T.C.

DİCLE ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

LİNEER REGRESYONDA KÜÇÜLTME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Erdem KALKAN

YÜKSEK LİSANS TEZİ

MATEMATİK ANABİLİM DALI

DİYARBAKIR

Ocak-2020

TEŞEKKÜR

Bu uzun ve yorucu tez çalışmam süresince beni daima destekleyen bilgisiyle, tecrübesiyle yol gösteren değerli danışmanım Doç. Dr. Pakize TAYLAN'a yardım ve katkılarından dolayı teşekkürü borç bilirim.

İnançlarını ve sevgilerini benden eksik etmeyen anneme, babama, abime, Cevat'a ve bugünlere gelmemde emeği geçen tüm öğretmenlerime teşekkürlerimi sunarım.

Biricik halam Nafiye YAĞCI' nın değerli anısına…

İÇİNDEKİLER

Sayfa

TEŞEKKÜR I

İÇİNDEKİLER II

ÖZET IV

ABSTRACT V

ÇİZELGE LİSTESİ VI

ŞEKİL LİSTESİ VII

KISALTMA VE SİMGELER VIII

1. GİRİŞ 1

2. ÖNCEKİ ÇALIŞMALAR 3

3. MATERYAL ve METOT 5

3.1. Lineer Regresyon Modeli 5

3.1.1. Basit Doğrusal Regresyon Modeli 5

3.1.2. Çoklu Doğrusal Regresyon Modeli 5

3.1.2.1. Çoklu Doğrusal Regresyon Modelindeki Varsayımlar 6

3.1.3. Çoklu Doğrusal Regresyonda Parametre Tahmin Yöntemleri 7

3.1.3.1. Tahmin Edicilerde Aranan Özellikler 8

3.1.3.2. En Küçük Kareler Yöntemi (EKK) 8

3.1.3.3. En Çok Olabilirlik Yöntemi (MLE: Maksimum Likehood) 11

3.2. Lineer Regresyonda İç İlişki Problemi 13

3.2.1. İç İlişki Problemini Saptama Yöntemleri 14

3.2.1.1. Korelasyon Katsayısı 14

3.2.1.2. R^2 Belirleme Katsayısı 15

3.2.1.3. Varyans Şişirme Faktörü (VIF) 15

3.2.1.4. $X^T X$ Korelasyon Matrisinin Öz Değerlerinin İncelenmesi 16

3.2.1.5. $X^T X$ Korelasyon Matrisinin Determinantı 17

3.2.1.6. Tolerans Değeri 17

3.2.1.7. F ve t Testi Değeri 17

3.2.2. İç İlişki Problemi İçin Çözümler..... 18
3.3. Lineer Regresyonda Küçültme Yöntemleri..... 18
3.3.1. Ridge Regresyon..... 19
3.3.1.1. Ridge Ceza Parametresinin Seçimi..... 23
3.3.2. En Küçük Mutlak Küçültme ve Seçim Operatörü (LASSO)..... 24
3.3.2.1. Least Angle Regression (LARS)..... 25
3.3.2.2. Lasso Ceza Parametresinin Seçimi..... 26
3.3.3. Naïve Elastik Net (NEN)..... 28
3.3.4. Elastik Net (EN)..... 31
3.3.4.1. Least Angle Regression-EN (LARS-EN)..... 32
3.3.4.2. LARS-EN İçin Ceza Parametrelerinin Seçimi..... 33
3.3.5. Liu Küçültme Yöntemi..... 33
3.3.5.1. Liu Ceza Parametresinin Seçimi..... 35
4. ARAŞTIRMA BULGULARI..... 39
4.1. İç İlişkinin Saptanması..... 39
4.2. Modelin Tahmini..... 42
4.3. Ridge Regresyon İle Modelin Çözümü..... 46
4.3.1. Ridge İzi..... 47
4.3.2. Çapraz Değerlendirme Yöntemi..... 48
4.4. LASSO Küçültme Yöntemi İle Modelin Çözümü..... 50
4.5. Elastik Net İle Modelin Çözümü..... 54
4.6. Liu Küçültme Yöntemi İle Modelin Çözümü..... 57
5. TARTIŞMA VE SONUÇ..... 61
6. KAYNAKLAR..... 63
ÖZGEÇMİŞ..... 67

ÖZET

LİNEER REGRESYONDA KÜÇÜLTME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

ERDEM KALKAN

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MATEMATİK ANABİLİM DALI

2020

İstatistiksel araştırmalar ve analizlerde veri kümesine dayalı olarak elde edilen regresyon modelleri çok önemli rol oynar. Bu modellerden biride çoklu doğrusal regresyon modelidir. Ancak çoklu doğrusal regresyon analizinde, iç ilişki problemi ortaya çıktığında en küçük kareler tahmin edicileri yansız olmasına rağmen büyük varyanslı olarak elde edilirler. Bu durum yapılan istatistiksel analizin doğruluğunu olumsuz yönde etkilemektedir. İstatistiksel araştırmalarda küçültme yöntemlerinin birinin kullanılması ile tahmin edicilerin varyansı küçültülerek iç ilişki problem ortadan kalkabilmektedir. Bu tez çalışmasında iç ilişki problemi ele alınarak, iç ilişkinin giderilmesi için Ridge Regresyon, LASSO, Elastik Net ve Liu gibi küçültme yöntemleri incelenmiştir.Uygulama çalışması olarak da, iç ilişkinin saptandığı veri üzerinden adı geçen küçültme yöntemlerinin hata kareler toplamı, hata kareler ortalamaları ve R^2 belirleme katsayıları karşılaştırılmıştır. Bu karşılaştırmalarda yukarıdaki küçültme yöntemlerinin tamamı en küçük karelere göre daha iyi sonuçlar vermekle birlikte, LASSO hem en küçük karelere göre hem de ele alınan bütün küçültme yöntemlerine göre daha iyi sonuçlar vermiştir. Sonuç olarak iç ilişki problemin varlığı durumunda en küçük kareler yöntemi yerine küçültme yöntemlerinin, özelliklede LASSO nun kullanılmasının daha doğru sonuçlar verdiği kararına varılmıştır.

Anahtar kelimeler: İç ilişki problemi, Ridge Regresyon, LASSO, Elastik Net, Liu Regresyon.

ABSTRACT

A COMPARISON OF SHRINKAGE METHODS IN LINEAR REGRESSION

MSc THESIS

ERDEM KALKAN

DEPARTMENT OF MATHEMATICS
INSTITUTE OF NATUREL AND APPLIED SCIENCES
UNIVERSITY OF DICLE

2020

Regression models based on data set play a very important role in statistical research and analysis. One of these models is multiple linear regression model. However, in multiple linear regression analysis, when multicolinearity problem arises, the least squares estimators are obtained with large variance although they are unbiased. This situation negatively affects the accuracy of statistical analysis. In statistical research, by using one of the shrinkage methods, the variance of the estimators can be reduced and the multicolinearity problem can be eliminated. In this thesis, colinearity problem is investigated and Ridge Regression, LASSO, Elastic Net and Liu shrinkage methods are examined to eliminate multicolinearity problem. As an application study, sum of squares errors, mean squares erros mean and determination coefficients of shrikage methods were compared over the data in which colinearity was determined. In these comparisons, all of the above shrinkage methods yielded better results than the least squares, while the LASSO yielded better results than both the least squares and all the shrinkage methods discussed for this data set. As a result, it is concluded that the shrinkage methods instead of least squares method, especially LASSO, give more accurate statistical results in case of multicolinearity problem.

Keywords: Multicollinearity, Ridge Regression, LASSO, Elastic Net, Liu Regression.

ÇİZELGE LİSTESİ

Çizelge No		Sayfa
Çizelge 4.1.	2017 yılında yaşam maliyetini etkileyen etkenler	39
Çizelge 4.2.	VIF değerleri	41
Çizelge 4.3.	Modelin EKK sonuçları	43
Çizelge 4.4.	Modelin varyans analiz tablosu	43
Çizelge 4.5.	Eğitim (train) veri kümesi için EKK sonuçları	45
Çizelge 4.6.	Deneme (test) veri kümesi İçin EKK sonuçları	45
Çizelge 4.7.	Bazı K değerlerine ait ridge parametre tahminleri	47
Çizelge 4.8.	K_{best} değeri için ridge regresyon katsayıları	49
Çizelge 4.9.	Eğitim veri kümesi için ridge regresyon sonuçları	50
Çizelge 4.10.	Deneme veri kümesi için ridge regresyon sonuçları	50
Çizelge 4.11.	Bazı ψ değerlerine ait LASSO parametre tahminleri	51
Çizelge 4.12.	ψ_{best} değeri için LASSO parametre tahminleri	53
Çizelge 4.13.	Eğitim veri kümesi için LASSO sonuçları	53
Çizelge 4.14.	Deneme veri kümesi için LASSO sonuçları	53
Çizelge 4.15.	Farklı δ_2 değerleri için Elastik net parametre tahminleri	54
Çizelge 4.16.	Elastik net parametre tahminleri	56
Çizelge 4.17.	Eğitim veri kümesi için Elastik net sonuçları	57
Çizelge 4.18.	Deneme veri kümesi için Elastik net sonuçları	57
Çizelge 4.19.	Bazı d değerleri için Liu parametre tahminleri	58
Çizelge 4.20.	Liu parametre tahminleri	59
Çizelge 4.21.	Eğitim veri kümesi için Liu sonuçları	59
Çizelge 4.22.	Deneme veri kümesi için Liu sonuçları	59
Çizelge 5.1.	Eğitim veri kümesi için genel sonuçlar	61
Çizelge 5.2.	Deneme veri kümesi için genel sonuçlar	61

ŞEKİL LİSTESİ

Şekil No		Sayfa
Şekil 3.1.	LASSO ve Ridge regresyona ait tahminlerin grafiksel gösterimi	25
Şekil 4.1.	Değişkenlerin matris diyagramları	44
Şekil 4.2.	Artıkların grafikleri	45
Şekil 4.3.	Ridge izi	48
Şekil 4.4.	K değerine göre Ridge parametre tahminlerinin değişimi	48
Şekil 4.5.	K ceza parametresine göre MSE değişimi	49
Şekil 4.6.	ψ değerine göre LASSO parametre tahminlerinin değişimi	51
Şekil 4.7.	ψ ceza parametresine göre MSE değişimi	52
Şekil 4.8.	δ_2 değerine göre Elastik net parametre tahminlerinin değişimi	55
Şekil 4.9.	δ_2 ceza parametresine göre MSE'nin değişimi	56
Şekil 4.10.	d ceza parametresine göre minimum MSE' nin yeri	58

KISALTMA VE SİMGELER

EKK : En küçük kareler

k : Bağımsız değişken sayısı

n : Gözlem sayısı

ε : Rastgele hata vektörü

Y : nx1 boyutlu yanıt vektörü

X : $n \times (k+1)$ boyutlu açıklayıcı değişkenler matrisi

$\boldsymbol{\beta}$: $(k+1)$ boyutlu parametre vektörü

λ_k : $\boldsymbol{X}^T\boldsymbol{X}$ korelasyon matrisinin öz değerleri

KS : Koşul sayısı

KI : Koşul indeksi

VIF : Varyans şişirme faktörü

$Var(.)$: Varyans

Cov (.) : Kovaryans

RR : Ridge regresyon

LASSO : En küçük mutlak küçülme ve seçim operatörü

EN : Elastik Net

LR : Liu Regresyon

R^2 : Belirleme katsayısı

MSE : Hata kareler ortalaması

RSS : Hata kareler toplamı

$\hat{\boldsymbol{\beta}}$: En küçük kareler tahmin edicisi

$\hat{\boldsymbol{\beta}}_R$: Ridge tahmin edicisi

$\hat{\boldsymbol{\beta}}_{LASSO}$: LASSO tahmin edicisi

$\hat{\boldsymbol{\beta}}_{EN}$: Elastik Net tahmin edicisi

$\hat{\boldsymbol{\beta}}_{Liu}$: Liu tahmin edicisi

K :Ridge ceza parametresi

d : Liu ceza parametresi

ψ :LASSO ceza parametresi

LARS :Least angle regression

NEN : Naïve Elastik Net

1.GİRİŞ

Tıp, mühendislik, ekonomi, ziraat gibi birçok alanda kullanılan regresyon analizi, çok faktörlü verileri analiz etmek için kullanılan en yaygın tekniklerden biridir. Cazibesi ve kullanışlılığı, ilgilenilen bir yanıt değişken ve ilgili açıklayıcı değişkenler arasındaki ilişkiyi ifade etmek için denklem kullanabilme, oluşturulan denklemi anlayabilme, yorumlayabilme, değişkenlerden birinin değeri bilindiğinde diğeri hakkında tahminler yapabilmenin mantıksal işlemlerinden kaynaklanmaktadır. Bununla birlikte regresyon analizi, temel matematiğin inceliklerini ve iyi geliştirilmiş bir istatistiksel teoriyi kullanması nedeniyle teorik olarak da ilginçtir. Regresyonun başarılı bir şekilde kullanılması, gerçek dünya verilerinde kullanıldığında ortaya çıkan teori ve pratik problemlerin değerlendirmesini gerektirir. Regresyon analizinde değişkenler arasındaki ilişkiyi açıklayabilmek için oluşturulan modelde parametrelerin tahmin edilmesinde en küçük kareler yöntemi kullanılmaktadır. Ancak çoklu doğrusal regresyon analizinde en küçük kareler yönteminin kullanılabilmesi için bazı varsayımların sağlanması gerekmektedir. Bu varsayımlardan biri de bağımsız değişkenler arasında iç ilişki olmamasıdır (Neter ve ark. 1990). Bağımsız değişkenler arasında iç ilişki (multicollinearity) olması durumunda en küçük kareler yöntemiyle elde edilen sonuçlar yansız olsa bile varyansın büyük çıkmasından dolayı sonuçlar güvenilir olmayacaktır. Bu durumda varyansı azaltmak için yansızlıktan vazgeçip modeldeki parametrelerin tahmini için küçültme yöntemleri kullanılmaktadır (Webster 1975).

