

RİDGE REGRESYON ÜZERİNE BİR ÇALIŞMA

Mücahit KURTULUŞ

**YÜKSEK LİSANS
İSTATİSTİK**

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

114736

Kasım 2001

ANKARA

Mücahit KURTULUŞ tarafından hazırlanan RİDGE REGRESYON ÜZERİNE BİR ÇALIŞMA adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

gönen bne

Prof.Dr. Soner GÖNEN

Tez Yöneticisi

Bu çalışma jürimiz tarafından İstatistik Ana Bilim Dalı Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Soner GÖNEN *gönen bne*

Üye : Prof. Dr. Hamza GÜNGÖR *H. Güngör*

Üye : Doç. Dr. Aydın ERAR *A. Erar*

Üye : _____

Üye : _____

Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

Elhan Demir

İÇİNDEKİLER

Sayfa

ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
1. GİRİŞ	1
2. ÇOKLU LİNEER REGRESYON MODELLERİ VE ÖZELLİKLERİ.....	2
2.1. Lineer Modelin Tanımlanması.....	2
2.2. Regresyon Denklemlerinin Bulunması ve Özellikleri	2
2.2.1. En küçük kareler ölçütü	3
2.2.2. Parametrelerin en küçük kareler kestirimleri ve özellikleri	3
2.3. Çoklu Bağlantı Kavramı	8
2.3.1. Çoklu bağlantı ölçütü	8
2.4. Yanlı Kestirim Kavramı.....	9
2.5. Ridge Regresyon	9
2.5.1. Ridge kestiricisinin en küçük kareler kestiricisi ile ilişkisi.....	10
2.5.2. Z ve W matrislerinin özellikleri	11
2.5.3. Ridge kestiricisinin beklenen değeri	12
2.5.4. Ridge kestiricisinin varyansı	12

2.6. Ridge Kestiricisi ile En Küçük Kareler Kestiricisinin Karşılaştırılması.....	13
2.7. Ridge Kestiricisinin Hata Kareler Ortalaması	14
2.7.1. Ridge regresyonda hata kareler ortalaması fonksiyonu ile ilgili teoremler ..	15
2.8. Çoklu Bağlantının Ridge Kestiricisi Üzerindeki Etkileri	19
2.9. Ridge Regresyonda k Parametresinin Belirlenmesi	20
2.9.1. Ridge izi yöntemi	20
2.9.2. Tekrarlayıcı kestirim yöntemi	21
2.9.3. k'nın nokta kestirimi yöntemi	23
3. UYGULAMA	24
4. SONUÇ VE ÖNERİLER.....	27
KAYNAKLAR	28
EKLER.....	29
ÖZGEÇMİŞ	35

RIDGE REGRESYON ÜZERİNE BİR ÇALIŞMA
(Yüksek Lisans Tezi)

Mücahit KURTULUŞ

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ
Kasım 2001

ÖZET

Bu çalışmada, bağımsız değişkenler arasında çoklu bağlantı olması durumunda ortaya çıkan sorunların giderilmesi yöntemlerinden biri olan Ridge Regresyon yöntemi üzerinde çalışılmış ve Ridge Regresyon yöntemi ile En Küçük Kareler yöntemi karşılaştırılmıştır.

Ridge Regresyon yönteminde k parametresinin bulunmasında kullanılan birkaç yöntem tanıtılmış ve bir yöntem önerilmiştir. Son bölümde örneğimizden elde edilen Ridge Regresyon sonuçları ile, En Küçük Kareler yönteminin sonuçları karşılaştırılmıştır.

Bilim Kodu : 406.01.01

Anahtar Kelimeler : Ridge izi, Çoklu bağlantı, Ridge Regresyon modelinde k parametresinin belirlenmesi, Ridge Kestiricisinin hata kareler ortalaması

Sayfa Adedi : 35

Tez Yöneticisi : Prof.Dr.Soner GÖNEN

A STUDY ON RIDGE REGRESSION**M.Sc. Thesis****Mücahit KURTULUŞ****GAZİ UNIVERSITY****INSTITUTE OF SCIENCE AND TECHNOLOGY****NOVEMBER 2001****ABSTRACT**

In this study we worked on ridge regression method and especially choosing ridge parameters which is detoted by k and also suggested a heuristic way to calculate k . We also compared ridge regression method by Least Squares Method.

Science Code : 406.01.01

Key Word : Ridge trace, Multicollinearity, at to choose to find ridge parameter of Ridge Regression Method, Mean squared Error of Ridge Estimator.

Paper Number : 35

Advisor : Prof.Dr.Soner GÖNEN

TEŞEKKÜR

Çalışmalarım boyunca yardım ve katkılarıyla beni yönlendiren hocam Prof.Dr.Soner GÖNEN'e, Yrd.Doç.Dr.Celal AYDIN'a ve Araş.Gör.Mustafa SEMİZ'e, manevi desteklerinden dolayı aileme teşekkürü bir borç bilirim.



1. GİRİŞ

İstatistiksel yöntemlerin en çok kullanılanlarından birisi de Çoklu Doğrusal Regresyon yöntemidir. Yöntem, bilim ve teknolojinin hemen hemen her alanında model kurmak, parametreleri kestirmek ve ileriye yönelik kestirimler yapmak için kullanılır.

Önkestirimler (prediction) için veri vektörleri ortogonal olmadığında özellikle çoklubağlantının varlığında bazı sorunlar ortaya çıkmaktadır. Katsayılar mutlak değerce oldukça büyük olma eğilimindedir. Bazı katsayıların da yanlış işaretli olması olasıdır. Verilerde yapılacak çok küçük bir düzeltme, bir-iki gözlemin atılması ya da aynı kitleden aynı yöntemle yeni örneklemelerin alınması regresyon katsayılarında ani değişikliklere neden olabilir. Önkestirim için veri vektörlerinin çoğu diklikten saparsa bu sorunların ortaya çıkma olasılığı da artar.

Önkestirim için veri vektörleri arasındaki korelasyon yüksek ise yani çoklu bağlantı (multicollinearity) var ise “En Küçük Kareler Yöntemi” doğru yargıya varacak sonuçlara götürmez.

2. ÇOKLU LİNEER REGRESYON MODELLERİ VE ÖZELLİKLERİ

2.1. Lineer Modelin Tanımlanması

n , gözlem sayısı ve p bağımsız değişken sayısı olmak üzere Y , $n \times 1$ boyutlu açıklanan rastgele değişkenin gözlenebilir bir vektörü; X , $n \times p$ ($n > p$) boyutlu açıklayıcı değişkenlerin bazı koşullar altında sabitleştirilmiş ya da önceden seçilmiş değerlerinin matrisi; β , $p \times 1$ boyutlu bilinmeyen parametrelerin vektörü; ε , $n \times 1$ boyutlu Y ile aynı dağılıma sahip gözlenemeyen rastgele değişkenlerin vektörü, $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 I$ ve bu nicelikler arasındaki bağıntı

$$Y = X\beta + \varepsilon \quad (2.1)$$

denklemleriyle belirlenen modele Genel Lineer Model veya Gauss-Markov Modeli denir (Hoerl, 1962).

2.2. Regresyon Denklemlerinin Bulunması ve Özellikleri

Bir regresyon modelinin kurulması verileri temsil edebilecek bir denklemin bulunmasıdır. (2.1) denkleminde β parametreleri yerine kestirimleri yazılır ise, örneklem regresyon denklemi;

$$Y = X\hat{\beta} + e \quad (2.2)$$

biçiminde yazılır.

