta beş değişkenin farklı gözlem sayılarıyla (70, 29 ve 33) yer aldığı bir veri kümesi oluşturulmuştur. Bu yapay veri kümesi her bir grup için farklı oranlarda aykırı gözlem içermektedir. Önceki bölümlerde de bahsedildiği gibi veri kümesinde aykırı gözlemlerin varlığında klasik istatistik yöntemlerin yani dayanıklı olmayan yöntemlerin kullanımı oldukça sıkıntılı ve yanlı sonuçlara sebebiyet vermektedir. Veri kümesinde bulunan aykırı gözlemlerin tespit edilmesi için literatürde güvenilir sonuçlar verdiği yapılan çalışmalarla vurgulanmış uyarlanmış kuantil (adjusted quantile) metodu kullanılmıştır. Bu metot ile 14 gözlem aykırı gözlem olarak belirlenmiştir. Ayrıca Grup 1’deki gözlemlerin %2’si, Grup 2’deki gözlemlerin %27’si ve Grup 3’deki gözlemlerin %12’si aykırı gözlem olarak belirlenmiştir. Aykırı gözlemler Şekil 1’de kırmızı renkle gösterilmektedir.



**Şekil 1.** Uyarlanmış kuantil grafiğine göre aykırı gözlemlerin (kırmızı noktalar) dağılımları

Bu yapay veri kümesi sırasıyla önerilen yaklaşım, Yöntem-1 ve Yöntem-2 ile değerlendirilmiş ve bulgular Tablo 1’de verilmiştir. Tablo 1 incelendiğinde, görünüşte hata oranı tahmini açısından önerilen yaklaşım ve Yöntem-1’nin aynı tahmin oranına sahip olduğu görülmüştür. En düşük görünüşte hata oranı tahmini 0.04 oranı ile Yöntem-2’ye aittir. En büyük özdeğerin toplam varyans açıklama oranına katkısı açısından Tablo 1 incelendiğinde sırasıyla, önerilen yaklaşım %73, Yöntem-1’in %72 ve Yöntem-2’ün %70 olduğu görülmektedir.

**Tablo 1.** Önerilen ve diğer iki yöntem için görünüşte hata oranı (APER: Apparent Error Rate) tahmini

|  |  |  |
| --- | --- | --- |
| **Yöntem** | **Görünüşte Hata Oranı (APER:** **Apparent Error Rate)** **Tahmini** | **En Büyük Özdeğerin Toplam Varyans Açıklama Oranı** |
| Önerilen Yaklaşım | 0.08 | 0.73 |
| Yöntem-1 | 0.08 | 0.72 |
| Yöntem-2 | 0.04 | 0.70 |

**3.2. Benzetim çalışması için elde edilen bulgular**

Benzetim çalışması için çok değişkenli normal dağılımdan sayı üretilmiştir. Çalışmanın tasarımı %10, %20, %40 aykırı gözlem oranları için Matlab program kodları kullanılarak aşağıda sunulan şekilde oluşturulmuştur.

*%10 Aykırı gözlem durumu:*

Grup 1: mvnrnd (-ones(3,1),eye(3),90)

 mvnrnd (10\*ones(3,1),eye(3),10)

Grup 2: mvnrnd (ones(3,1),eye(3),90)

 mvnrnd (-10\*ones(3,1),eye(3),10)

*%20 Aykırı gözlem durumu:*

Grup 1: mvnrnd (-ones(3,1),eye(3),80)

 mvnrnd (10\*ones(3,1),eye(3),20)

Grup 2: mvnrnd (ones(3,1),eye(3),80)

 mvnrnd (-10\*ones(3,1),eye(3),20)

*%40 Aykırı gözlem durumu:*

Grup 1: mvnrnd (-ones(3,1),eye(3),60)

 mvnrnd (10\*ones(3,1),eye(3),40)

Grup 2: mvnrnd (ones(3,1),eye(3),60)

 mvnrnd (-10\*ones(3,1),eye(3),40)