Bu tez çalışmasının ilk bölümünde basit doğrusal regresyon ve çoklu doğrusal regresyon verildikten sonra, parametreleri tahmin etmede kullanılan en küçük kareler yöntemi ve en çok olabilirlik yöntemleri tanıtılıp, tahmin edicilerde aranan özellikler sıralanmıştır.

Çalışmamızın ikinci bölümünde iç ilişki problemi tanıtılıp, iç ilişki probleminin nedenleri ve en küçük kareler yöntemi üzerindeki olumsuz etkileri açıklanmıştır. İç ilişki problemini saptamak için R^2 belirleme katsayısı, Varyans Şişirme Faktörü (VIF), X^TX korelasyon matrisinin öz değerlerinin İncelenmesi ve X^TX korelasyon matrisinin determinantı gibi yöntemler kullanılmıştır. İç ilişki probleminin çözümü için; ek veri toplama, iç ilişkiye neden olan değişken veya değişkenlerin modelden atılması,

aralarında iç ilişki olan bağımsız değişkenlerin tek bir değişken olarak kabul edilip modelin yeniden tanımlanması ve en küçük kareler yöntemine alternatif olarak kullanılabilecek küçültme yöntemleri verilmiştir.

Çalışmamızın üçüncü bölümünde iç ilişki problemi durumunda kullanılan Ridge Regresyon (RR) (Hoerl ve Kennard 1970a), En küçük Mutlak Küçülme ve Seçim Operatörü (LASSO) (Usai ve ark. 2012), Elastik Net (EN), Liu Regresyon(LR) (Akdeniz ve Kaçıranlar 1995), küçültme yöntemlerinin özellikleri ayrıntılı olarak verilip, her yöntem için parametre tahminlerinin nasıl yapılacağı açıklanmıştır.

Çalışmamızın dördüncü bölümünde, https://dasl.datadescription.com/ adresinden alınan değerler ile, 2017 yılına ait 511 şehrin yiyecek, konaklama masrafları ve tüketim malları fiyatlarına göre bir veri oluşturulup bu şehirlerin yaşam maliyeti hesaplanmıştır. Kurulan modelde iç ilişki problemi farklı yöntemlerle belirlenmiştir. İç ilişki probleminin en küçük kareler yöntemi üzerindeki olumsuz etkileri gösterildikten sonra modelimiz sırasıyla RR, LASSO, EN, LR küçültme yöntemleri kullanılarak çözülmüştür.

Tez çalışmamızın sonuç ve tartışma bölümünde; hesaplamalar sonucu en iyi parametre tahmin yönteminin seçilmesi için En Küçük Kareler (EKK), RR, LASSO, EN ve LR yöntemlerinden elde edilen hata kereler toplamı (RSS), hata kareler ortalaması (MSE) ve belirleme katsayısı R^2 karşılaştırma kriteri olarak kullanılmıştır.

2. ÖNCEKİ ÇALIŞMALAR

Bağımsız değişkenler arasında iç ilişki olması durumunda en küçük kareler yöntemiyle elde edilen sonuçlar hatalı olabilmektedir (Orhunbilge 2000). Bu nedenle EKK yöntemine alternatif olarak küçültme yöntemlerini kullanarak daha doğru sonuçlar elde etmek mümkündür (Draper ve Smith 1998).

İlk olarak Hoerl ve Kennard 1970 yılında EKK yöntemine alternatif bir yöntem olarak Ridge regresyonu önermiştir. Ridge regresyon, değişkenler arasındaki iç ilişki problemini çözerek EKK yöntemine göre daha küçük hatalar ile modeli tahmin etmektedir. Hoerl ve Kennard (1970a, 1970b), her zaman en küçük MSE'yi verecek bir $K>0$ ceza parametresinin olduğunu belirtmiştir. K ceza parametresinin belirlenmesi için çok sayıda çalışmalar yapılmıştır (Golub ve ark. 1979).

Liu (1993), Ridge regresyon ve Stein tahmin edicisinin (Stein,1956) olumlu özelliklerini bir araya getirerek yeni bir küçültme yöntemi önermiştir. Daha sonra Akdeniz ve Kaçıranlar (1995) yapmış oldukları çalışmada bu yöntemi Liu tahmin edicisi diye adlandırmıştır. Liu (1993), d ceza parametresinin (yanlılık katsayısı) 0 ile 1 arasında değer aldığını ifade etmiştir. Ancak Liu (2011) d ceza parametresinin 0'dan küçük veya 1'den büyük bir değer alabileceğini belirtmiştir. Burada asıl önemli nokta minimum MSE'nin hesaplanması için d ceza parametresinin belirlenmesidir (Akdeniz ve Kaçıranlar 1995).

Ridge regresyon ve Liu yöntemleri EKK yöntemine göre daha küçük hatalar verip çok daha doğru sonuçlar elde edilmesini sağlasa da hataların hiçbir zaman sıfıra eşit bulunamaması bu konu üzerinde çalışmaların devam etmesine neden olmuştur. Tibshirani (1996), Ridge regresyonundan farklı olarak, bazı büyük katsayı tahminlerini sıfıra eşitleyip modelin yorumlanmasını kolaylaştıran En küçük Mutlak Küçülme ve Seçim Operatörü (LASSO) adında bir yöntem geliştirmiştir.

Zou ve Hastie (2005) tarafından Ridge regresyon ile LASSO ya göre daha iyi sonuçlar veren Elastik Net yöntemini önermiştir.

3. MATERYAL VE METOT

3.1. Lineer Regresyon Modeli

Bu bölümünde basit doğrusal regresyon ve çoklu doğrusal regresyon verildikten sonra, parametreleri tahmin etmede kullanılan en küçük kareler yöntemi ve en çok olabilirlik yöntemleri tanıtılıp, tahmin edicilerde aranan özellikler sıralanmıştır.

3.1.1. Basit Doğrusal Regresyon Modeli

Bir bağımlı (Y) ve bir bağımsız (X) değişken arasındaki doğrusal ilişkiyi inceleyen model basit doğrusal regresyon olarak adlandırılır. Burada Y bağımlı değişkenine yanıt, X bağımsız değişkenine de açıklayıcı değişken denir. Regresyon analizinde amaç Y bağımlı değişkenini en iyi tahmin eden modelin bulunması ve hangi bağımsız değişkenden daha çok etkilendiğini bulmaktır.

$(X_i, Y_i)\ (i = 1, 2, ..., n)$, i. gözlem değerine ait yanıt ve açıklayıcı değişkeni göstermek üzere, bu gözlemlere ait basit doğrusal regresyon modeli

$$Y_i = \alpha + \beta X_i + \varepsilon_i \ (i = 1, 2, ..., n), \tag{3.1}$$

olarak tanımlanır. Burada α, X_i sıfır olduğunda regresyon doğrusunun Y_i eksenini kestiği değer, β, bilinmeyen regresyon katsayısıdır ve regresyon doğrusunun eğimidir, ε_i ise bilinmeyen rastgele hata değişkenidir.

3.1.2. Çoklu Doğrusal Regresyon Modeli

Bir bağımlı değişken Y ile birçok bağımsız değişken $(X_1, X_2, ..., X_k)$ arasındaki doğrusal ilişkiyi bulmak için çoklu doğrusal regresyon denklemi kullanılır.

Çoklu doğrusal regresyon modeli,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i \quad (i = 1, 2, ..., n), \tag{3.2}$$

denklemi ile tanımlanır. n gözlem sayısını göstermek üzere, her bir gözlem değeri için bu model,

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_k X_{1k} + \varepsilon \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \varepsilon_n \end{aligned} \tag{3.3}$$

denklemleri ile veya matris formunda

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1\ X_{11}\ X_{12} \ldots X_{1k} \\ 1\ X_{21}\ X_{22} \ldots X_{2k} \\ \vdots \\ 1\ X_{n1}\ X_{n2} \ldots X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \tag{3.4}$$

olarak ifade edilir. (3.4) Eşitliği ile verilen çoklu doğrusal regresyon modeli, $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$, $n-$boyutlu yanıt vektörünü, $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, ..., X_{ik})^T$ olmak üzere $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)^T$, $n \times (k+1)-$boyutlu açıklayıcı değişkenler matrisini, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_k)^T$, $(k+1)-$boyutlu bilinmeyen regresyon katsayıları veya parametre vektörünü, $\boldsymbol{\varepsilon}$, $n-$ boyutlu gözlenemeyen rastgele hata vektörünü göstermek üzere, kısaca

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.5}$$

şeklinde yazılabilir.

3.1.2.1. Çoklu Doğrusal Regresyon Modelindeki Varsayımlar

Çoklu doğrusal regresyon analizinin uygulanması aşağıdaki varsayımların geçerliliğine bağlıdır.

1. Hata terimlerinin ortalamaları sıfıra eşittir: $E(\varepsilon_i) = E(\varepsilon_i \mid \boldsymbol{X}_i) = 0$.

   Bu varsayımın sağlanmaması, yani $E(\varepsilon_i) \neq 0$ olması yanlı parametre tahminlerin bulunmasına neden olur.

2. ε_i hata teriminin varyansı bağımsız değişkenlere bağlı olmayıp, bütün gözlem değerleri için sabittir: $Var(\varepsilon_i) = E\left[\varepsilon_i - E(\varepsilon_i)\right]^2 = E\left[\varepsilon_i^2\right] = \sigma^2 \ (i = 1, 2, ..., n)$.

3. Hata terimleri arasında ve hata terimleri ile bağımsız değişkenler arasında ilişki yoktur: $Cov(\varepsilon_i, \varepsilon_j) = 0$, $Cov(\varepsilon_i, X_i) = 0$.
4. Gözlem sayısı (n), açıklayıcı değişken sayısından (k) büyüktür ve $\text{Rank}(\boldsymbol{X}) = k$ dir. Model oluşturulurken incelenen bağımlı değişkenin açıklanmasına yarayan tüm önemli değişkenlerin modele alındığı varsayılır.
5. Hata terimleri 0 ortalamalı ve σ^2 varyanslı normal dağılıma sahiptir: $E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2$, $\varepsilon_i \sim N(0, \sigma^2)$ $(i = 1, 2, ..., n)$. Buna göre, gözlenemeyen rastgele hata vektörü $\boldsymbol{\varepsilon}$ için, $E(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ yazılabilir.
6. Bağımsız değişkenler arasında anlamlı ilişki yoktur: $Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) = 0$. Bu varsayıma göre çoklu doğrusal regresyon modellerinde bağımsız değişkenler arasında doğrusal ya da doğrusala yakın bir ilişki yoktur. Değişkenler arasında iç ilişki olması durumunda $(\boldsymbol{X}^T \boldsymbol{X})^{-1}$ matrisinin tersi bulunamayacağı için parametre tahmini de yapılamayacaktır. Tersinin alınabildiği durumlarda ise parametre tahminlerinin varyansı büyük çıkacaktır. Bu durum da tahminlerin tutarsız olmasına neden olacaktır.

Yukarıda ifade edilen varsayımlardan birinin sağlanmaması durumunda parametre tahmincileri yanlı, tutarsız veya etkisiz olacaktır.

3.1.3. Çoklu Doğrusal Regresyonda Parametre Tahmin Yöntemleri

Çoklu doğrusal regresyonda parametre tahmininde amaç bilinmeyen parametre vektörü $\boldsymbol{\beta}$'yı tahmin etmektir. $\boldsymbol{\beta}$ parametresinin tahmin edicisi $\hat{\boldsymbol{\beta}}$'nın bulunması için en küçük kareler (EKK) ve en çok olabilirlik (ML) yöntemleri en çok kullanılan yöntemlerdir. Bu yöntemleri açıklamadan önce aşağıda iyi bir tahmin edicide bulunması gereken özelliklere göz atacağız.