Burada $\hat{\beta}$, $(p \times 1)$ boyutlu parametre katsayılarının kestirim vektörüdür.

e ise $n \times 1$ boyutlu artık vektörü olarak tanımlanır.

$\varepsilon = Y - E(Y)$ iken $e = Y - \hat{Y}$ yazılabilir. Buradan hata teriminin x_i gözlem kümesi için bağımlı değişkenin gözlenen değeri ile beklenen değeri arasındaki fark olduğu söylenir.

2.2.1. En küçük kareler ölçütü

Artık kareler yöntemi artık kareler toplamının en küçüklenmesidir :

$$\text{enk} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{enk} \sum_{i=1}^n e_i^2 \quad (2.3)$$

şeklinde yazılabilir. Artık kareler toplamını en küçük yapmasından dolayı en uygun ölçüt olarak bilinir.

2.2.2. Parametrelerin en küçük kareler kestirimleri ve özellikleri

(2.1) modelindeki β parametresinin (2.3) nolu bağıntıdaki en küçük kareler ölçütü ile kestirilebilmesi için $\sum_{i=1}^n e_i^2$ vektörel olarak,

$$\begin{aligned} e' e &= \sum_{i=1}^n e_i^2 = (Y - \hat{Y})' (Y - \hat{Y}) \\ &= (Y - X \hat{\beta})' (Y - X \hat{\beta}) \\ &= Y' Y - 2 \hat{\beta}' X' Y + \hat{\beta}' X' X \hat{\beta} \end{aligned}$$

yazılabilir. $\hat{\beta}$ 'ya göre kısmi türev alınarak sıfıra eşitlenirse,

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial \hat{\beta}'} = -2X' Y + 2X' X \hat{\beta} = 0$$

$$X' X \hat{\beta} = X' Y$$

ve buradan $\hat{\beta}$ 'nin en küçük kareler kestirimi,

$$\hat{\beta} = (X' X)^{-1} X' Y$$

olarak elde edilir.

$\hat{\beta}$ 'ya göre ikinci kısmi türev alınarak sifıra eşitlenirse

$$\frac{\partial^2 \sum_{i=1}^n \varepsilon_i^2}{\partial \hat{\beta}' \partial \hat{\beta}} = 2 X' X > 0$$

olarak elde edilir. $X' X$ matrisi, tam kesin çoklu bağlantının yokluğunda daima pozitif tanımlı olması sebebiyle bir en küçükleme modelidir. $\hat{\beta}$ en küçük hatayı veren parametre tahminidir.

$\hat{\beta}$ 'nin iki özelliği varyans-kovaryans matrisi ve kendi beklenen değerinden uzaklığı yardımıyla bulunur. $\hat{\beta}$ 'nin beklenen değer ve varyans özelliğini inceleyelim :

$$\begin{aligned} E(\hat{\beta}) &= E[(X' X)^{-1} X' Y] \\ &= E[(X' X)^{-1} X' (X\beta + \varepsilon)] \\ &= (X' X)^{-1} X' X\beta + (X' X)^{-1} X' E(\varepsilon) \\ &= \beta \end{aligned}$$

Buradan , $\hat{\beta}$, β 'nin yansız bir kestirimidir. $V(\hat{\beta})$ ise,

$$\begin{aligned} V(\hat{\beta}) &= V[(X' X)^{-1} X' Y] \\ V(\hat{\beta}) &= \sigma^2 (X' X)^{-1} X' X (X' X)^{-1} \\ V(\hat{\beta}) &= \sigma^2 (X' X)^{-1} \end{aligned}$$

olarak bulunur. $X' X$ matrisinin özdeğerleri λ_i , özvektörleri u_i olmak üzere,

$$V(\hat{\beta}) = \sigma^2 \sum_{i=1}^p \lambda_i^{-1} u_i' u_i$$

biçiminde ifade edilir.

$V(\hat{\beta}) = \Sigma = \sigma^2 (X' X)^{-1}$ olarak tanımlanırsa $\hat{\beta} \sim N(\beta, \Sigma)$ biçiminde yazılır. β 'dan $\hat{\beta}$ 'ya uzaklığın karesi L_1^2 olarak tanımlanır ve aşağıdaki gibi bulunur :

$$L_1^2 = (\hat{\beta} - \beta)' (\hat{\beta} - \beta)$$

β 'dan $\hat{\beta}$ 'ya olan uzaklığın moment çıkaran fonksiyonu t bir skaler olmak üzere,

$$\begin{aligned} M_{L_1}(t) &= E[\exp t' L_1] \\ &= \int \int \frac{1}{2\pi^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\hat{\beta} - \beta)' \Sigma^{-1} (\hat{\beta} - \beta) \right. \\ &\quad \left. + (\hat{\beta} - \beta)' t (\hat{\beta} - \beta) \right] d\beta \\ &= \int \int \frac{1}{2\pi^{p/2} |\Sigma|^{1/2}} \exp \{ -\frac{1}{2} (\hat{\beta} - \beta)' [\Sigma^{-1} - 2tI] (\hat{\beta} - \beta) \} d\beta \\ &= \frac{1}{2\pi^{p/2} |\Sigma|^{1/2}} (2\pi)^{L/2} |\Sigma^{-1} - 2tI|^{-1/2} \\ &= |\Sigma|^{-1/2} |\Sigma^{-1} - 2tI|^{-1/2} \\ &= |(\Sigma)(\Sigma^{-1} - 2tI)|^{-1/2} \\ &= |I - 2t\Sigma|^{-1/2} \end{aligned}$$

olarak bulunur.

$$|\Sigma - \theta I| = 0 \text{ ise } \theta, \Sigma \text{'nin özdeğerleridir.}$$

O halde,

$$|I - 2t\Sigma - \lambda I| = 0 \text{ ise } \lambda, (I - 2t\Sigma) \text{'nin özdeğerleridir.}$$

$$|-2t\Sigma + (1 - \lambda)I| = 0$$

$$|-2t[(\Sigma - (1 - \lambda I)/2t)]| = 0 \quad t \neq 0 \text{ olmak üzere}$$

$$(-2t)^p \left| \Sigma - \frac{1-\lambda}{2t} I \right| = 0 \text{ ise}$$

$$\frac{1-\lambda}{2t}, \Sigma \text{'nin özdeğerleridir.}$$

$$\frac{1-\lambda}{2t} = \theta \text{ 'dır. Buradan } \lambda = 1 - 2\theta t \text{ olarak yazılır. Bir matrisin determinatı o}$$

matrisin özdeğerleri çarpımı olduğundan L_1 'in moment çıkaran fonksiyonu;

$$ML_1(t) = \pi \prod_{i=1}^p (1 - 2\theta_i t)^{-1/2} \text{ şeklinde ifade edilir.}$$

$$= (1 - 2\theta_1 t)^{-1/2} (1 - 2\theta_2 t)^{-1/2} \dots (1 - 2\theta_p t)^{-1/2}$$