Bu tasarıma göre elde edilen veri kümeleri üzerinden benzetim çalışması yapılmış ve üç yöntem için elde edilen sonuçlar Tablo 2’de sunulmuştur. Tablo 2 incelendiğinde, %10 aykırı gözlem oranı durumunda, görünüşte hata oranı tahmini açısından, önerilen yaklaşım en küçük hata oranına sahip olduğu görülmüştür. Yöntem-1 ve Yöntem-2 için bu tahmin oranının aynı olduğu görülmektedir. En büyük özdeğerin toplam varyans açıklama oranına katkısı açısından Tablo 2 incelendiğinde, önerilen yaklaşımın %43.83 ile diğer iki yönteme göre daha iyi bir performansa sahip olduğu görülmektedir. %20 aykırı gözlem oranı durumunda, görünüşte hata oranı tahmini açısından, tüm yöntemlerin aynı tahmin oranına sahip olduğu görülmüştür. En büyük özdeğerin toplam varyans açıklama oranına katkısı açısından Tablo 2 incelendiğinde, önerilen yaklaşım ve Yöntem-1 sırasıyla %40.66 ve %40.11 oranlarına sahipken, Yöntem-2 ise %39.66 oranlarını aldıkları görülmektedir. %40 aykırı gözlem oranı durumunda, görünüşte hata oranı tahmini açısından, önerilen yaklaşım en küçük hata oranına sahip olduğu görülmektedir. Yöntem-1 ve Yöntem-2 için bu tahmin oranının aynı olduğu görülmektedir. En büyük özdeğerin toplam varyans açıklama oranına katkısı açısından Tablo 2 incelendiğinde, önerilen yaklaşımın %41.41 ile diğer iki yönteme göre daha iyi bir performansa sahip olduğu görülmektedir.

**Tablo 2.** Önerilen ve diğer iki yöntem için görünüşte hata oranı (APER: Apparent Error Rate) tahmini simülasyon denemesi sonuçları

|  |  |  |
| --- | --- | --- |
| **Yöntem** | **Görünüşte Hata Oranı (APER: Apparent Error Rate) Tahmini**  | **En Büyük Özdeğerin Toplam Varyans Açıklama Oranı** |
|  | %10 Aykırı Gözlem | %20 Aykırı Gözlem | %40 Aykırı Gözlem | %10 Aykırı Gözlem | %20 Aykırı Gözlem | %40 Aykırı Gözlem |
| Önerilen Yaklaşım | 0.1400 | 0.2150 | 0.4250 | 0.4383 | 0.4066 | 0.4141 |
| Yöntem-1 | 0.1450 | 0.2150 | 0.4300 | 0.4094 | 0.4011 | 0.4112 |
| Yöntem-2 | 0.1450 | 0.2150 | 0.4300 | 0.4382 | 0.3966 | 0.4124 |

**4. Tartışma ve Sonuç**

Diskriminant analizi gibi çoğu çok değişkenli istatistiksel yöntem, veri kümesinde aykırı gözlemlerin olması durumundan oldukça fazla etkilenir. Böyle durumlarda klasik istatistiksel yöntemler yerine onların dayanıklı versiyonları olan yöntemler bir alternatif olarak kullanılabilir. Bu yöntemlerden bazıları S tahmin edicilerini, bazıları M tahmin edicilerini ve bazıları da en küçük kovaryans determinant yöntemiyle elde edilen tahmin edicilere dayanmaktadır. Son yıllarda dayanıklı çok değişkenli istatistik yöntemlerden, minimum kovaryans determinant tahmin edicisi (MKD) oldukça yaygın kullanılmaktadır. Bu nedenle MKD’nin sonuçlarının çeşitli modifikasyonlarla iyileştirilmesi ve dolasıyla MKD’ye dayanan yöntemlerin de sonuçlarının böyle modifikasyonlardan pozitif olarak etkilenip etkilenmediğinin araştırılması önemlidir.

Bu çalışmada, önerilen dayanıklı LDA yaklaşımı ile Croux ve Dehon (2001) tarafından önerilen (Yöntem 1), Hawkins ve McLachlan (1997) tarafından önerilen (Yöntem-2) yaklaşımların aykırı gözlem sayısındaki değişmelerden nasıl etkilendiği yapay veri uygulaması ve benzetim çalışmaları üzerinden Görünüşte Hata Oranı ve En Büyük Özdeğerin Toplam Varyans Açıklama Oranı kriterleri çerçevesinde değerlendirildiğinde önerilen yaklaşımın performansının biraz daha iyi (ya da en az onlar kadar iyi) olduğu görülmektedir.