3.1.3.1. Tahmin Edicilerde Aranan Özellikler

Örneklemden elde edilen değerler yardımıyla kitlenin bilinmeyen $\boldsymbol{\beta}$ parametrelerine ilişkin tahmin yapılabilmekle birlikte, $\boldsymbol{\beta}$ parametresinin birden fazla tahmin edicisi de elde edilebilir. Burada amaç, bu tahmin edicileri arasında kitle

parametresi $\boldsymbol{\beta}$'ya en yakın değere sahip olanının elde edilmesidir. Buda, tahmin edicinin yansızlık, tutarlılık ve etkinlik özellerine sahip olması ile mümkün olmaktadır. Bu özellikler aşağıdaki eşitlikler ile verilir

i) Yansızlık**:** $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\beta}$'nın tahmin edicisi olmak üzere, $\hat{\boldsymbol{\beta}}$'nın beklenen değeri ortalama olarak bilinmeyen parametre vektörü $\boldsymbol{\beta}$'ya eşit ise, yani

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \tag{3.6}$$

ise $\hat{\boldsymbol{\beta}}$'ya $\boldsymbol{\beta}$'nın yansız bir tahmin edicisi, $E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$ farkına sapma veya yanlılık miktarı denir.

ii) Tutarlılık: *n* gözlem sayısı artarken $\hat{\boldsymbol{\beta}}$ tahmin edicisi , $\boldsymbol{\beta}$'nın gerçek değerine yaklaşıyorsa $\hat{\boldsymbol{\beta}}$ tutarlı bir tahmin edicidir. Bu özellik matematiksel olarak,

$$\begin{aligned} &\lim_{n\to\infty} E(\hat{\beta}) = \beta, \\ &\lim_{n\to\infty} Var\left(\hat{\boldsymbol{\beta}}\right) = \mathbf{0} \end{aligned} \tag{3.7}$$

eşitlikleri ile ifade edilir (Rawlings ve ark. 1998)

iii) Etkinlik: $\boldsymbol{\beta}$ parametresinin tutarlı tahmin edicileri arasında yansız ve en küçük varyanslı olanına $\boldsymbol{\beta}$'nın etkin tahmin edicisi denir.

3.1.3.2. En Küçük Kareler Yöntemi (EKK)

Regresyon analizinde doğrusal modellerin parametre tahminlerinde, kullanımı ve de yorumlaması kolay olduğu için daha çok EKK yöntemi tercih edilir. Bir deneyin *n* kez tekrarlanması sonucunda, $(Y_i, X_i)\ (i = 1, 2, ..., n)$'nin elde edilen gözlem değeri $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$ olsun. EKK yöntemi, bu gözlem değerlerine bağlı olarak hata karelerinin toplamı

$$\begin{aligned} RSS(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 &= \sum_{i=1}^{n} \left(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}\right)^2 \\ &= \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right)^T \left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \end{aligned} \tag{3.8}$$

RSS'yi en küçük yapan değeri belirleyerek $\boldsymbol{\beta}$'yı tahmin etmeyi amaçlar. Yani, $\boldsymbol{\beta}$'nın EKK tahmini $\hat{\boldsymbol{\beta}}$,

$$\min_{\beta} RSS(\boldsymbol{\beta}) \tag{3.9}$$

en iyileme probleminin çözümünden elde edilir. $RSS(\boldsymbol{\beta})$ reel değerli, convex ve diferansiyellenebilir olduğunda, $RSS(\boldsymbol{\beta})$'yı minimum yapan değer her zaman mevcut olacaktır ve bu değer

$$\left.\frac{\partial RSS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\beta=\hat{\beta}} = \mathbf{0}, \tag{3.10}$$

normal denkleminin çözümünden elde edilir. (3.8) Eşitliğinin (3.10) denkleminde yerine yazılarak, $\boldsymbol{\beta}$'ya göre türev alınması sonucunda

$$-2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \mathbf{0},$$

veya

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^{T'}\boldsymbol{Y} \tag{3.11}$$

denklemi elde edilir. (3.11) Eşitliğinin her iki tarafı $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ ile çarpılırsa $\boldsymbol{\beta}$'nın en küçük kareler tahmin edicisi $\hat{\boldsymbol{\beta}}$,

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{3.12}$$

olarak elde edilir. Hata karelerinin toplamı kullanılarak σ^2'nin de tahmin edicisi bulunabilir. Eğer (3.8) eşitliğinde $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ alınarak $\hat{\boldsymbol{\beta}}$'nın değeri yerine yazılırsa (3.12),

$$\begin{aligned} RSS\left(\hat{\boldsymbol{\beta}}\right) &= \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) \\ &= \left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right)^T\left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right) \\ &= \left[\boldsymbol{Y} - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right]^T\left[\boldsymbol{Y} - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right] \\ &= \boldsymbol{Y}^T\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\boldsymbol{Y} \end{aligned} \tag{3.13}$$

eşitliği elde edilir. Burada $\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]$ simetrik ve idempotent bir matristir ve $RSS\left(\hat{\boldsymbol{\beta}}\right)/\sigma^2$, *n-p* serbestlik dereceli χ^2 dağılımına sahiptir. Yani,

$$\frac{RSS\left(\hat{\boldsymbol{\beta}}\right)}{\sigma^2} = \frac{\boldsymbol{Y}^T\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\boldsymbol{Y}}{\sigma^2} \sim \chi^2_{n-p}. \tag{3.14}$$

Hata karelerinin toplamının beklenen değeri $E\left[RSS\left(\hat{\boldsymbol{\beta}}\right)\right]$,

$$\begin{aligned} E\left(RSS\left(\hat{\boldsymbol{\beta}}\right)\right) &= E\left(\boldsymbol{Y}^T\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\boldsymbol{Y}\right) \\ &= trace\left(\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\sigma^2\boldsymbol{I}_n\right) + E\left(\boldsymbol{Y}^T\right)\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]E\left(\boldsymbol{Y}\right) \\ &= \left(n-p\right)\sigma^2, \end{aligned} \tag{3.15}$$

olduğundan, σ^2 EKK tahmin edicisi $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{E\left(RSS\left(\hat{\beta}\right)\right)}{\left(n-p\right)}, \tag{3.16}$$

olarak elde edilir. $\hat{\boldsymbol{\beta}}$ 'nın varyans –covaryans matrisi,

$$\begin{aligned} \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}\right) = \mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) &= E\left\{\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]^T\right\} \\ &= E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T\right] \\ &= MSE\left(\hat{\boldsymbol{\beta}}\right), \end{aligned} \tag{3.17}$$

j. köşegen elemanı $\hat{\beta}_j$ in varyansına, *ij*. elemanı $\hat{\beta}_i$ ve $\hat{\beta}_j$ arasındaki kovaryansa karşılık gelen $k \times k$ tipinde simetrik bir matristir. (3.12) Eşitliği (3.17) da yerine yazılırak , $Var\left(\boldsymbol{Y}\right) = \sigma^2\mathbf{I}_n$ olduğu göz önüne alınırsa,

$$\begin{aligned}\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) &= \mathrm{Var}\left[\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right] \\ &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T Var\left(\boldsymbol{Y}\right)\left[\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]^T \\ &= \sigma^2 \mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},\end{aligned} \tag{3.18}$$

olarak elde edilir. $\boldsymbol{A} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = \left[\left(a_{ij}\right)\right](i,j=1,2,...,k)$ olarak alınırsa,

$\mathrm{Var}\left(\hat{\beta}_j\right) = \sigma^2 a_{jj}$, $\mathrm{Cov}\left(\hat{\beta}_i,\hat{\beta}_j\right) = \sigma^2 a_{ij}$ olacaktır.

3.1.3.3. En Çok Olabilirlik Yöntemi (MLE: Maksimum Likelihood)

Parametre tahmininde kullanılan en yaygın yöntemlerden biride en çok olabilirlik yöntemidir. Bu yöntem doğrusal regresyon modelindeki hata vektörünün $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0},\sigma^2\mathbf{I}_n\right)$ olduğu varsayımı altında çalışır (Montgomery ve ark. 2012).

(3.5) Eşitliği ile verilen doğrusal regresyon modelindeki i. hata terimi için normal olasılık fonksiyonu,

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2}, i=1,2,...,n \tag{3.19}$$

şeklindedir. Bu model için en çok olabilirlik fonksiyonu,

$$L\left(\boldsymbol{\varepsilon},\boldsymbol{\beta},\sigma^2\right) = \prod_{i=1}^{n} f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}} \tag{3.20}$$

olarak elde edilir. (3.20) Eşitliğinde hata vektörü, $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$ olarak yazılırsa, en çok olabilirlik fonksiyonu

$$L\left(\boldsymbol{\varepsilon},\boldsymbol{\beta},\sigma^2\right) = \prod_{i=1}^{n} f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}(Y-X\boldsymbol{\beta})^T(Y-X\boldsymbol{\beta})} \tag{3.21}$$

olur. (3.21) Eşitliğinin logaritması alınırsa,

$$\ln L(\boldsymbol{\varepsilon},\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right) \tag{3.22}$$

eşitliği veya eşdeğer olarak,

$$\ln L(\varepsilon,\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\left(\boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}\right) \quad (3.23)$$

şeklinde de yazılabilir. $\boldsymbol{\beta}$'nın ve σ^2'nin MLE tahmin edicileri sırası ile $\tilde{\boldsymbol{\beta}}$ ve $\tilde{\sigma}^2$ olmak üzere bu tahmin ediciler, (3.23) denkleminin sırasıyla σ^2 ve $\boldsymbol{\beta}$'ya göre türevinin sıfıra eşitlenmesiyle elde edilen

$$\begin{aligned} \frac{\partial\left(\ln L\left(\boldsymbol{\varepsilon},\boldsymbol{\beta},\sigma^2\right)\right)}{\partial\boldsymbol{\beta}} &= \mathbf{0}, \\ \frac{\partial\left(\ln L\left(\boldsymbol{\varepsilon},\boldsymbol{\beta},\sigma^2\right)\right)}{\partial\sigma^2} &= \mathbf{0}. \end{aligned} \quad (3.24)$$

denklem sisteminin, yani

$$\begin{aligned} -\frac{1}{2\sigma^2}\left(-2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}\right) &= \mathbf{0}, \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right)^T\left(Y - X\boldsymbol{\beta}\right) &= \mathbf{0}, \end{aligned} \quad (3.25)$$

denklemlerinin ortak çözümünden elde edilir. Bu çözümler sırası ile

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \\ \tilde{\sigma}^2 &= \frac{(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})}{n}, \end{aligned} \quad (3.26)$$

eşitlikleri ile verilir. (3.12) ve (3.26) eşitlikleri karşılaştırıldığında, bilinmeyen $\boldsymbol{\beta}$ parametresinin EKK tahmin edicisi ile MLE tahmin edicisinin birbirine eşit ve

$$\begin{aligned} \mathrm{E}\left(\hat{\beta}\right) = \boldsymbol{E}\left(\tilde{\beta}\right) &= \mathrm{E}\left[\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right] \\ &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathrm{E}\left(\boldsymbol{Y}\right) \\ &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned} \quad (3.27)$$

olduğundan yansız olduğu kolayca görülebilir. Bu ise $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0},\sigma^2\boldsymbol{I}_n\right)$ olmasının bir sonucudur.

Montgomery ve ark.(2012), En çok olabilirlik yönteminin en küçük kareler yöntemi ile kıyaslandığında çok daha iyi bir yöntem olduğuna, MLE tahmin edicilerinin etkin tahmin ediciler olduğuna (yani, yansız, tutarlı ve diğer tahmin ediciler arasında en küçük varyansa sahip), ayrıca hata vektörlerinin normal dağıldığı varsayımı hariç EKK yönteminde kabul edilen varsayımlara bu yöntemde gerek duyulmadığına dikkat çekmişlerdir. Ancak günümüzde doğrusal regresyon modelindeki parametrelerin tahmininde, pratik ve kullanışlı olması nedeni ile EKK daha çok tercih edilir.