$$\begin{aligned} \frac{\partial ML_1(t)}{\partial t} &= (-\frac{1}{2}) (1 - 2\theta_1 t)^{-1/2-1} (-2\theta_1) (1 - 2\theta_2 t)^{-1/2} \dots (1 - 2\theta_p t)^{-1/2} \\ &+ (-\frac{1}{2}) (1 - 2\theta_2 t)^{-1/2-1} (-2\theta_2) (1 - 2\theta_1 t)^{-1/2} (1 - 2\theta_3 t)^{-1/2} \\ &\dots (1 - 2\theta_p t)^{-1/2} + \\ &\dots \\ &+ (-\frac{1}{2}) (1 - 2\theta_p t)^{-1/2-1} (-2\theta_p) (1 - 2\theta_1 t)^{-1/2} (1 - 2\theta_2 t)^{-1/2} \\ &\dots (1 - 2\theta_{p-1} t)^{-1/2} \\ &= (-\frac{1}{2}) (1 - 2\theta_1 t)^{-1} (-2\theta_1) (1 - 2\theta_1 t)^{-1/2} (1 - 2\theta_2 t)^{-1/2} \\ &\dots (1 - 2\theta_p t)^{-1/2} + (-\frac{1}{2}) (1 - 2\theta_2 t)^{-1} (-2\theta_2) (1 - 2\theta_1 t)^{-1/2} \\ &(1 - 2\theta_2 t)^{-1/2} \dots (1 - 2\theta_p t)^{-1/2} + \\ &\dots \\ &+ (-\frac{1}{2}) (1 - 2\theta_p t)^{-1} (-2\theta_p) (1 - 2\theta_1 t)^{-1/2} (1 - 2\theta_2 t)^{-1/2} \\ &\dots (1 - 2\theta_p t)^{-1/2} \\ &= (-\frac{1}{2}) (1 - 2\theta_1 t)^{-1} (-2\theta_1) ML_1(t) + (-\frac{1}{2}) (1 - 2\theta_2 t)^{-1} \\ &(-2\theta_2) ML_1(t) + \dots + (-\frac{1}{2}) (1 - 2\theta_p t)^{-1} (-2\theta_p) ML_1(t) \\ &= ML_1(t) \left[\sum_{i=1}^p \theta_i (1 - 2\theta_i t)^{-1} \right] \end{aligned}$$

$$\left. \begin{array}{c} \frac{\partial ML_1(t)}{\partial t} \\ \hline \end{array} \right|_{t=0} = (1) \left[\sum_{i=1}^p \theta_i (1)^{-1} \right]$$

$$\left. \begin{array}{c} \frac{\partial ML_1(t)}{\partial t} \\ \hline \end{array} \right|_{t=0} = \Sigma \theta_i = iz \Sigma$$

$$= \sigma^2 iz (X' X)^{-1}$$

$$= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

$$\begin{aligned} \frac{\partial^2 ML_1(t)}{\partial t^2} &= \frac{\partial ML_1(t)}{\partial t} \left[\sum_{i=1}^p \theta_i (1-2\theta_i t)^{-1} \right] \\ &\quad + ML_1(t) \left[\sum_{i=1}^p \theta_i (-1) (1-2\theta_i t)^{-2} (-2\theta_i) \right] \\ &= \frac{\partial ML_1(t)}{\partial t} \left[\sum_{i=1}^p \theta_i (1-2\theta_i t)^{-1} \right] \\ &\quad + ML_1(t) \left[\sum_{i=1}^p 2\theta_i^2 (1-2\theta_i t)^{-2} \right] \end{aligned}$$

$$\left. \begin{array}{c} \frac{\partial^2 ML_1(t)}{\partial t^2} \\ \hline \end{array} \right|_{t=0} = (iz \Sigma) \left(\sum_{i=1}^p \theta_i \right) + \left(\sum_{i=1}^p 2\theta_i^2 \right)$$

$$= (iz \Sigma)^2 + 2 \sum_{i=1}^p \theta_i^2$$

$$\left. \begin{array}{c} \frac{\partial^2 ML_1(t)}{\partial t^2} \\ \hline \end{array} \right|_{t=0} = (iz \Sigma)^2 + 2 iz \Sigma^2$$

$$\begin{aligned}
V(L_1) &= E(L_1^2) - [E(L_1)]^2 \\
&= (\text{iz } \Sigma)^2 + 2 \text{iz } \Sigma^2 - (\text{iz } \Sigma)^2 \\
&= 2 \text{iz } \Sigma^2 \\
&= 2 \text{iz } [\sigma^2 (X'X)^{-1}]^2
\end{aligned}$$

2.3. Çoklu Bağlantı Kavramı

$Y = X\beta + \varepsilon$ modelinde hipotez testleri için en çok olabilirliği en iyileme koşulu olarak aldığımızda β parametre vektörünün en küçük kareler kestiricisi

$$\hat{\beta} = (X'X)^{-1} X'Y$$

olur.

Y 'nin gözlenen değerlerinin bir olasılık dağılımından gelmiş olması nedeniyle elde edilecek kestirimler en küçük kareler çözümün getirdiği sorunlarla karşı karşıya kalacaktır. $X'X$ matrisinin sağlıksız koşullu olması durumunda en küçük kareler çözümlerindeki kararsızlık kendini en küçük kareler kestiricilerinin varyanslarında gösterecektir. $X'X$ matrisinin determinantının sıfıra yaklaşması (sağlıksız koşullu olması) nedeniyle parametre tahminlerinin varyansları büyük olacaktır. En küçük kareler kestiricilerinin en küçük varyansa sahip olma özelliğine güvenilerek yapılacak olan istatistiksel yorumlar bizi yanıltmış olacaktır. $X'X$ matrisinin sağlıksız koşullu olması “Çoklu Bağlantı” olarak tanımlanan özelliğe bağlanmaktadır.

2.3.1. Çoklu bağlantı ölçütü

$X'X$ matrisinin sağlıksız koşullu olmasına neden veya sağlıksız koşulluğunun bir türü olan çoklu bağlantının bir ölçütü de koşul sayısıdır ve

$$\text{enb. } K(X'X) = \sqrt{\frac{\max \lambda_i}{\min \lambda_i}} \quad (2.4)$$

olarak ifade edilir.

Koşul sayısının 10'dan küçük olması durumunda verilerde çok az bir çoklu bağlantı vardır. Koşul sayısı 10 ile 30 arasında ise orta dereceden çoklu bağlantıyı belirlemektedir. Eğer koşul sayısı 30'dan büyük ise şiddetli bir çoklu bağlantının varlığını göstermektedir.

2.4. Yanlı Kestirim Kavramı

En küçük kareler kestirim yöntemi, lineer model varsayımlarının sağlanması durumunda yansız ve en küçük varyanslı β ve $E(Y)$ kestirimleri verir. Bağımsız değişkenler arasındaki çoklu bağlantı varlığında en küçük kareler kestirimleri ile yine yansız fakat çok büyük varyanslı kestirimler elde edilir. Bu kestirimlerin kullanımı yanlış bulgulara ve yanlış model kullanımına neden olur.

Çoklu bağlantının etkisini gidermede kullanılan yöntemlerden birisi de “Ridge Regresyon” yöntemidir. Ridge regresyon yöntemi yanlı kestirim yöntemlerinden biridir.

Yanlı kestirim yöntemlerinde genel amaç, en küçük kareler yönteminde parametre kestirimleri için büyük olan varyans alanını küçük bir yan karşılığında daraltmaktır (Hoerl ve Kennard, 1970a).

2.5. Ridge Regresyon

İlk kez Hoerl 1962'de $k > 0$ için $\hat{\beta}^*$ kestiricisini karesel yanıt fonksiyonlarının grafikleri ile matematiksel bir benzerlik göstermesi nedeniyle $\hat{\beta}^*$ 'ı “Ridge” kestiricisi olarak tanımlanmıştır (Hoerl, 1962).

Lineer regresyon modelinde en küçük kareler kestiricisi,

$$\hat{\beta} = (X'X)^{-1} X'Y$$

olarak ifade edilmişti.