Çalışmadan elde edilen bulgular ışığında, diğer çok değişkenli dayanaklı istatistiksel yöntemler için de önerilen yaklaşımın uygunluğunun araştırılabilecek olmasının lisansüstü tez çalışmalarına ve mühendislik gibi uygulama alanlarında yapılacak gelecekteki çalışmalara katkı sağlayacağı düşünülmektedir.

**Teşekkür**

Bu çalışma, Sinop Üniversitesi tarafından FEF-1901-16-06 proje numarası ile desteklenmiştir.

**Kaynakça**

1. Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of human genetics, 7(2), 179-188.
2. Johnson, R. A., & Wichern, D. 2002. Multivariate analysis. John Wiley & Sons, Ltd
3. Alrawashdeh, M. J., Radwan, T. R., & Abunawas, K. A. 2018. Performance of linear discriminant analysis using different robust methods. European Journal of Pure and Applied Mathematics, 11(1), 284-298.
4. Todorov V. and Filzmoser P. 2009. An object oriented framework for robust multivariate analysis. Journal of Statistical Software, Vol. 32(3), pp. 1-47.
5. Hubert M, Debruyne M, Rousseeuw PJ. 2017. Minimum covariance determinant and extensions. WIREs Comput Stat.;e1421.
6. Rousseeuw, P. J., & Hubert, M. 2018. Anomaly detection by robust statistics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8:e1236.
7. Anderson, T.W. 1984. An introduction to multivariate statistical analysis. Second edition, New York, John Wiley & Sonsc Inc.
8. Todorov, V. and Pires, A.M. 2007. Comparative Performance of Several Robust Linear Discriminant Analysis Methods, REVSTAT Statistical Journal
9. Rousseeuw, P.J. and van Zomeren, B.C. 1991. Robust distances: Simulation and cutoff values. In: Directions in Robust Statistics and Diagnostics, Part II (W. Stahel and S. Weisberg, Eds.), Springer Verlag, New York.
10. Alkan, B. B. 2016. Aykırı Gözlemlerin Varlığında Uyarlanmış En Küçük Kovaryans Determinant Tahminine Dayalı Dayanıklı Temel Bileşenler Analizi. Alphanumeric Journal, 4(2), 85-94.
11. Atakan, C. 2003. Diskriminant Analizinde Gerçek Hata Oranına İlişkin Güven Aralığı için Bir Simülasyon Çalışması. Selçuk Üniversitesi Fen Fakültesi Fen Dergisi, 1(22), 89-96.
12. Croux, C. and Dehon, C. 2001. Robust linear discriminant analysis using s-estimators, The Canadian Journal of Statistics, 29, 473–492.
13. Croux, C., Filzmoser, P. and Joossens, K. 2008. Classification efficiencies for robust linear discriminant analysis. Statistica Sinica, 18(2):581–599.
14. Hawkins, D.M. and McLachlan, G.J. 1997. High-Breakdown Linear Discriminant Analysis, Journal of the American Statistical Association, 92, 136–143.
15. Hubert, M., Rousseeuw, P. J., & Verdonck, T. 2012. A deterministic algorithm for robust location and scatter. Journal of Computational and Graphical Statistics, 21(3), 618-637.
16. Maronna, R. A., & Zamar, R. H. 2002. Robust estimates of location and dispersion for high-dimensional datasets. Technometrics, 44(4), 307-317.
17. Rousseeuw, Peter J, and Driessen, KV. 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3):212–223.
18. Rousseeuw, P. J. 1984. Least median of squares regression. Journal of the American statistical association, 79(388), 871-880.
19. Todorov, V. 2007. Robust selection of variables in linear discriminant analysis. Statistical Methods and Applications, 15(3), 395-407.
20. Wiegand, P., Pell, R., & Comas, E. 2009. Simultaneous variable selection and outlier detection using a robust genetic algorithm. Chemometrics and Intelligent Laboratory Systems, 98(2), 108-114.