3.2. Lineer Regresyonda İç İlişki Problemi

Çeşitli uygulama alanlarında kullanılan çoklu doğrusal regresyon modelleri, bağımsız değişkenler arasında doğrusal bir ilişkinin olmadığı varsayımı altında kullanılır. Ancak bu modeldeki bazı açıklayıcı değişkenler arasında iç ilişki bulunması durumunda modelin faydalarını olumsuz bir şekilde etkileyen ciddi bir problem ortaya çıkar. Genel olarak, iç ilişki olarak adlandırılan bu problemin mevcut olduğu durumda EKK parametre tahmincilerinin yansızlığı korunur ancak bu tahmincilerin varyans ve kovaryans değerleri artacaktır. İç ilişki, bağımsız değişkenler matrisi X'in sütunları arasında doğrusala yakın bir ilişkinin, $\boldsymbol{X}^T\boldsymbol{X}$'in singular olması durumda ise tam bir doğrusal ilişkinin olduğunu ifade eder. Bu durumda da model üzerinde yapılacak yorum ve tahminler iç ilişkinin olumsuz etkileri nedeni ile hatalı olacağından, regresyon analizi de faydalı sonuçlar üretmeyecektir. Bu nedenle regresyon analizi yapılmadan önce bağımsız değişkenler arasında iç ilişkinin olup olmadığı büyük önem taşır ve mutlaka saptanması gerekir. İç ilişkinin olması durumunda, iç ilişkinin zararlı etkilerini ortadan kaldıracak çözümler üretilmelidir. Aşağıdaki alt bölümde Çoklu doğrusal regresyonda iç ilişki saptama yöntemleri ele alınacaktır. Bu yöntemler ele alınırken, açıklayıcı değişkenler ve yanıt değişkenin birim uzunluğa göre standartlaştırıldığını (ortalandığını ve ölçeklendiğini) kabul ediyoruz. Yani, sırası ile *j*. açlıklayıcı değişken ve yanıt değişkene ilişkin varyanslar

$$s_X^2 = \frac{\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2}{n-1},\ s_y^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}{n-1},$$

olmak üzere, X ve Y'nin standartlaştırılmış hali,

$$X_i^* = \frac{X_{ij} - \overline{X}_j}{s_X} \qquad \text{ve} \qquad Y_i^* = \frac{Y_i - \overline{Y}}{s_y} \tag{3.28}$$

şeklindedir. Bu standartlaştırmanın sonucunda $\beta_0 = \overline{Y} = 0$ olacağından β_0 parametresi modelde yer almayacaktır ve $k \times k$ tipindeki $\boldsymbol{X}^T\boldsymbol{X}$ matrisi açıklayıcı değişkenler arasındaki korelasyon matrisi, $k \times 1$ tipindeki $\boldsymbol{X}^T\boldsymbol{Y}$ vektörü ise açıklayıcı ve yanıt değişkenleri arasındaki korelasyon matrisine karşılık gelecektir.

3.2.1. İç İlişki Problemini Saptama Yöntemleri

İç ilişki probleminin varlığını tespit etmek için çeşitli teknikler vardır (Belsley ve ark. 1980). Korelasyon katsayısı, belirleme katsayısı $\left(R^2\right)$, varyans şişirme faktörü (VIF), $\boldsymbol{X}^T\boldsymbol{X}$ korelasyon matrisinin öz değerleri ve korelasyon matrisinin determinantı, tolerans değeri, *F*-test ve *t*-test değerlerine bağlı olarak kullanılan bu teknikler aşağıda detaylı olarak ele alınacaktır.

3.2.1.1. Korelasyon Katsayısı

İç ilişki sorununun saptanmasında kullanılan tekniklerden ilki ve en basit olanı $\boldsymbol{X}^T\boldsymbol{X}$ in köşegen elemanları dışında bulunan korelasyon katsayılarıdır (Gujarati 2003). X_i ve X_j bağımsız değişkenleri için korelasyon katsayısı r_{ij} olmak üzere, X_i ve X_j bağımsız değişkenleri arasında doğrusala yakın bir ilişki söz konusu ise, $\left|r_{ij}\right|$ basit korelasyon katsayısının mutlak değeri 1'e yakın değer alacaktır. Bu yaklaşım ile sadeceiki bağımsız değişken arasındaki ilişki açıklanabilmektedir. İkiden fazla değişken arasındaki ilişkiyi açıklamada yetersiz kaldığından, ikiden fazla değişken arasındaki ilişkiyi açıklamak için bu yaklaşım yerine kısmi korelasyon katsayıları incelenir. İç ilişki olması durumunda kısmi korelasyon katsayıları küçük çıkacaktır. Ancak kısmi korelasyon katsayılarının büyük olduğu durumlarda da iç ilişki saptanabilmektedir. Bu nedenle bu yöntem iç ilişki saptamada fazla tercih edilen bir yöntem değildir.

3.2.1.2. R^2 Belirleme Katsayısı

İç ilişki problemini saptamak için kullanılan ikinci teknik, regresyona bağlı karelerin toplamı S_R nin toplam karelerin toplamı S_T ye oranı olarak tanımlanan R^2 belirleme katsayısıdır:

$$R^2 = \frac{S_R}{S_T} = \frac{\hat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{Y} - n\overline{Y}}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2} = 1 - \frac{RSS\left(\hat{\boldsymbol{\beta}}\right)}{S_T}. \tag{3.29}$$

Burada,

$$S_T = S_R + RSS\left(\hat{\boldsymbol{\beta}}\right) \text{ ve } \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \tag{3.30}$$

şeklindedir. Modele yeni bir bağımsız değişken eklendiğinde R^2'de büyük bir değişimin olmaması değişkenler arasında çoklu doğrusal ilişki problemini gösterir (Albayrak 2005).

3.2.1.3. Varyans Şişirme Faktörü (VIF)

İç ilişki problemini saptamak için en çok kullanılan yöntem Marquart (1970) tarafından önerilen varyans şişirme faktörüdür (Marquart ve Snee 1975). Bu yöntemde ele alınan regresyon modelinde her bir bağımsız değişken için VIF hesaplanır. Bu yöntemi

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \tag{3.31}$$

regresyon modelini kullanarak açıklamaya çalışalım. İlk aşamada X_{i1} bağımlı X_{i2} ve X_{i3} bağımsız değişkenler olarak kabul edilerek R_1^2 bulunur. X_{i1} için varyans şişirme faktörü,

$$\text{VIF}\left(X_{i1}\right) = \frac{1}{\left(1 - R_1^2\right)} \tag{3.32}$$

hesaplanır. Benzer şekilde, X_{i2} bağımlı, X_{i1} ve X_{i3} bağımsız değişkenler olarak kabul edilerek VIF değeri ,

$$\mathrm{VIF}\left(X_{i2}\right)=\frac{1}{\left(1-R_2^2\right)} \tag{3.32}$$

bulunur (Albayrak 2005). Son adımda ise X_{i3} bağımlı X_{i1} ve X_{i2} bağımsız değişkenler olarak kabul edilerek VIF değeri,

$$\mathrm{VIF}\left(X_{i3}\right)=\frac{1}{\left(1-R_3^2\right)} \tag{3.33}$$

bulunur. Daha genel bir ifadeyle, (3.18) verilen $\boldsymbol{A}=\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ kovaryans matrisinin köşegen elemanları VIF değerine eşittir ve bu değer,

$$\mathrm{VIF}_j=a_{jj}=\frac{1}{(1-R_j^2)} \tag{3.34}$$

eşitliği ile verilir. R_j^2 ile gösterilen belirleme katsayısı *j.* değişken için iç ilişki probleminin olmaması durumunda sıfır çıkacaktır ve VIF_j değeri de 1'e eşit olacaktır. İç ilişki olması durumunda ise R_j^2 1'e yakın değer alarak, VIF_j'nin de büyük değer almasına sebep olur. İç ilişki probleminin saptanmasında en çok kullanılan yöntem olan VIF in değerlerinin yorumlarına ilişkin ortak bir görüş bulunmamaktadır. Ancak "VIF değerinin 10'a eşit veya 10'dan büyük olduğu durumlar da iç ilişki probleminin varlığı Webster (1992) tarafından gösterilmiştir.

3.2.1.4. X^TX Korelasyon Matrisinin Öz Değerlerinin İncelenmesi

İç ilişki saptamada kullanılan $\boldsymbol{X}^T\boldsymbol{X}$ korelasyon matrisinin öz değerleri $\left(\lambda_1,\lambda_2,...,\lambda_k\right)$ çoklu ilişkinin hangi ölçüde olduğu hakkında bilgi vermektedir. $\left(\lambda_1,\lambda_2,...,\lambda_k\right)$ öz değerlerinden bir veya daha fazlasının sıfıra yakın değer alması iç ilişki probleminin olduğunu gösterir. Ancak, iç ilişki problemini saptamada her bir öz değeri incelemek yerine en büyük öz değerin en küçük öz değere oranı olarak tanımlanan koşul sayısı (KS),

$$KS = \frac{\lambda_{max}}{\lambda_{min}} \tag{3.35}$$

kullanılır. Koşul sayısının 100 ile 1000 arasında olması orta, 1000'den büyük olması güçlü doğrusal ilişki olduğunu gösterir. Bazı araştırmacılar tarafından, çoklu ilişki belirlenmesinde KS nın karekökü olarak tanımlanan koşul indeksi (KI)

$$KI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \tag{3.36}$$

en iyi ölçüt olarak kabul edilmektedir. Koşul indeksi 10 ile 30 arasında bir değer alırsa orta, 30'dan büyük olursa güçlü çoklu ilişki problemi vardır (Gujarati 2003).

3.2.1.5. $X^T X$ Korelasyon Matrisinin Determinantı

$X^T X$ korelasyon matrisinin determinantı hangi bağımsız değişkenler arasında iç ilişki olduğunu göstermez ancak iç ilişki probleminin derecesi hakkında bilgi verir (Farrar ve Glauber 1967). Determinantın değeri sıfıra yaklaştıkça iç ilişki probleminin derecesi de artacaktır. Determinantın sıfır olması tam iç ilişki problemi olduğunu, 1 olması ise iç ilişki olmadığını gösterir.

3.2.1.6. Tolerans Değeri

İç ilişki probleminin saptanmasında kullanılan bir diğer teknik, tolerans değeri (TD) dir. Bu değer

$$TD = (1 - R_i^2) \tag{3.37}$$

eşitliği ile tanımlanır. Tolerans değerinin sıfır olması iç ilişki problemi olduğunu, 1'e eşit olması iç ilişki probleminin olmadığını gösterir (Kutner ve ark. 2004).

3.2.1.7. F ve t Testi İle Belirlenmesi

Doğrusal regresyon modelindeki açıklayıcı değişkenler arasında iç ilişki probleminin olup olmadığı F ve t istatistik değerleri kullanılarak da belirlenebilir. İç ilişki probleminin olması durumunda F istatistik değerleri büyük ve t istatistik değerleri küçülerek anlamsız sonuçlar verir. Dolayısıyla yüksek F değeri ile küçük t değeri çoklu doğrusal ilişki probleminin varlığını gösterir. Ancak, F istatistik değerlerinin büyük

olup tüm t istatistik değerlerinin küçük hesaplanmadığı zamanlarda da çoklu ilişki problemi görülebilmektedir (Johnston 1972). Bu nedenle, bilimsel araştırmalarda bu yöntem çok etkili bir yöntem değildir.

3.2.2. İç İlişki Problemi İçin Çözümler

Doğrusal regresyonda karşılaşılan iç ilişki problemini gidermek için çeşitli yöntemler geliştirilmiştir (Gujarati 2003) . Bu yöntemler, aşağıdaki şekildedir:

i) Ek veri toplama işlemi. Bu yöntemdeki amaç örneklem büyüklüğünü artırıp, tahmin edicilerin varyansını azaltmaktır. Böylece standart hata azaltılarak, daha kesin tahminler elde edilecektir. Ancak hem ekonomik kısıtlamalar hem de uygun olmayan çalışma süreçleri bu yöntemin kullanılmasını zorlaştırmaktadır.

ii) İç ilişki saptanması durumunda yapılabilecek en basit işlem, aralarında iç ilişki olan değişken veya değişkenlerin modelden atılmasıdır. Ancak modelden çıkarılan değişkenin bağımlı değişken üzerinde güçlü etkisi olması durumunda modelin açıklama gücüne zarar vermektedir.

iii) Bir başka yöntem iç ilişki problemine neden olan bağımsız değişkenleri yeniden tanımlayarak sorunun ortadan kaldırılmasıdır. Örneğin, aralarında ilişki bulunan X_1 ve X_2 bağımsız değişkenlerini tek tek modele almak yerine bu değişkenlerin içeriğini korumak koşuluyla, $X_1 + X_2$ veya $\frac{X_1 + X_2}{2}$ gibi fonksiyonları modele dahil edilebilir.

iv) İç ilişki problemi durumunda artan varyanslar karşısında yetersiz kalan en küçük kareler yöntemi yerine küçük bir yanlılık katsayısı kullanarak daha kesin tahminlerin elde edilmesini sağlayan düzenleme veya küçültme yöntemleri olarak adlandırılan yöntemler kullanılır. RR (Hoerl ve Kennard 1970a), LASSO (Usai ve ark. 2012), EN, LR (Akdeniz ve Kaçıranlar 1995) , gibi küçültme yöntemleri en çok bilinen ve bilimsel araştırmalarda yaygın olarak tercih edilen küçültme yöntemleridir. Bir sonraki bölümde bu yöntemler ele alınacaktır.