Ridge regresyon yönteminde $X'X$ matrisi korelasyon formunda olmak üzere $X'X$ matrisinin köşegen öğelerine küçük k değerlerinin (genelde $0 \leq k \leq 1$) eklenmesiyle ridge kestiricileri elde edilir. Bağımsız değişkenler standartlaştırıldığında,

$$\hat{\beta}^* = (X'X + kI)^{-1} X'Y \quad (2.5)$$

olarak ifade edilir (Hoerl, 1962).

2.5.1. Ridge kestiricisinin en küçük kareler kestiricisi ile ilişkisi

En küçük kareler kestiricisi,

$$\hat{\beta} = (X'X)^{-1} X'Y$$

olarak verilmişti. Buradan

$$X'X \hat{\beta} = X'Y$$

olarak yazılabilir. Ridge kestiricisi,

$$\hat{\beta}^* = (X'X + kI)^{-1} X'Y$$

olarak verilmişti. $X'Y$ yerine eşiti yazıldığında buradan

$$\hat{\beta}^* = (X'X + kI)^{-1} X'X \hat{\beta}$$

elde edilir.

$X'X$ matrisinin tersinin tersi kendisi olduğundan,

$$\hat{\beta}^* = (X'X + kI)^{-1} [(X'X)^{-1}]^{-1} \hat{\beta}$$

yazılabilir. Her iki matris tekil olmadıklarından,

$$\hat{\beta}^* = [(X'X)^{-1} (X'X + kI)]^{-1} \hat{\beta}$$

yazılabilir. Buradan da,

$$\hat{\beta}^* = [(X'X)^{-1} X'X + k(X'X)^{-1}]^{-1} \hat{\beta}$$

yazılabilir. Gerekli işlemlerden sonra,

$$\hat{\beta}^* = [I + k(X'X)^{-1}]^{-1} \hat{\beta}$$

olur. $Z = [I + k(X'X)^{-1}]^{-1}$ olarak tanımlanırsa,

$$\hat{\beta}^* = Z \hat{\beta}$$

olarak yazılır. Bu eşitlik ridge kestiricisinin en küçük kareler kestiricisinin bir dönüşümü olduğunu göstermektedir.

Aynı zamanda $\hat{\beta}^*$ 'nin boyu $k \neq 0$ için $\hat{\beta}$ 'dan daha kısadır ve aşağıdaki şekilde gösterilir :

$$\hat{\beta}^{*'} \hat{\beta}^* < \hat{\beta}' \hat{\beta}$$

2.5.2. Z ve W matrislerinin özellikleri

Ridge kestiricisi $\hat{\beta}^* = (X'X + kI)^{-1} X'Y$ olarak verilmişti. W matrisini,

$$W = (X'X + kI)^{-1}$$

şeklinde tanımladığımızda,

$$\hat{\beta}^* = W X'Y$$

olur. Z matrisini,

$$Z = [I + k(X'X)^{-1}]^{-1}$$

şeklinde tanımlamıştık. Buradan,

W matrisinin özdeğerleri,

$$\tau_i(w) = \frac{1}{\lambda_i + k}$$

dır.

Z matrisinin özdeğerleri aşağıdaki gibidir :

$$\tau_i(z) = \frac{\lambda_i}{\lambda_i + k}$$

Z matrisi aynı zamanda aşağıdaki eşitliklerle de verilebilir :

$$Z = I - k (X'X + kI)^{-1} = I - k W$$

2.5.3. Ridge kestiricisinin beklenen değeri

$$\begin{aligned} E(\hat{\beta}^*) &= E[(Z\hat{\beta})] \\ &= E[(I - kW)\hat{\beta}] \\ &= E[(I - k(X'X + kI)^{-1})\hat{\beta}] \\ &= E[\hat{\beta} - k(X'X + kI)^{-1}\hat{\beta}] \end{aligned}$$

u_i matrisin özdeğerlerine karşılık gelen özvektörler olmak üzere $\hat{\beta}^*$ beklenen değeri,

$$E(\hat{\beta}^*) = \beta - k \sum_{i=1}^p (\lambda_i + k)^{-1} u_i u_i' \beta$$

olarak yazılır.

2.5.4. Ridge kestiricisinin varyansı

$$V(\hat{\beta}^*) = V[(X'X + kI)^{-1} X'Y]$$

$$= \sigma^2 (X'X + kI)^{-1} X'X (X'X + kI)^{-1}$$

özdeğerler ve özvektörler cinsinden,

$$= \sigma^2 \sum_{i=1}^P \lambda_i (\lambda_i + k)^{-2} u_i u_i'$$

olarak yazılır.

2.6. Ridge Kestiricisi ile En Küçük Kareler Kestiricisinin Karşılaştırılması

En küçük kareler kestiricisinin beklenen değeri, $E(\hat{\beta}) = \beta$; ridge kestiricisinin beklenen değeri, $E(\hat{\beta}^*) = \beta - k \sum_{i=1}^P (\lambda_i + k)^{-1} u_i u_i' \beta$ olarak bulunmuştur.

$E(\hat{\beta}^*) - \beta = -k \sum_{i=1}^P (\lambda_i + k)^{-1} u_i u_i' \beta$ olarak yazıldığında ridge kestiricisi en küçük kareler kestiricisine göre $[-k \sum_{i=1}^P (\lambda_i + k)^{-1} u_i u_i' \beta]$ kadar yanlıdır.

En küçük kareler kestiricisinin varyansı, $V(\hat{\beta}) = \sigma^2 \sum_{i=1}^P \lambda_i^{-1} u_i u_i'$; ridge

kestiricisinin varyansı, $V(\hat{\beta}^*) = \sigma^2 \sum_{i=1}^P \lambda_i (\lambda_i + k)^{-2} u_i u_i'$ olarak bulunmuştur. $k = 0$

olması durumunda ridge kestiricisi en küçük kareler kestiricisine eşit olmaktadır.

$k \neq 0$ olmak üzere;

$$\lambda_i (\lambda_i + k)^{-2} < \lambda_i^{-1}$$

eşitsizliği sağlanır. O halde beklenen değerler ve varyanslar bakımından yapılan karşılaştırmada, ridge kestiricisi küçük bir yana karşı en küçük kareler kestiricisine göre daha küçük varyansa sahiptir.

$$\gamma_1(k) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2},$$

$$\gamma_2(k) = k^2 \beta' (X'X + kI)^{-2} \beta$$

olarak tanımlayalım.

$$E[L_1^2(k)] = \gamma_1(k) + \gamma_2(k)$$

olarak yazılır. $\gamma_1(k)$ toplam varyansı, $\gamma_2(k)$ 'da yan kareleri göstermektedir.

2.7.1. Ridge regresyonda hata kareler ortalaması fonksiyonu ile ilgili teoremler

Teorem 1 : Toplam varyans $\gamma_1(k)$ sürekli ve k 'nın monoton azalan bir fonksiyonudur (Hoerl ve Kennard, 1970b).

İspat : $k \rightarrow 0^+$ a ve $\lambda_p \rightarrow 0$ 'a giderken toplam varyans $\gamma_1(k)$ 'nın türevi $-\infty$ 'a yaklaşır.

$$\gamma_1(k) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2}$$

olarak verilmişti.

$$\begin{aligned} \frac{\partial \gamma_1(k)}{\partial k} &= \sigma^2 \sum_{i=1}^p \frac{[-2(\lambda_i + k)\lambda_i]}{(\lambda_i + k)^4} \\ &= -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} \\ \lim_{k \rightarrow 0^+} \frac{\partial \gamma_1(k)}{\partial k} &= \lim_{k \rightarrow 0^+} -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} \\ &= -2\sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i^2} \\ &= \lim_{\lambda_p \rightarrow 0} -2\sigma^2 \left(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \dots + \frac{1}{\lambda_p^2} \right) \end{aligned}$$

$$\frac{\partial \gamma_1(k)}{\partial k} = -\infty$$

Teorem 2 : Yan kareler $\gamma_2(k)$ sürekli ve k 'nın monoton artan bir fonksiyonudur (Hoerl ve Kennard, 1970b).

$$\text{İspat : } \gamma_2(k) = k^2 \beta' (X'X + kI)^{-2} \beta$$

olarak verilmişti.