3.3. Lineer Regresyonda Küçültme Yöntemleri

EKK tahmin edicileri regresyon modelindeki bütün katsayıları tahmin eder ve bu tahmin ediciler genellikle küçük yanlılık miktarına, ancak büyük varyansa sahiptirler. Regresyon modelindeki katsayıların sayısının fazla olması tahmin edilen modelin

yorumlanmasını zorlaştırırken, tahmin edicilere ilişkin küçük yanlılık miktarı ve büyük varyans tahminin doğruluğunu olumsuz yönde etkilemektedir. Regresyon modelindeki bazı β_j katsayıların küçültülmesi veya 0 olarak elde edilmesi tahminin doğruluğuna ilişkin olumsuzluğu ortadan kaldırır. Bunun içinde tahmin edilen regresyon katsayılarının yanlılığı küçük miktar da artırılır ve varyansları ise düşürülür, ve böylece genelde tahminin doğruluk miktarı artırılır. Regresyon modelindeki daha güçlü etkilere sahip katsayıların bir kümesinin seçilmesi ise modelin doğru yorumlanmasına katkıda bulunur. EKK tahmin edicilerinin bu iki olumsuzluğunu ortadan kaldıran en bilinen iki yöntem sırası ile Ridge (Hastie ve ark. 2008) ve alt küme seçim yöntemidir. Ancak her iki yönteminde bazı sakıncaları vardır. Ridge, iç ilişki probleminin olması durumunda bazı katsayıları sadece küçülten, onları ortadan kaldırmayan sürekli bir yöntemdir. Buda modelin daha kolay bir şekilde yorumlanmasına katkıda bulunmaz. Alt küme seçimi, yorumlanabilir modeller sunar ancak, bu yöntem kesikli bir prosedür olduğundan açıklayıcıların modelde tutulup veya düşürülmesi aşırı değişken modeller sunabilir. Verilerdeki küçük değişiklikler çok farklı modellerin seçilmesine neden olabilir ve bu durum tahmin doğruluğunu azaltabilir. LASSO bu iki yöntemin iyi özelliklerini birleştirerek tahmin yapar (Tibshirani 1996).

3.3.1. Ridge Regresyon

Bilindiği gibi, $\boldsymbol{\beta}$ nın EKK tahmin edicisi $\hat{\boldsymbol{\beta}}$ nın yansız ve minimum varyanslı olması istenen bir durumdur. Ancak iç ilişki probleminin söz konusu olması durumunda varyans ve kovaryanslar artabilmektedir. Bu durumda $\hat{\boldsymbol{\beta}}$ tahmin edicisinin varyansını küçültmek için yansızlık özelliğinden vazgeçilmesi gerekebilir. Yani, varyansı $\hat{\boldsymbol{\beta}}$'dan daha küçük olan yanlı tahmin ediciler tercih edilebilir. Eğer $\hat{\boldsymbol{\beta}}_Y$ böyle bir tahmin edici ise, $\hat{\boldsymbol{\beta}}_Y$ nin hata karelerinin ortalaması (MSE) (Marquardt ve Snee 1975)

$$\begin{aligned} \mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_Y\right) = E\left\|\hat{\boldsymbol{\beta}}_Y - \boldsymbol{\beta}\right\|^2 &= \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_Y\right) + \left[E\left(\hat{\boldsymbol{\beta}}_Y\right) - \boldsymbol{\beta}\right]^2 \\ &= \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_Y\right) + \left[Yan\left(\hat{\boldsymbol{\beta}}_Y\right)\right]^2 \end{aligned} \tag{3.38}$$

şeklindedir. Burada $Yan\left(\hat{\boldsymbol{\beta}}_Y\right) = E\left(\hat{\boldsymbol{\beta}}_Y\right) - \boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_Y$'daki yanlılığı ifade etmektedir. (3.38) eşitliğinde de görüldüğü gibi $\hat{\boldsymbol{\beta}}_Y$'daki küçük bir yanlılık miktarı, $\hat{\boldsymbol{\beta}}_Y$'nın varyansını küçültmektedir. Buda $\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_Y\right)$nın, EKK tahmin edicisi $\hat{\boldsymbol{\beta}}$'nın varyansı $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$'dan daha küçük olduğunu gösteriyor. $\hat{\boldsymbol{\beta}}_Y$ yanlı tahmin edicisinin daha küçük varyanslı olması, yansız EKK tahmin edicisi $\hat{\boldsymbol{\beta}}$ya göre daha tutarlı bir tahmin edici olduğunu gösteriyor. Ancak, iyi bir tahminde bulunmak için (3.38) eşitliğindeki varyans ve $(\mathrm{Yan})^2$ miktarının mümkün olduğunca küçük olması gerekmektedir. Varyans, fonksiyonun hangi ölçüde tahmin edildiği hakkında bilgi verirken $(\mathrm{Yan})^2$, modelin gerçek f fonksiyonundan ne kadar saptığını ölçer. MSE dengesini koruyarak varyansı küçültmek için (Yan) ın artırılması gerekebilir.

Yanlı tahmin edicileri elde etmek için bazı yöntemler geliştirilmiştir. Bunların en çok bilineni ve tercih edileni, Hoerl ve Kennard (1970a, 1970b) tarafından önerilen ridge regresyondur.

Ridge regresyonu, EKK da kullanılan hata kareler toplamına ilaveten bir ceza terimi aracılığıyla, β_j parametrelerini sıfira doğru küçülterek tahmin etmeyi amaçlayan cezalandırılmış en küçük kareler (PRSS) eniyileme problemini çözer. $\boldsymbol{\beta}$ Parametresinin ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$ olmak üzere, $\hat{\boldsymbol{\beta}}_R$,

$$\hat{\boldsymbol{\beta}}_R = \arg\min_{\boldsymbol{\beta}} PRSS\left(\boldsymbol{\beta}\right) = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}\right)^2 + K \sum_{j=1}^{k} \beta_j^2 \right\} \qquad (3.39)$$

probleminin çözümünden elde edilir. (3.39) Eşitliğindeki K sabitine ceza, ayar, küçültme, yanlılık veya ridge parametresi denir. $K \geq 0$ ceza parametresi, ceza terimin etkinliğini kontrol eder. Bir diğer ifadeyle, tahmin edicilerin küçülme miktarlarını kontrol eder. Bu parametre için üç durum söz konusu olabilir:

i) $K = 0$ olduğunda, $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}$ olacaktır. Yani; $\boldsymbol{\beta}$'nın Ridge tahmin edicisi EKK tahmin edicisini verecektir.

ii) $K = \infty$ olduğunda, $\hat{\boldsymbol{\beta}}_R = \mathbf{0}$ olacaktır.

iii) $K \in (0, \infty)$ olduğunda, K ceza parametresi korelasyon matrisi $(\boldsymbol{X}^T\boldsymbol{X})$'in köşegen elemanlarına eklenerek, β_j parametrelerinin değerlerini küçültür ve $\boldsymbol{X}$ e bağlı olarak $\boldsymbol{Y}$ nin doğrusal modelini oluşturur (Hastie ve ark. 2008).

(3.39) en iyilime problemi matris formunda aşağıdaki şekilde yazılır:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= \arg\min_{\boldsymbol{\beta}} \left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)+K\boldsymbol{\beta}^T\boldsymbol{\beta} \\ &= \arg\min_{\boldsymbol{\beta}} \left\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right\|_2^2 + K\left\|\boldsymbol{\beta}\right\|_2^2.\end{aligned} \tag{3.40}$$

(3.40) Eşitliği ile verilen en iyileme probleminin çözümü, $PRSS$'nin β'ya göre türevi alınıp sıfıra eşitlenmesi ile elde edilir:

$$\begin{aligned}\frac{\partial PRSS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -2\boldsymbol{X}^T\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)+2K\boldsymbol{\beta} \\ &= 0.\end{aligned} \tag{3.41}$$

(3.41) Eşitliği kullanılarak,

$$\begin{aligned}\boldsymbol{X}^T\boldsymbol{Y} &= \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}+K\boldsymbol{\beta}, \\ \boldsymbol{\beta}\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right) &= \boldsymbol{X}^T\boldsymbol{Y},\end{aligned} \tag{3.42}$$

eşitlikleri yazılabilir. Bu eşitliklerden hareketle, $\boldsymbol{\beta}$'nın ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$,

$$\hat{\boldsymbol{\beta}}_R = \left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \tag{3.43}$$

olarak elde edilir. Ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$'in beklenen değeri:

$$\begin{aligned}\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right) &= E\left\{\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right\} \\ &= E\left\{\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\right\} \\ &= \left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)\boldsymbol{\beta},\end{aligned} \tag{3.44}$$

olarak elde edilir. $\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)=\boldsymbol{Z}_K$ olarak alınırsa, ridge tahmin edicisinin beklenen değeri:

$$\begin{aligned} E\left(\hat{\boldsymbol{\beta}}_R\right) &= \boldsymbol{Z}_K\boldsymbol{\beta} \\ &\neq \boldsymbol{\beta} \end{aligned} \tag{3.45}$$

olarak elde edilir ki, bu da $\hat{\boldsymbol{\beta}}_R$ in yanlı olduğunu gösterir. Ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$'in varyansı $\mathrm{Var}\left(\hat{\beta}_R\right)$:

$$\begin{aligned} \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_R\right) &= \mathrm{Var}\left\{\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right\} \\ &= \left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}\right)^{-1}\sigma^2 \end{aligned} \tag{3.46}$$

eşitliği ile verilir (Montgomery ve Peck 2001). (3.46) Eşitliğinden yararlanarak ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$ için hata kareler ortalaması $\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_R\right)$;

$$\begin{aligned} \mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_R\right) &= \mathrm{E}\left\|\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}\right\|^2 \\ &= \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}\right)^T\left(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}\right)\right] \\ &= \mathrm{E}\left\|\hat{\boldsymbol{\beta}}_R-\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)\right\|^2+2\mathrm{E}\left[\left(\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)-\boldsymbol{\beta}\right)^T\left(\hat{\boldsymbol{\beta}}_R-\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)\right)\right]+\left\|\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)-\boldsymbol{\beta}\right\|^2 \\ &= \mathrm{E}\left\|\hat{\boldsymbol{\beta}}_R-\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)\right\|^2+\left\|\mathrm{E}\left(\hat{\boldsymbol{\beta}}_R\right)-\boldsymbol{\beta}\right\|^2 \\ &= \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_R\right)+\left\|Yan\left(\hat{\boldsymbol{\beta}}_R\right)\right\|^2 \\ &= \sigma^2 iz\left[\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\boldsymbol{X}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-1}\right]+K^2\boldsymbol{\beta}^T\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-2}\boldsymbol{\beta} \\ &= \sigma^2\sum_{j=1}^{k}\frac{\lambda_j}{\left(\lambda_j+K\right)^2}+K^2\boldsymbol{\beta}^T\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-2}\boldsymbol{\beta} \end{aligned} \tag{3.47}$$

olarak elde edilir. Eğer (3.47) eşitliğindeki ridge tahmin edicisi $\hat{\boldsymbol{\beta}}_R$ in varyansı $\gamma_1(K)$,

$$\gamma_1(K)=\sigma^2\sum_{j=1}^{k}\lambda_j/\left(\lambda_j+K\right)^2$$

ve yanlılık miktarının karesi $\gamma_2(K)$,

$$\gamma_2(K)=K^2\beta^T\left(\boldsymbol{X}^T\boldsymbol{X}+K\boldsymbol{I}_k\right)^{-2}\boldsymbol{\beta}$$

olarak alınırsa, ridge tahmin edisi $\hat{\boldsymbol{\beta}}_R$ için hata kareler ortalaması $\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_R\right)$:

$$\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_R\right)=\gamma\left(K\right)=\gamma_1\left(K\right)+\gamma_2\left(K\right) \tag{3.48}$$

olarak yazılabilir. (3.48) Eşitliğinde de açık olarak görülebileceği gibi, ridge tahmin edisi $\hat{\boldsymbol{\beta}}_R$ için hata kareler ortalaması $\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}_R\right)$, ceza parametresinin bir fonksiyonudur.