$X'X$ 'in özdeğerler matrisi Λ ve özvektörler matrisi P olarak tanımlansın. Buradan $X'X = P' \Lambda P$ yazılır. $\alpha = P \beta$ olmak üzere;

$$\gamma_2(k) = k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}$$

olur. Bütün i 'ler için $\lambda_i > 0$ ve $k \geq 0$ olduğunda her bir element; $(\lambda_i + k)$ pozitiftir.

$\gamma_2(0) = 0$ 'dır. O halde $k \geq 0$ için $\gamma_2(k)$, k 'nın sürekli bir fonksiyonudur.

$\gamma_2(k)$ 'yı aşağıdaki şekilde de yazabiliriz :

$$\gamma_2(k) = \sum_{i=1}^p \frac{\alpha_i^2}{(1 + \frac{\lambda_i}{k})^2}$$

Bütün i 'ler için $\lambda_i > 0$ olduğunda; λ_i / k , k 'nın artışı ile monoton olarak azalır ve $\gamma_2(k)$ 'nın her bir terimi monoton olarak artar.

Sonuç : Yan kareler $\gamma_2(k)$; β' , β 'nın üst limitine yaklaşmaktadır (Hoerl ve Kennard, 1970b).

$$\text{İspat : } \gamma_2(k) = k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}$$

$$\lim_{k \rightarrow \infty} \gamma_2(k) = \lim_{k \rightarrow \infty} k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}$$

$$\begin{aligned}
&= \sum_{i=1}^p \alpha_i^2 \\
&= \alpha' \alpha \\
&= \beta' P' P \beta \\
&= \beta' \beta
\end{aligned}$$

Sonuç : $k \rightarrow 0^+$ 'a giderken $\gamma_2(k)$ 'nin türevi sıfırdır.

$$\gamma_2(k) = \sum_{i=1}^p \frac{k^2 \alpha_i^2}{(\lambda_i + k)^2}$$

$$\begin{aligned}
\frac{\partial \gamma_2(k)}{\partial k} &= \sum_{i=1}^p \frac{2k \alpha_i^2 + (\lambda_i + k)^2 - 2(\lambda_i + k) k^2 \alpha_i^2}{(\lambda_i + k)^4} \\
&= \frac{(\lambda_i + k) [2k(\lambda_i + k) \alpha_i^2 - 2k^2 \alpha_i^2]}{(\lambda_i + k)^4} \\
&= \frac{2k \lambda_i \alpha_i^2}{(\lambda_i + k)^3}
\end{aligned}$$

$\gamma_2(k)$ 'nin türevinin her bir terimi k 'nın sürekli fonksiyonudur.

$k \rightarrow 0^+$ 'a giderken her bir terimin limiti sıfırdır.

Teorem 3 (Varlık Teoremi) : Her bir $k > 0$ değeri için ;

$$E[L_1^2(k)] < E[L_1^2(0)] = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

olur (Hoerl ve Kennard, 1970b).

$$\begin{aligned}
\frac{\partial E[L_1^2(k)]}{\partial k} &= \frac{\partial \gamma_1(k)}{\partial k} + \frac{\partial \gamma_2(k)}{\partial k} \\
&= -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3}
\end{aligned}$$

Burada

$$\gamma_1(0) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i^2} \text{ ve } \gamma_2(0) = 0$$

dır.

$\gamma_1(k)$ 'nin birinci türevi pozitif olmayan ve $\gamma_2(k)$ 'nin birinci türevi negatif olmayandır.

$$\frac{\partial E[L_i^2(k)]}{\partial k} < 0$$

bu sonucu $k > 0$ olduğunu göstermek için kullanalım.

$$-2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3} < 0$$

$$2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3} < 2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3}$$

$$k < \frac{\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3}}{\sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3}}$$

$$k < \frac{\sigma^2 \frac{\lambda_1}{(\lambda_1 + k)^3} + \frac{\lambda_2}{(\lambda_2 + k)^3} + \dots + \frac{\lambda_p}{(\lambda_p + k)^3}}{\frac{\lambda_1 \alpha_1^2}{(\lambda_1 + k)^3} + \frac{\lambda_2 \alpha_2^2}{(\lambda_2 + k)^3} + \dots + \frac{\lambda_p \alpha_p^2}{(\lambda_p + k)^3}}$$

α_i^2 yerine $\alpha_{\max}^2$ alınır ise,

$$k < \frac{\sigma^2 \left[\frac{\lambda_1}{(\lambda_1 + k)^3} + \frac{\lambda_2}{(\lambda_2 + k)^3} + \dots + \frac{\lambda_p}{(\lambda_p + k)^3} \right]}{\alpha_{\max}^2 \left[\frac{\lambda_1}{(\lambda_1 + k)^3} + \frac{\lambda_2}{(\lambda_2 + k)^3} + \dots + \frac{\lambda_p}{(\lambda_p + k)^3} \right]}$$

$$k < \frac{\sigma^2}{\alpha_{\max}^2}$$

olur.

2.8. Çoklu Bağlantının Ridge Kestiricisi Üzerindeki Etkileri

En küçük kareler kestiricisini hesaplamada kullanılan $X'X$ matrisinin özdeğerleri ($\lambda_{\max} = \lambda_1 > \lambda_2 > \dots > \lambda_p = \lambda_{\min}$) ile gösterilsin.

$$|X'X - \lambda I| = 0$$

yardımıyla bulunurlar. Ridge kestiricisinde ise köşegenlere eklenen küçük k değerlerinden ötürü $X'X$ matrisinin özdeğerleri,

$$|X'X + kI - \theta I| = 0$$

$$|X'X + (k - \theta)I| = 0$$

$$|X'X - (\theta - k)I| = 0$$

olarak yazıldığında $(\theta - k)$ 'dır.

Eşitliklerden yararlanarak koşul sayılarını yazdığımızda,

$$K(X'X + kI) = \sqrt{\frac{\lambda_1 + k}{\lambda_p + k}} < \sqrt{\frac{\lambda_1}{\lambda_p}} = K(X'X); k \neq 0$$

DUŞÜNMEK İSTİYEN MERKEZİ

görüreceği gibi verilerde çoklu bağlantı olması durumunda ridge kestiricisinin koşul sayısı, en küçük kareler kestiricisinin koşul sayısından daha küçüktür. Bundan dolayı ridge kestirimleri, en küçük kareler kestirimlerinden daha kararlı olacaktır.

2.9. Ridge Regresyonda k Parametresinin Belirlenmesi

Bu bölümde k parametresinin belirlenmesinde kullanılan yöntemlerden bir kaçını inceleyeceğiz.

k parametresi bilinmeyen hata varyansı ile, bilinmeyen parametre vektörüne bağlıdır. $\hat{\beta}^*$ 'ya ilişkin hata kareler ortalaması ise iki kısma ayrıldığında; toplam varyans k'nın monoton azalan bir fonksiyonu, yan kareler ise k'nın monoton artan bir fonksiyonudur. Amacımız $\hat{\beta}^*$ için hata kareler ortalamasını minimum yapacak k parametresini belirlemektir.

2.9.1. Ridge izi yöntemi

Ridge izi k için bir çözüm elde edilmesinde kullanılan grafiksel yöntemdir. Regresyon katsayıları $\hat{\beta}^*$ 'lar düşey eksen, k değerleri yatay eksen olmak üzere iki boyutlu uzayda grafik çizilir. $X'X$ korelasyon matrisi çok sayıda büyük korelasyon içerdiğinde, basit korelasyon katsayılarına bakarak açıklayıcı değişkenler arasındaki ilişkiyi ortaya çıkarmak güçtür. k'nın değişik değerlerine göre $\hat{\beta}^*$ katsayıları hesaplanarak çizilen grafikte, hangi katsayıların verilere göre hassas olduğu, yani çok değiştiği belirlenebilir. Ridge regresyon aynı zamanda katsayıların duyarlılık analizinde de kullanılabilir. Katsayıların toplam varyansı k'nın azalan bir fonksiyonu, yan kareler ise k'nın artan bir fonksiyonudur. Böylece k artarken $\hat{\beta}^*$ katsayılarının hata kareler ortalaması minimuma gider ve daha sonra artar. Amacımız ridge kestiricisi $\hat{\beta}^*$ için, en küçük kareler kestiricisi $\hat{\beta}$ 'dan daha küçük hata kareler

ortalaması veren k parametresini bulmak ve k parametresine karşılık gelen “kararlı” $\hat{\beta}^*$ katsayılarının setini oluşturmaktır. Kararlılıktaki amaç verilerdeki küçük değişikliklere karşılık katsayıların hassas olmamasıdır.

Bağımsız değişkenler arasında yüksek ilişki var ise katsayılar k ’nın küçük artımlarında çok hızlı değişecektir. k ’nın büyük değerlerinde ise katsayılar belli bir kararlılık gösterecektir. Katsayıların kararlılık gösterdiği k değeri katsayıların istenen setini verecektir.

Eğer değişkenler ortogonal sayılabilirse k ’nın artımlarında $\hat{\beta}$ katsayıları çok az değişecektir. Bu durumda en küçük kareler çözümü katsayıların en iyi setini oluşturacaktır.

2.9.2. Tekrarlayıcı kestirim yöntemi

Arthur E.Hoerl ve Robert W.Kennad standartlaştırılmış $\hat{\beta}$ ’lar olmak üzere k için tek iterasyon kestirimini;

$$k_a = \frac{p \sigma^2}{\hat{\beta}' \hat{\beta}} \quad (2.6)$$

olarak göstermişlerdir (Hoerl ve Kennard, 1976). $\hat{\beta}' \hat{\beta}$, $\beta' \beta$ ’dan daha büyük olmaya başlarken koşul daha kötüye gider ve $X'X$ ’in koşulunu da daha zayıflatır. Bu etki k_a ’yı çok küçük yapabilir. Böylece $[\hat{\beta}^*(k_a)]' [\hat{\beta}^*(k_a)]$, $\hat{\beta}' \hat{\beta}$ ’dan daha küçük olmaya başlar. Bu durumda “ $\hat{\beta}$ ’nın yerine $\hat{\beta}^*(k_a)$ kullanılırsa k ’nın kestirimi iyileşir mi?” sorusu ortaya çıkacaktır.

k ’nın tekrarlayıcı kestirimi için β ve k ’nın bir dizisini düşünelim. Yani,

$$\beta, \quad k_{a0} = \frac{p \hat{\sigma}^2}{\hat{\beta}' \hat{\beta}}, \quad \hat{\beta}^*(k_{a0})$$

$$k_{a1} = \frac{p \hat{\sigma}^2}{[\hat{\beta}^*(k_{a0})]' [\hat{\beta}^*(k_{a0})]}, \quad \hat{\beta}^*(k_{a1})$$

$$k_{a2} = \frac{p \hat{\sigma}^2}{[\hat{\beta}^*(k_{a1})]' [\hat{\beta}^*(k_{a1})]}, \dots, \hat{\beta}^*(k_{at})$$

olsun. Algoritmik kestirim elde edebilmek için diziyi sonuçlandıracak bir yöntem olarak;

$$(k_{a,i+1} - k_{a,i}) / k_{a,i} > \delta$$

ise kestirime devam edilecek, değil ise sonuçlandırılarak $\hat{\beta}^*(k_{a,i})$ kestirim seti olarak kullanılır. İlk önce $k_{a,i+1} > k_{a,i}$ ve dolayısıyla $\delta \geq 0$ 'dır.

İkinci olarak sonuçlandırma kuralı regresyon vektörünün karesel uzunluğu cinsinden,

$$[\hat{\beta}^*(k_{a,i})]' [\hat{\beta}^*(k_{a,i})] < \left(\frac{1}{1+\delta} \right) [\hat{\beta}^*(k_{a,i-1})]' [\hat{\beta}^*(k_{a,i-1})]$$

olarak ifade edilir. Bu durumda yukarıdaki eşitsizlik sağlanmıyor ise işlem sonuçlandırılır ve $\hat{\beta}^*(k_{a,i})$ kestirim seti kullanılır. $\left(\frac{1}{1+\delta} \right)$ miktarı 1'den küçüktür ve $X'X$ matrisinin sağlıksız koşullu olmasının derecesi arttıkça δ azalmalıdır.

Deneyisel yaklaşımla,

$$T = \frac{\text{iz}(X'X)^{-1}}{p} \quad (2.7)$$

olarak tanımlayalım. Ortogonal matris için $T = 1$ 'dir. Verilen bir p boyutunda $X'X$ matrisinin özdeğer yayılımı (genişleme) ile T değeri büyür. Birkaç simülasyon (benzetim) kestirim dizisinin sona erdirilmesi için,

$\delta = A T^{-b}$ $b > 0$ kullanılarak elde edilir. Özellikle $\delta = 20 T^{-1.30}$ değerinin δ hesaplamak için pratik bir fonksiyon olduğu gösterilmiştir (Hoerl, Kennard, 1976).

2.9.3. k 'nın nokta kestirimi yöntemi

Ridge regresyonda koşul sayısı,

$$k_2 (X'X + kI) = \sqrt{\frac{\lambda_{\max} + k}{\lambda_{\min} + k}} \quad (2.8)$$

olarak tanımlanmıştı. Koşul sayısının 10'dan küçük olması çoklu bağlantının olmadığını gösteriyordu.

Bu çalışmada,

$$\sqrt{\frac{\lambda_{\max} + k}{\lambda_{\min} + k}} \geq 10$$

eşitsizliğinden hareket ederek,

$$k \leq \frac{\lambda_{\max} - 100 \lambda_{\min}}{99}$$

olarak k değeri elde edildi ve buna k 'nın nokta kestirimi yöntemi adı verildi. Eşitlik olduğunda elde edilen k değeri çoklu bağlantının ölçütü olan koşul sayısını 10 yapan değerdir. k 'nın bir nokta kestirimini verir.

3. UYGULAMA

Ridge regresyonun işlerliğini göstermek amacıyla bir model üzerinde çalışılmıştır. Örnek problemimizde öğrencilerin sınavlardan aldığı notlar ile başarı notları tahmin edilmeye çalışıldı. Bu değişkenler arası ilişki aşağıdaki model denklemiyle verilsin.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

$X'X$ matrisi korelasyon formunda olmak üzere standartlaştırılmış model denklemi;

$$Y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

olarak yazılır.

Deneysel olarak oluşturulan modelimizde değişkenler aşağıdaki gibi tanımlanmıştır.