3.3.1.1. Ridge Ceza Parametresinin Seçimi

Ridge regresyonunda uygun bir ceza parametresi K'nın seçilmesi durumunda elde edilen hata kareler ortalaması, EKK yöntemi kullanılarak elde edilen hata kareler ortalamasına göre daha küçük olarak elde edilir. Bu nedenle, ceza parametresi K'nın seçilmesi ridge regresyonunda büyük önem taşır. ceza parametresi K'nın seçimi için çeşitli yöntemler mevcuttur. Hoerl ve Kennard (1970b) K parametresinin değerini araştırmak için ridge çözümlerinin izini önermiştir. Ancak birçok araştırmacı K parametresinin seçimi için analitik işlemler önermiştir. Bu işlemlerden biri, Hoerl ve ark. (1975) tarafından önerilen ve EKK tahmin edicilerine dayalı olarak ifade edilen aşağıdaki eşitlik ile verilir:

$$K=\frac{k\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}}=\frac{k\hat{\sigma}^2}{\left\|\hat{\boldsymbol{\beta}}\right\|^2}, \tag{3.49}$$

burada $\hat{\boldsymbol{\beta}}$ ve $\hat{\sigma}^2$ sırası ile $\boldsymbol{\beta}$ ve σ^2'nin EKK tahmin edicileridir. Anderson(1998), VIF değerlerini 1'e yaklaştıran K değerinin alınabileceğini belirtmiştir. McDonald ve Galarneau (1975) K parametresinin seçimi için aşağıdaki eşitliği önermiştir:

$$\hat{\boldsymbol{\beta}}_R^T\hat{\boldsymbol{\beta}}_R=\hat{\boldsymbol{\beta}}^T\hat{\boldsymbol{\beta}}-\hat{\sigma}^2\sum_{j=1}^{k}\left(1/\lambda_j\right). \tag{3.50}$$

Marquart ve Snee (1970) ise, VIF değerinin 1 ile 10 arasında olmasını sağlayan K değerinin kullanılabileceğini belirtmiştir. K Parametresinin belirlenmesine ilişkin en bilinen ve yaygın olarak kullanılan yöntem Genelleştirilmiş çapraz değerlendirme yöntemidir (GCV) (Golub ve ark. 1979). Bu yöntemde,

$$R\left(K\right)=\frac{1/n\left\|\left(\boldsymbol{I}-\boldsymbol{A}\left(K\right)\right)\boldsymbol{Y}\right\|^2}{\left[1/n iz\left(\boldsymbol{I}-\boldsymbol{A}\left(K\right)\right)\right]^2}, \tag{3.51}$$

eşitliğini minimum yapan değer ceza parametresi K'nın değeri olarak kullanılır. Burada $A(K) = X\left(X^T X + nKI_k\right)^{-1} X^T$ şeklindedir.

3.3.2. En Küçük Mutlak Küçültme ve Seçim Operatörü (LASSO)

Ridge regresyon β_j katsayılarını küçülterek daha istikrarlı modeller elde eden bir küçültme yöntemi olmakla birlikte, β_i katsayılarını 0 olarak belirlemez. Bu durum iç ilişki problemini de bulunduran modellerin yorumlanmasını oldukça zorlaştırır. Tibshirani (1996) tarafından önerilen, En Küçük Mutlak Küçültme ve Seçim Operatörü (LASSO) yöntemi regresyon modelindeki bazı β_i katsayılarını küçültüp diğerlerini de 0 olarak elde ederek hem tahmin edicilerin varyanslarını küçültür hem daha kolay yorumlanabilir modeller sunar.

(3.5) lineer regresyon modelindeki $\boldsymbol{\beta} = \left(\beta_0, \beta_1 ..., \beta_k\right)^T$ parametresinin LASSO tahmin edicisi (Tibshirani 1996) tarafından aşağıdaki kısıtlı (K.S) bir eniyileme problemi olarak tanımlamıştır:

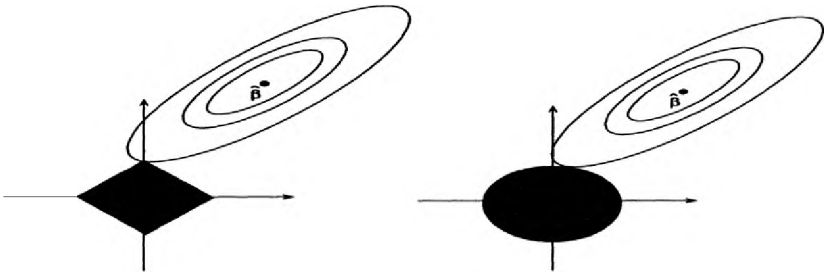
$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg\min \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2$$

$$\text{K.S} \quad \sum_{j=1}^{k} \left|\beta_j\right| \le t \quad \text{koşulu ile} \tag{3.52}$$

$\hat{\beta}_j^0$, EKK tahmin edicileri ve $t_0 = \sum_{j=1}^{k} \left|\beta_j^0\right|$ olsun. $t < t_0$ değerleri çözümleri 0'a doğru küçültürken bazı katsayıları da tam olarak 0'a eşit hesaplayacaktır.Ridge regresyonunda olduğu gibi β_j regresyon katsayılarının bulunması için standartlaştırma işlemi yapılıp ψ ceza sabitinin de modele eklenmesi ile LASSO problemi Lagrange formunda, $\|\cdot\|_1$, L_1 normu göstermek üzere aşağıdaki şekilde yazılabilir:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{LASSO} &= \arg\min \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 + \psi \sum_{j=1}^{k} \left|\beta_j\right| \right\} \\ &= \arg\min \frac{1}{2} \left\| y - X\boldsymbol{\beta} \right\|_2^2 + \psi \left\| \boldsymbol{\beta} \right\|_1 . \end{aligned} \tag{3.53}$$

Ridge regresyonda L_2 normu $\|\cdot\|_2$ kullanılarak verilen $\sum_{j=1}^{k}\beta_j^2$ ceza fonksiyonunun LASSO regresyonda karşılığı $\sum_{j=1}^{k}|\beta_j|$ fonksiyonudur. LASSO çözümleri bu şekliyle doğrusal olmayan forma dönüştüğünden hesaplamalar için karesel programlar kullanılmıştır. Lasso yöntemi bazı katsayıları sıfıra doğru küçültürken bazılarını da tamamen sıfır yapmayı sağlayan bir yöntem olmasının yanında, $n > k$ olması durumunda değişkenler arasında yüksek ilişki olması bu yöntemin dezavantajlarındandır (Hastie ve ark. 2003). Şekil 3.1.' de $k = 2$ için Ridge regresyon ve Lasso yöntemlerinin hata ve kısıt fonksiyonları karşılaştırılmıştır (Hastie ve ark. 2008).



(a) (b)

Şekil 3.1. LASSO (a) ve Ridge Regresyona (b) ait tahminlerin grafiksel gösterimi (Hastie ve ark. 2008).

Yukarıdaki şekilde LASSO için kısıtlı alan $|\beta_1|+|\beta_2| \le t$ bir eşkenar dörtgen, ridge regresyon için ise $\beta_1^2 + \beta_2^2 \le t$ olup $\sqrt{t}$ yarıçaplı bir daire şeklindedir. Her iki yöntem de eliptik eğrilerin kısıtlama bölgesine çarptığı ilk noktayı bulur. LASSO da çözümün köşede olması durumunda sıfıra eşit bir parametre bulunur. Ancak ridge regresyonun köşeleri olmadığından parametreler sıfıra eşit olmaz (Hastie ve ark. 2008).

3.3.2.1. Least Angle Regression (LARS)

LASSO düzenlemesi bir model seçim yöntemi olduğundan, yanıt değişkenin uygun bir şekilde tahmin edilmesi için modele girecek en uygun değişkenleri hedefler.

Bunun için de mevcut olan geniş bir açıklayıcı değişkenler kümesinden, yanıt değişkenin en etkili şekilde tahmin edilmesinde kullanılacak daha küçük bir kümenin seçimi ile yapılır. Efron ve ark (2004), LASSO problemini çözmek için, geleneksel ileri seçim yöntemlerinin daha kullanışlı ve daha az değişkeni modele alan formu olan Least Angle Regression (LARS) u geliştirdi. LARS algoritması nın basit bir modifikasyonu, mutlak regresyon katsayılarının toplamını sınırlayarak sıradan en küçük kareler yöntemine göre daha cazip olan LASSO problemini başarılı bir şekilde çözer. LARS modifikasyonu, belirli bir problem için olası bütün LASSO tahminlerini, önceki yöntemlerden daha az bilgisayar zamanı büyüklüğünde bir sıra kullanarak hesaplar. LASSO problemini çözmek için LARS algoritması aşağıdaki adımlarda gerçekleşir:

***Adım*1.** Açıklayıcı değişken değerleri sıfır ortalama ve birim uzunluğa sahip olacak şekilde standartlaştırılır. $\boldsymbol{R} = \boldsymbol{Y} - \bar{\boldsymbol{Y}}$, $\beta_1 = 0$, $\beta_2 = 0, ..., \beta_k = 0$ alınarak başlanır.

***Adım*2.** $\boldsymbol{R}$ ile ençok ilişkili açıklayıcı değişken $\boldsymbol{X}_j$ bulunur.

***Adım*3.** Bazı diğer $\boldsymbol{X}_k$ açıklayıcı değişkenlerinin $\boldsymbol{R}$ ile $\boldsymbol{X}_j$ kadar ilişkili (Korelasyon) oluncaya kadar, β_j 0 noktasından, en küçük kareler katsayısı $< \boldsymbol{X}_j, \boldsymbol{R} >$ ye doğru taşınır.

***Adım*4.** Bazı diğer $\boldsymbol{X}_l$ açıklayıcı değişkenlerinin mevcut kalan ile $\left(\boldsymbol{X}_j, \boldsymbol{X}_k\right)$ kadar ilişkili oluncaya kadar β_j ve β_k değerleri, $\left(\boldsymbol{X}_j, \boldsymbol{X}_k\right)$ üzerindeki mevcut kalanların ortak en küçük kareler katsayısı olarak tanımlanan yönde hareket ettirilir.

***Adım*5.** Bütün (k) açıklayıcı değişkenler kontrol edilinceye kadar bu şekilde işleme devam edilir. $\text{Min}\left(N-1, k\right)$ adımdan sonra en küçük kareler çözümünün tamamına ulaşılır (Hastie ve ark. 2001, Efron ve ark. 2004).

3.3.2.2. Lasso Ceza Parametresinin Seçimi

LASSO problemindeki ceza parametresi ψ'nin 'yanıt değerlerini en yüksek doğrulukla tahmin edilmesini sağlayacak', değeri bir veri kümesi verildiğinde Ridge regresyonda olduğu gibi çapraz doğrulama yöntemi ile elde edilebilir (James ve ark.

2013, Hastie ve ark. 2015, Efron ve ark. 2016). ψ'nin çok küçük değerleri, model veri kümesindeki gürültüyü tanımlamaya yöneldiğinde aşırı uyuma neden olur (düşük yanlılık yüksek varyans). Aksine, ψ'nin çok büyük değerleri, 'kullanılan yöntemin bağımlı değişken ve bağımsız değişken arasındaki ilişkiyi yakalayamadığı durumda' yetersiz uyuma neden olur. Ceza parametresi ψ'nin en uygun değerinin belirlenmesinde çapraz doğrulama yöntemi kullanılırken, olası bir aşırı uyumu engellemek ve modelin daha önce görmediği veri kümesinde üzerinde nasıl bir başarı gösterdiğini saptayabilmek için, ilk adımda veri kümesi eğitim ve test kümesi olarak iki alt gruba ayrılır. İkinci adımda eğitim kümesi kullanılarak regresyon modelindeki katsayılar hesaplanır. Üçüncü adımda ise, ikinci adımda elde edilen tahminler test kümesine uygulanarak başarıları değerlendirilir. Ancak, oluşturulacak model için eğitim ve test aşamasındaki dağılım kaynaklı bazı hatalar ortaya çıkabilir. Bu nedenle, eğitim kümesi eşit uzunluklu M bloğa ayrılır. Bu bloklardan $M-1$ tanesi eğitim kümesini oluştururken 1 tanesi test kümesi olarak göz önüne alınır. Uygulamada blok sayısı M genellikle 5 veya 10 olarak seçilir. Daha sonra $\psi = [\psi_v]$ değerlerinin bir ızgarasını seçilir ve her bir ψ_v değeri için regresyon katsayıları β_j ler hesaplanır. Bu katsayılar kullanılarak, m test kümesi olarak kullanılan bloğun indisi olmak üzere

$$RSS_{\psi_v}^{m}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{k} \hat{\beta}_j(m,\psi_v) x_{ij} \right)^2 \tag{3.54}$$

hata kareleri toplamı hesaplanır. Bu işlem M defa tekrarlanır ve her defasında ilgili test kümesi için $RSS_{\psi_v}^{m}(\boldsymbol{\beta})$ değerleri hesaplanır. Daha sonra hesaplanan $RSS_{\psi_v}^{m}(\boldsymbol{\beta})$ değerlerinin ortalaması

$$MSE_{\psi_v}(\boldsymbol{\beta}) = \frac{1}{M}\sum_{m=1}^{M} RSS_{\psi_v}^{m}(\boldsymbol{\beta}), \tag{3.55}$$

hesaplanır. MSE_{ψ_v}'nı en küçük veren ψ_v değeri ψ olarak alınır. ψ nin elde edilmesi için diğer bir popüler yaklaşım ise "bir- standart hata kuralı" dır (Dettling ve ark. 2004). Bu kurala göre, her MSE_{ψ_v} için ortalamanın standart hatası hesaplanır. Daha sonra, MSE_{ψ_v} değeri en küçük MSE değerinin bir standart hatası içinde olan en büyük ψ_v

değeri seçilir. Bu şekilde, MSE değeri birden fazla standart hatayla arttırılmazken daha düzenli bir model elde edilir.