Bağımlı değişken (Y) : öğrencinin başarı notu

Bağımsız Değişkenler :

(X_1) : Öğrencinin 1.vize notu

(X_2) : Öğrencinin 2.vize notu

(X_3) : Öğrencinin final notu

(X_4) : Öğrencinin final notunun iki katı

Deneysel olarak oluşturduğumuz model denkleminde çoklu bağlantı varlığını yaratmak amacıyla (X_4) değişkeni (X_3) değişkeninin bir fonksiyonu olarak tanımlanmıştır.

Model denklemi “SPSS” paket programı yardımıyla çözülmüştür. Modeldeki veriler standartlaştırılarak en küçük kareler ve ridge regresyon çözümleri elde edilmiştir. En

küçük kareler yönteminde elde edilen $\hat{\beta}$ katsayılarının olması gerektiğinden büyük

olduğu ve bazı katsayıların ters işaretli olduğu (0.287894, 0.4649, 60.06090, -59.4939) gözlenmiştir. Çoklu bağlantıyı incelemek amacıyla;

$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{\min}$ olmak üzere özdeğerler sırasıyla,

3.30203 , 0.46138 , 0.23568 , 0.000001

olarak bulunmuştur.

Buradan koşul sayısı,

$$\begin{aligned} k(X'X) &= \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \\ &= \sqrt{\frac{3.30203}{0.000001}} \\ &= \sqrt{3302030} \\ &= 1817,148866 \end{aligned}$$

olarak hesaplanmıştır. Koşul sayısının 10'dan büyük olması nedeniyle çoklu bağlantı olduğuna karar verilmiştir.

En küçük kareler için hata kareler ortalamasının beklenen değeri,

$$\begin{aligned} E[L_1^2] &= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \\ &= 1000006.715 \sigma^2 \end{aligned}$$

olarak hesaplanmıştır. Verilerimiz ortogonal bir sistemin özelliğini göstermiş olsaydı. Özdeğerler birbirine çok yakın olacak ve hata kareler ortalamasının beklenen değeri; $E[L^2] = 4 \sigma^2$ olacaktı. Buna göre verilerimiz ortogonal bir sisteme göre 250000 kat daha geniş bir alana yayılmıştır.

Örneğimizde $k = 0$ 'dan başlayarak k 'nın artan değerlerine karşılık $\hat{\beta}^*$ katsayıları hesaplanarak (Tablo 1) verilmiştir. Ridge izi grafiği (Şekil-1)'de çizilmiştir. k 'nın artan değerlerine karşılık R^2 grafiği (Şekil-2)'de verilmiştir.

Örneğimizde k 'nın nokta kestirimi yöntemine göre çözümü;

$$k \leq \frac{3.30293 - 100(0.000001)}{99}$$

$$k \leq 0.033361919$$

olarak bulunmuştur.



4. SONUÇ VE ÖNERİLER

Örneğimizde en küçük kareler yöntemi ile tahmin edilen $\hat{\beta}$ katsayıları (0.287894, 0.4649, 60.06090, -59.4939) incelendiğinde katsayıların olduğundan daha büyük ve bazı katsayıların ters işaretli olduğu gözlemlenmiştir. Buna karşın küçük bir yana karşılık Ridge Regresyon yöntemi ile elde edilen $\hat{\beta}^*$ katsayıları olduğundan büyük olması ve ters işaretli durumu düzelmiştir. Artık kareler toplamının k artarken artması çok büyük bir sorun olarak ele alınmamalıdır; çünkü amacımız kestirim verilerine mümkün olan en yakın uydurma değildir. En küçük kareler yöntemine göre $\hat{\beta}^*$ için daha küçük hata kareler ortalaması veren, daha dar bir bölgeye yayılan ve gelecekteki gözlemlerin kestirimini iyi yapabilecek “sabit” bir katsayı seti oluşturmaktır. En küçük kareler kestirimleri ile ridge kestirimleri hata kareler ortalaması bakımından karşılaştırılması amaçlanmıştır. Örneğimizde k için; ridge izinden ve önerdiğimiz k'nın nokta kestirimi yönteminden yaklaşık, $k = 0.03336$ olarak bulunmuştur. Ridge izi grafiği (Şekil-1) incelendiğinde de k'nın yaklaşık olarak 0.04 değerinde $\hat{\beta}^*$ katsayılarının durağanlaştığı gözlemlenmiştir. En küçük kareler için $E(L_1^2) = 1000006.715 \sigma^2$ idi. Ridge Regresyon için $E[L_1^2(k)] = 12.503 \sigma^2$ olarak hesaplanmıştır. Görüleceği gibi nokta kestiriminden elde edilen $k=0,03336$ değeri için hata kareler ortalamasının beklenen değeri ortogonal sisteme göre 3 kat daha geniş bir alana yayılmıştır.

Eğer bağımsız değişkenler arasında çoklu bağlantı var ise, çoklu bağlantının etkisini azaltmak ve güvenilir katsayılar setini oluşturmak için ridge regresyon yöntemi önerilebilir.

KAYNAKLAR

Hoerl, A.E and Kennard, R.W. (1970a). "Ridge Regression : Biased Estimation for Nonorthogonal Problems," **Technometrics**, 12, 55-67.

Hoerl, A.E. and Kennard, R.W. (1970b). "Ridge Regression : Applications to Nonorthogonal Problems", **Technometrics** 12, 69-82.

Hoerl, A.E. and Kennard, R.W. and Baldwin, K.F (1975). "Ridge Regression : Some Simulations", **Communications in Statistics**, 4, 105-123.

Hoerl, A.E. (1962) "Application of Ridge Analysis to Regression Problems", **Chemical Engineering Progress**. 58, 54-59.

Hoerl, A.E., and Kennard R.W. (1976) "Ridge Regression : Iterative Estimation of the Baising Parameter." **Communications in Statistics**, 5, 77-88.