3.3.3. Naïve Elastik Net (NEN)

LASSO tahmin edicileri birçok katsayıyı sıfıra eşitlediğinden daha kolay yorumlanabilir modeller sunar. Bu durum veri analizinde Ridge tahmin edicilerine göre daha büyük avantaj sağlar ve LASSO tahmin edicilerinin daha başarılı olduğunu gösterir. Ancak, Zou ve Hastie (2005) tarafından önerilen elastik net (EN), LASSO gibi bir değişken seçimi yapan bir düzenleme yöntemidir. Bu yöntem tasarım matrisinin seyrekliğini kullanarak, LASSO dan daha iyi sonuçlar sunmakla birlikte, kuvvetli bir şekilde ilişkili olan tahmin edicilerin modelin içinde veya dışında olma eğiliminde olduğu bir gruplama etkisini teşvik eder (Zou ve Hastie 2005). EN, özelliklede LASSO nun yetersiz kaldığı aşağıdaki durumlarda oldukça faydalıdır:

i) $(k > n)$ durumunda, LASSO konvex en iyileme probleminin doğası nedeniyle, en fazla n açıklayıcı değişkeni seçer. Bu da açıklayıcı değişken seçim yöntemi için sınırlayıcı bir özelliktir ve katsayıların L_1 normu üzerindeki sınır belirli bir değerden küçük olmadığı sürece LASSO iyi tanımlı değildir.

ii) İkili korelasyonların çok yüksek olduğu bir grup değişken varsa, LASSO gruptan yalnızca bir değişkeni seçme eğilimindedir ve hangisinin seçildiği ile ilgilenmez.

iii) Genel $(n > k)$ durumlarında, açıklayıcı değişkenler arasında yüksek korelasyonlar varsa, LASSO nun tahmin performansının ridge regresyonu tarafından bastırıldığı deneysel olarak gözlemlenmiştir (Tibshirani 1996).

Bu düzenlemenin hesaplanması için, Zou ve Hastie (2005), LASSO'nun hesaplanmasında kullanılan En Küçük Açı Regresyonu (LARS) algoritmasını önermiş ve bu algoritmayı LARS-EN olarak adlandırmıştır. EN tahmin edicisi, yeni bir EN penaltısı kullanılarak elde edilen cezalandırılmış en küçük kareler yöntemi olan naïve elastik net (NEN) kriterinden faydalanılarak elde edilir. Bunun için, konum ve ölçek dönüşümünden aşağıda verilen merkezileştirme ve standartlaştırma işlemlerinin yapıldığı kabul edilir:

$$\sum_{i=1}^{n} y_i = 0,\ \sum_{i=1}^{n} x_{ij} = 0,\ \sum_{i=1}^{n} x_{ij}^2 = 1 \left(j = 1,2,...,k \right) \tag{3.56}$$

Negatif olmayan δ_1 ve δ_2 sabitleri için NEN tahmin edicisi $\hat{\boldsymbol{\beta}}_{\text{NEN}}$,

$$\min_{\boldsymbol{\beta}} RSS(\boldsymbol{\beta}) + \delta_1 \sum_{j=1}^{k} |\beta_j| + \delta_2 \sum_{j=1}^{k} \beta_j^2 \tag{3.57}$$

eniyileme probleminin çözümünden elde edilir (Zou ve Hastie 2005, Bai ve Ng 2007). (3.57) eşitliğinde de görüldüğü gibi, NEN eniyileme problemi, LASSO ve Ridge problemlerinin bir birleşimidir ve cezalandırılmış en küçük kareler yöntemi olarak göz önüne alınabilir. Eğer $\tau = \delta_2 / (\delta_1 + \delta_2)$ olarak alınırsa, eşdeğer olarak $\hat{\boldsymbol{\beta}}_{\text{NEN}}$,

$$\begin{aligned} &\min_{\boldsymbol{\beta}} RSS(\boldsymbol{\beta}) \\ &K.S \quad (1-\tau) \sum_{j=1}^{k} |\beta_j| + \tau \sum_{j=1}^{k} \beta_j^2 \le t,\ t \in \mathbb{R}, \end{aligned} \tag{3.58}$$

kısıtlı eniyileme probleminin çözümünden elde edilir. Burada, $(1-\tau) \sum_{j=1}^{k} |\beta_j| + \tau \sum_{j=1}^{k} \beta_j^2$ toplamı NEN penaltısı olarak adlandırılır ve görüldüğü gibi LASSO ve Ridge penaltılarının konveks birleşimidir. (3.57) Eşitliğinde $\tau = 1$ olduğunda, NEN problemi ridge problemine dönüşür. NEN i içeren çalışmalar genellikle $\tau < 1$ durumunu göz önüne alır. Bütün $\tau \in [0,1)$ değerleri için EN penaltı fonksiyonu 0 noktasında tekildir (Türevi hemen-hemen her yerde sıfır olan ve sabit olmayan fonksiyon) ve $\tau > 0$ içinstrictly (kesinlikle) konveks (ikinci türevi pozitiftir) dir. Böylece hem LASSO'nun hem de Ridge'in özelliklerine sahiptir. $\tau = 0$ için LASSO konvekstir ancak strictly (kesinlikle) konveks değildir (Zou and Hastie 2005).

(3.57) eniyileme probleminin çözümünü elde etmek için LASSO probleminin çözüm yönteminden faydalanılır. Çünkü (3.57) nin minimize edilmesi LASSO problemine eşdeğerdir. Bunun için aşağıdaki lemma kullanılarak yeni veri kümesi oluşturulur.

Lemma1 $(\boldsymbol{Y}, \boldsymbol{X})$ veri kümesi ve (δ_1, δ_2) sabitleri verildiğinde,

$$\boldsymbol{X}^*_{(n+k)\times k} = (1+\delta_2)^{-1/2} \begin{pmatrix} \boldsymbol{X} \\ \sqrt{\delta_2}\mathbf{I}_k \end{pmatrix},\quad \boldsymbol{Y}^*_{(n+k)} = \begin{pmatrix} \boldsymbol{Y} \\ \mathbf{0}_k \end{pmatrix}, \tag{3.59}$$

olacak şekilde yapay veri kümesi $\left(\boldsymbol{Y}^{*}, \boldsymbol{X}^{*}\right)$ı tanımlayalım. Eğer

$$\eta = \frac{\delta_1}{\sqrt{1+\delta_2}}, \quad \boldsymbol{\beta}^{*} = \sqrt{1+\delta_2}\,\boldsymbol{\beta}, \tag{3.60}$$

olarak alınırsa, (3.57) eşitliğindeki minimum yapılacak amaç fonksiyonu,

$$\begin{aligned} RSS(\boldsymbol{\beta}) + \delta_1 \sum_{j=1}^{k} \left|\beta_j\right| + \delta_2 \sum_{j=1}^{k} \beta_j^2 &= RSS\left(\boldsymbol{\beta}^{*}\right) + \eta \sum_{j=1}^{k} \left|\beta_j^{*}\right| \\ &= \left\|\boldsymbol{Y}^{*} - \boldsymbol{X}^{*}\boldsymbol{\beta}^{*}\right\|^2 + \eta\left|\boldsymbol{\beta}^{*}\right|_1 \end{aligned} \tag{3.61}$$

olarak yazılabilir. Burada $\sum_{j=1}^{k}\left|\beta_j^{*}\right| := \left|\boldsymbol{\beta}^{*}\right|_1$ olarak alınmıştır. $\boldsymbol{\beta}^{*}$ ın LASSO tahmini $\boldsymbol{\beta}^{*}_{\text{LASSO}}$,

$$\hat{\boldsymbol{\beta}}^{*}_{\text{LASSO}} = \arg\min_{\boldsymbol{\beta}^{*}} \left\|\boldsymbol{Y}^{*} - \boldsymbol{X}^{*}\boldsymbol{\beta}^{*}\right\|^2 + \eta\left|\boldsymbol{\beta}^{*}\right|_1, \tag{3.62}$$

ve

$$\hat{\boldsymbol{\beta}}_{\text{NEN}} = \frac{1}{\sqrt{1+\delta_2}}\hat{\boldsymbol{\beta}}^{*}_{\text{LASSO}}, \tag{3.63}$$

dır (Zou and Hastie 2005).

(3.60) de yapılan dönüşümlerin (3.57) içinde kullanılması ve basit cebirsel işlemlerin yapılması ile Lemma1 ispatlanabilir. Lemma1, NEN probleminin, genişletilmiş verilere dayalı olarak eşdeğer bir LASSO problemine dönüştürülebileceğini ifade eder. (3.59) eşitliğinde de görüleceği gibi genişletilmiş LASSO probleminde kullanılan örneklem büyüklüğü $n+k$ dır ve açıklayıcı değişkenler matrisi $\text{rank}\left(\boldsymbol{X}^{*}\right) = k$ dır. $\text{rank}\left(\boldsymbol{X}^{*}\right) = k$ olması, NEN in her durumda bütün k açıklayıcı değişkenlerini seçebileceğini gösterir. Buda, bir değişken seçim yöntemi olan LASSO probleminin yukarıda *i*) ile verilen sınırlayıcı olma özelliğini ortadan kaldırır. Ayrıca Lemma 1, NEN in LASSO ya benzer bir şekilde otomatik değişken seçimi yapabildiğini gösterir.

Yüksek korelasyonlu değişkenlerden oluşan bir grubun regresyon katsayıları eşit olma eğilimindeyse, regresyon yöntemi gruplama etkisini gösterir (negatif korelasyon varsa işaret değişikliği kadar). Özellikle, bazı değişkenlerin tam olarak aynı olduğu sıra

dışı bir durumda, regresyon yöntemi aynı değişkenlere aynı katsayıları atamalıdır. (Zou and Hastie 2005) çalışmalarında, NEN in LASSO nun sahip olmadığı gruplanmış değişkenleri seçme özelliği olduğunu da göstermişlerdir. Çünkü açıklayıcı değişken sayısının büyük, örneklem sayısının küçük $(k > n)$ olduğu durumlarda "gruplanmış değişkenler" in durumu oldukça önemlidir ve endişe vericidir. Bu nedenle $(k > n)$ durumu ele alan çok sayıda çalışma literatürde mevcuttur (Hastie ve ark. 2000, Diaz-Uriarte 2003, Hastie ve ark. 2003, Segal ve Conklin 2003, Dettling and Bühlmann 2004).

3.3.4. Elastik Net (EN)

Otomatik değişken seçim yöntemi olarak, NEN, LASSO nun sahip olduğu (*i*) ve (*ii*) sınırlamalarının üstesinden gelir. Bununla birlikte Zou ve Hastie (2005), deneysel çalışmalarda NEN in ridge regresyona veya LASSO ya çok yakın olmadıkça tatmin edici performans vermediğini göstermiştir. Bu nedenle naïve (Toy-Saf) olarak adlandırılır. Regresyon çalışmalarında, doğru bir ceza yöntemi, yanlılık ve varyans arasındaki alış-veriş ile iyi bir öngörü performansı sağlar. NEN tahmin yöntemi iki adımlı bir yöntemdir. İlk adımda, her δ_2 sabiti için önce ridge regresyon katsayıları bulunur. İkinci adımda ise, LASSO katsayısı çözüm yolları boyunca küçülme yapılır. Bu da, model için iki kat küçülme yapıldığını ifade eder. Bu iki kat küçültme varyansı azaltmadığı gibi LASSO ve Ridge göre fazladan gereksiz bir yanlılık miktarı sunar. NEN in bu iki katlı küçültmesini ortadan kaldırmak ve öngörü performansını artırmak için Zou ve Hastie (2005) tarafından Elastik Net (EN) geliştirilmiştir. $(\boldsymbol{Y}, \boldsymbol{X})$ veri kümesi ve (δ_1, δ_2) sabitleri verildiğinde, Lemma1 de tanımlanan $(\boldsymbol{Y}^*, \boldsymbol{X}^*)$ genişletilmiş veri kümesi için NEN tahmin edicisi

$$\hat{\boldsymbol{\beta}}^*_{\text{LASSO}} = \arg\min_{\boldsymbol{\beta}^*} \left\| \boldsymbol{Y}^* - \boldsymbol{X}^* \boldsymbol{\beta}^* \right\|^2 + \frac{\delta_1}{\sqrt{1+\delta_2}} \left| \boldsymbol{\beta}^* \right|_1 \tag{3.64}$$