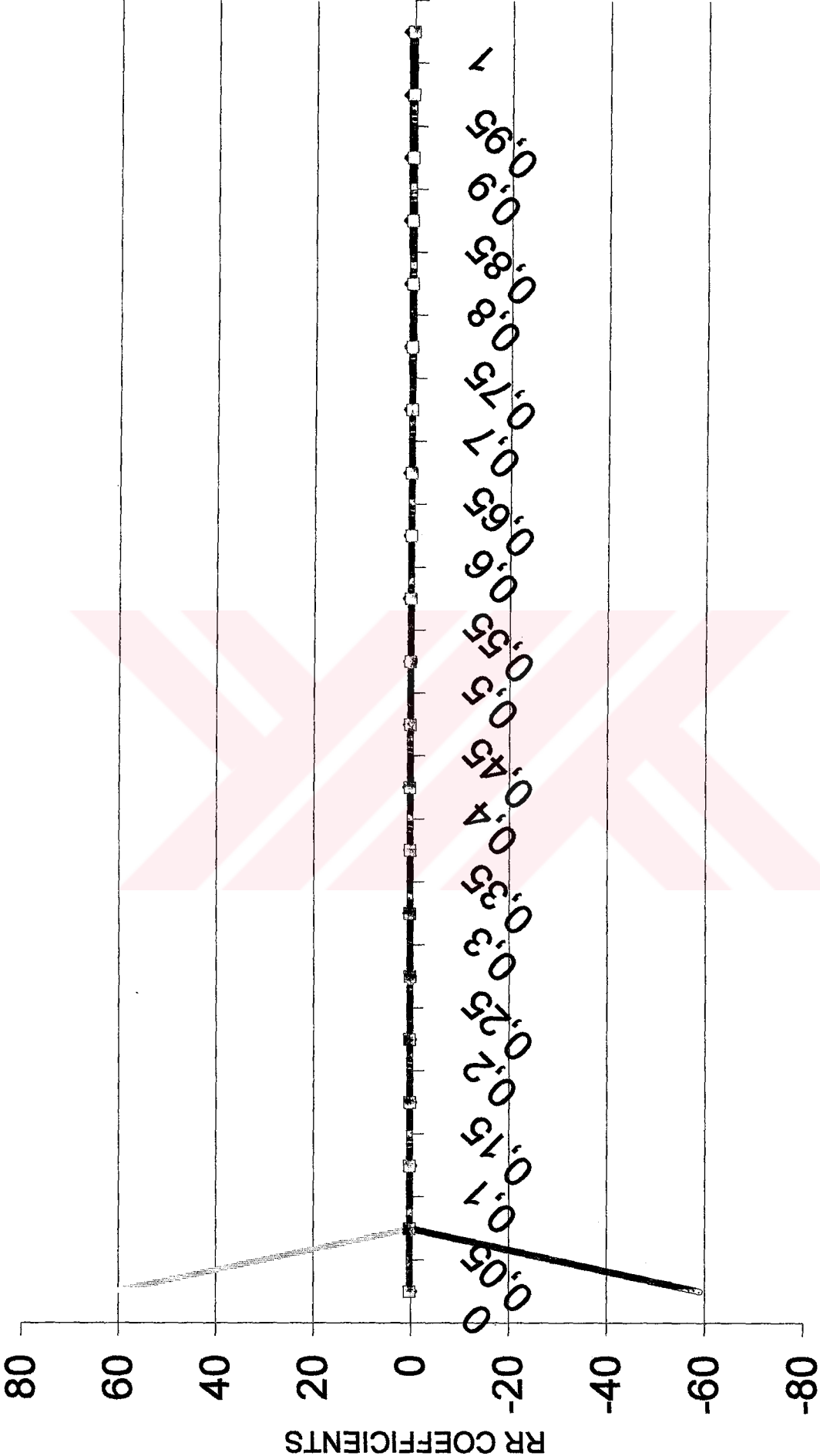
EKLER

Tablo-1

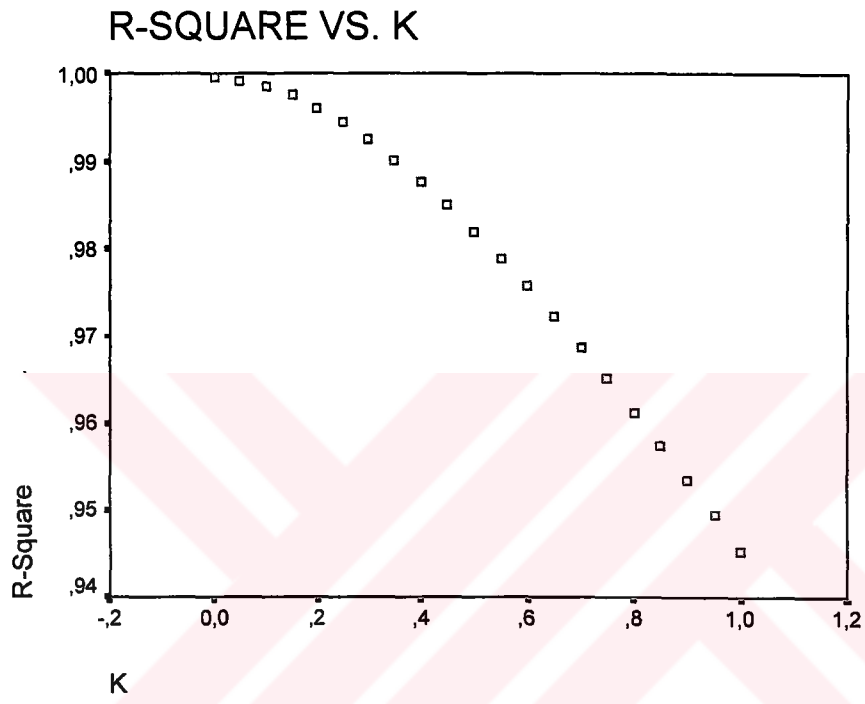
R-SQUARE AND BETA COEFFICIENTS FOR ESTIMATED VALUES OF K

K	RSQ	X1	X2	X3	X4
,00000	,99948	,287894	,244649	60,06090	-59,4939
,05000	,99922	,280935	,242738	,279689	,279676
,10000	,99856	,274215	,241303	,275797	,275790
,15000	,99753	,268328	,239357	,271977	,271973
,20000	,99614	,263028	,237110	,268236	,268233
,25000	,99444	,258166	,234685	,264579	,264577
,30000	,99245	,253645	,232156	,261006	,261004
,35000	,99019	,249401	,229570	,257518	,257516
,40000	,98769	,245387	,226961	,254113	,254112
,45000	,98498	,241571	,224349	,250790	,250789
,50000	,98206	,237927	,221749	,247548	,247547
,55000	,97897	,234434	,219171	,244385	,244384
,60000	,97571	,231077	,216623	,241297	,241296
,65000	,97230	,227843	,214110	,238283	,238283
,70000	,96875	,224723	,211635	,235342	,235341
,75000	,96509	,221706	,209201	,232470	,232469
,80000	,96131	,218785	,206809	,229665	,229664
,85000	,95743	,215954	,204460	,226925	,226925
,90000	,95347	,213207	,202154	,224249	,224249
,95000	,94943	,210539	,199892	,221634	,221634
1,0000	,94531	,207946	,197672	,219078	,219078

RIDGE TRACE



Şekil-2



Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Y	61,53	9,50	30
X1	60,13	13,99	30
X2	59,90	11,33	30
X3	62,33	8,98	30
X4	124,63	17,94	30

Correlations

	Y	X1	X2	X3	X4
Pearson Correlation	Y	X1	X2	X3	X4
	1,000	,879	,861	,945	,945
	,879	1,000	,758	,716	,715
	,861	,758	1,000	,702	,703
	,945	,716	,702	1,000	1,000
	,945	,715	,703	1,000	1,000
Sig. (1-tailed)	Y	X1	X2	X3	X4
	,000	,000	,000	,000	,000
	,000	,000	,000	,000	,000
	,000	,000	,000	,000	,000
	,000	,000	,000	,000	,000
N	Y	X1	X2	X3	X4
	30	30	30	30	30
	30	30	30	30	30
	30	30	30	30	30
	30	30	30	30	30
	30	30	30	30	30

Tablo-3

Y	X1	X2	X3	X4
64	65	58	65	130
65	60	54	70	140
58	45	62	60	120
59	55	60	60	120
73	72	65	75	150
82	80	75	85	170
58	65	60	55	110
63	70	65	60	120
66	72	75	60	120
81	80	85	80	160
53	45	55	55	110
51	40	35	60	120
56	50	64	55	110
62	54	60	65	130
61	65	60	60	120
62	70	60	60	120
67	75	65	65	130
57	48	55	60	120
61	59	50	65	130
61	65	60	60	120
72	72	78	70	140
75	70	80	75	150
45	30	45	50	100
36	25	35	40	80
52	58	50	50	100
56	44	56	60	120
61	64	60	60	120
65	72	60	65	130
61	70	55	60	120
63	64	55	65	130

ÖZGEÇMİŞ

Mücahit KURTULUŞ, 9.9.1964 tarihinde Ankara’da doğdu, İlk, Orta ve Lise eğitimini Ankara’da, Yüksek Öğrenimini ise Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde tamamladı.