probleminin çözümünden faydalanılarak (3.63) eşitliği ile elde edilmişti. EN tahmin edicisi $\hat{\boldsymbol{\beta}}_{\text{EN}}$, NEN tahmin edicisi $\hat{\boldsymbol{\beta}}_{\text{NEN}}$ nin yukarıda açıklanan yanlılığa ilişkin olumsuzluğu ortadan kaldıran düzeltilmiş formudur ve Zou ve Hastie (2005) tarafından

$$\hat{\boldsymbol{\beta}}_{\text{EN}} = \sqrt{1+\delta_2}\,\hat{\boldsymbol{\beta}}^{*}_{\text{LASSO}}, \tag{3.64}$$

olarak tanımlanır. (3.63) Eşitliğinden

$$\sqrt{1+\delta_2}\,\hat{\boldsymbol{\beta}}_{\text{NEN}} = \hat{\boldsymbol{\beta}}^{*}_{\text{LASSO}} \tag{3.65}$$

olduğu göz önüne alınarak, (3.64) de kullanılması ile $\hat{\boldsymbol{\beta}}_{\text{EN}}$,

$$\hat{\boldsymbol{\beta}}_{\text{EN}} = (1+\delta_2)\hat{\boldsymbol{\beta}}_{\text{NEN}} \tag{3.66}$$

olarak elde edilir ki, bu da EN tahmin edicisinin yeniden ölçeklenmiş NEN tahmin edicisi olduğunu gösterir. (3.66) Eşitliği ile verilen yeniden ölçekleme NEN in açıklayıcı değişken seçim özelliğini korur ve gereksiz yapılan küçülmeyi ortadan kaldırır. Dolayısı ile NEN in bütün iyi özellikleri EN inde geçerli olacağından, performans açısından EN de ridge ve LASSO ya göre daha iyi olacaktır. Gereksiz küçülmeyi ortadan kaldıran ölçeklendirme faktörünün $1+\delta_2$ seçilmesinin bir başka nedeni ise açıklayıcı değişken vektörlerinin dik olması durumunda LASSO nun en iyiliğinin minimum olduğu Donoho ve ark. (1995) tarafından saptanmıştır. Bu da NEN in en iyi (optimal) olmadığını gösterir. $1+\delta_2$ ölçeklemesinden sonra EN otomatik olarak minimax en iyiliğe ulaşır (Zou ve Hastie 2005).

3.3.4.1. Least Angle Regression-EN (LARS-EN)

EN probleminin çözümü için Zou ve Hastie (2005) tarafından LARS-EN olarak adlandırılan bir algoritma önerilmiştir. Zou ve Hastie yaptıkları simülasyon çalışmalarında bu algoritmanın EN probleminin çözümünde çok etkili olduklarını saptamışlardır. Bu algoritma 3.3.2.1 alt başlığında açıklanan, Efron ve arkadaşları tarafından (2004)'te geliştirilen LARS algoritmasına dayalı olarak geliştirilmiştir. Kesim 4.3 teki Lemma1 ile her sabit δ_2 değeri için EN problemi artırılmış veri kümesi üzerinde bir LASSO problemine eşdeğerdir. Bu nedenle, LARS algoritması, tek bir sıradan en küçük kareler uyumunun hesaplama işlemleri ile EN in çözüm yolunu verimli bir şekilde oluşturmak için doğrudan kullanılabilir. Bununla birlikte, Zou ve Hastie (2005), $k > n$ durumunda, artırılmış veri kümesinin $k+n$ sayıda gözlem içermesi nedeni ile LARS ın EN probleminin çözümünde hesaplamaların önemli ölçüde yavaşladığını belirtmişlerdir. Bu durumda, Efron (2004)'ten faydalanarak artırılmış

açıklayıcı değişken matrisi X^* seyrek yapısını kullanarak LARS-EN için hesaplamaların daha kolay ve hızlı yapılmasını sağlamışlardır.

3.3.4.2. LARS-EN İçin Ceza Parametrelerinin Seçimi

Artırılmış veri kümesine ilişkin oluşturulan ceza parametrelerinin seçimi için, çeşitli yöntemler vardır (Hastie ve ark. 2001). Sadece eğitim veri kümesinin mevcut olduğu durumda, çok bilinen ve kullanılan *On-Kat Çapraz Doğrulama* (CV) yöntemi farklı modelleri karşılaştırmak ve tahmin hatasını tahmin etmek ve için popüler bir yöntemdir. Ancak EN probleminde δ_1 ve δ_2 gibi iki penaltı parametresi bulunduğundan, iki boyutlu bir yüzeyde çapraz doğrulamaya ihtiyaç duyulur. Bunun için önce, δ_2 için (nispeten küçük) bir değer ızgarası seçilir, daha sonra her δ_2 değeri için LARS-EN algoritması EN için tüm çözüm yolunu üretir. Diğer ceza parametresi δ_1 on-kat çapraz doğrulama ile seçilir. Seçilen δ_2 en küçük CV hatasını vermelidir.

3.3.5. LIU Küçültme Yöntemi

Liu (1993), doğrusal regresyonda iç ilişki olması durumunda, ridge ve stein tahmincilerinin avantajlı yönlerini birleştirip bu iki küçültme yönteminin dezavantajlarını ortadan kaldıran Liu küçültme yöntemini önermiştir. Liu küçültme yöntemi, iç ilişkinin tam ya da tama yakın olduğu durumlarda başarılı sonuçlar vermekle birlikte ceza parametresi d'nin doğrusal fonksiyonudur. (3.5) eşitliği ile verilen doğrusal regresyon modeli için $\boldsymbol{\beta}$ nın liu tahmin edicisi $\hat{\boldsymbol{\beta}}_L$, Liu (1993) tarafından $(-\infty < d < 1)$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_L &= \left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_k\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{Y} + d\hat{\boldsymbol{\beta}}\right) \\ &= \left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_k\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X} + d\mathbf{I}_k\right)\hat{\boldsymbol{\beta}}\end{aligned} \tag{3.67}$$

olarak verilmiştir. Burada ceza parametresi d, $(0 < d < 1)$ aralığında değer alır (Liu 1993). Ancak daha sonra yapılan çalışmalarda d'nin 0'dan küçük veya 1'den büyük değer de alabileceği belirtilmiştir (Liu 2011). Liu tahmin edicisi $\hat{\boldsymbol{\beta}}_L$, $d = 1$ için en küçük kareler tahmin edicisi $\hat{\boldsymbol{\beta}}$'ya eşit olduğundan, $\hat{\boldsymbol{\beta}}_L$ $\hat{\boldsymbol{\beta}}$'nın doğrusal fonksiyonu

olarak yazılabilir (Liu 1993). Bu nedenle ceza parametresi d'nin seçimi, ridge küçültme yöntemindeki K'nın seçimine göre daha kolaydır.

$X^T Y = \left(X^T X\right)\hat{\boldsymbol{\beta}}$ eşitliği (3.67) de yerine yazılırsa,

$$\hat{\boldsymbol{\beta}}_L = \left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\hat{\boldsymbol{\beta}} \tag{3.68}$$

eşitliği elde edilir. (3.68) Eşitliğinde $\boldsymbol{T} = \left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)$ olarak yazılırsa,

$$\hat{\boldsymbol{\beta}}_L = \boldsymbol{T}\hat{\boldsymbol{\beta}} \tag{3.69}$$

eşitliği elde edilir. Liu tahmin edicisi $\hat{\boldsymbol{\beta}}_L$ nin beklenen değeri, EKK tahmin edicisinin yansız olması nedeni ile

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{\beta}}_L\right) &= \mathrm{E}\left(\boldsymbol{T}\hat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{T}\boldsymbol{\beta} \\ &= \left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\boldsymbol{\beta} \end{aligned} \tag{3.70}$$

olarak elde edilir. (3.70) Eşitliğinde de görüldüğü gibi $\mathrm{E}\left(\hat{\boldsymbol{\beta}}_L\right) \neq \boldsymbol{\beta}$ dır. Yani, Liu tahmin edicisi $\hat{\boldsymbol{\beta}}_L$, $\boldsymbol{\beta}$ nın yanlı tahmin edicisidir. (3.70) Eşitliğinin göz önüne alınması ile $\hat{\boldsymbol{\beta}}_L$ tahmin edicisinin yanlılık miktarı,

$$\begin{aligned} \mathrm{Yan}\left(\hat{\boldsymbol{\beta}}_L\right) &= \mathrm{E}\left(\hat{\boldsymbol{\beta}}_L\right) - \boldsymbol{\beta} \\ &= \left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\boldsymbol{\beta} - \boldsymbol{\beta} \\ &= \left[\left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right) - \mathbf{I}_k\right]\boldsymbol{\beta}, \end{aligned} \tag{3.71}$$

olarak elde edilir. (3.12) ve (3.68) eşitliklerinin göz önüne alınması ile Liu tahmin edicisi $\hat{\boldsymbol{\beta}}_L$ nin kovaryans matrisi $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_L)$,

$$\begin{aligned} \mathrm{Cov}(\hat{\boldsymbol{\beta}}_L) &= \mathrm{Cov}\left[\left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\hat{\boldsymbol{\beta}}\right] \\ &= \mathrm{Cov}\left[\left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\left(X^T X\right)^{-1} X^T Y\right] \\ &= \left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\left(X^T X\right)^{-1} X^T \mathrm{Cov}(Y) X(\left(X^T X\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\left(X^T X + \mathbf{I}_k\right)^{-1} \\ &= \sigma^2\left(X^T X + \mathbf{I}_k\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\left(X^T X\right)^{-1}\left(X^T X + d\mathbf{I}_k\right)\left(X^T X + \mathbf{I}_k\right)^{-1} \end{aligned} \tag{3.72}$$

olarak elde edilir. (3.71) ve (3.72) eşitliklerinin kullanılması ile $\hat{\boldsymbol{\beta}}_L$ için hata karelerinin ortalaması $\mathrm{MSE}(\hat{\boldsymbol{\beta}}_L)$,

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\beta}}_L) &= \mathrm{izCov}\left(\hat{\boldsymbol{\beta}}_L\right) + \left\|\mathrm{E}\left(\hat{\boldsymbol{\beta}}_L\right) - \boldsymbol{\beta}\right\|^2 \\ &= \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_{k+1}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X} + d\mathbf{I}_{k+1}\right)\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X} + d\mathbf{I}_{k+1}\right)\left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_{k+1}\right)^{-1} \\ &\quad + \left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_{k+1}\right)^{-1}\left(d\mathbf{I}_p - \mathbf{I}_{k+1}\right)\boldsymbol{\beta}\boldsymbol{\beta}^T\left(d\mathbf{I}_p - \mathbf{I}_{k+1}\right)\left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_{k+1}\right)^{-1},\end{aligned} \tag{3.73}$$

olarak elde edilir. Liu (1993), $\hat{\boldsymbol{\beta}}_L$ tahmin edicisinin genelleştirilmiş formunu $0 < d_i < 1$ ve $i = 1, 2, ..., k+1$ olmak üzere;

$$\hat{\boldsymbol{\beta}}_{GL} = \left(\boldsymbol{X}^T\boldsymbol{X} + \mathbf{I}_p\right)^{-1}\left[\boldsymbol{X}^T\mathbf{Y} + \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & d_{k+1} \end{pmatrix}\hat{\boldsymbol{\beta}}\right], \tag{3.74}$$

şeklinde elde etmiştir. Yukarıda $\boldsymbol{d}$ ceza parametresi matris formundadır.

3.3.5.1. Liu Ceza Parametresinin Seçimi

Ceza parametresinin seçimini kolaylaştırmak için kanonik model kullanılır. $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_k \geq 0$ Olmak üzere λ_i ler $\boldsymbol{X}^T\boldsymbol{X}$ matrisinin in öz değerleri olsun. $\boldsymbol{Q} = \left(q_1, q_2, ..., q_k\right)$ bu özdeğerlere ait öz vektörler matrisi,

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_k \end{pmatrix}$$

şeklinde ifade edilirse, $\hat{\boldsymbol{\beta}}_L$ tahmin edicisi kanonik model kullanılarak yeniden yazılabilir (Liu 1993). $\boldsymbol{Z} = \boldsymbol{XQ}$ ve $\boldsymbol{\alpha} = \boldsymbol{Q}^T\boldsymbol{\beta}$ olmak üzere

$$\begin{aligned}\boldsymbol{Z}^T\boldsymbol{Z} &= \left(\boldsymbol{XQ}\right)^T\left(\boldsymbol{XQ}\right) \\ &= \boldsymbol{Q}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)\boldsymbol{Q} \\ &= \boldsymbol{\Lambda}\end{aligned} \tag{3.75}$$

şeklinde ifade edilirse, $\boldsymbol{\alpha}$ nın en küçük kareler tahmin edicisi,

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^{T}\boldsymbol{y} \tag{3.76}$$

olarak, Liu tahmin edicisi ise

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{L} &= \left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\hat{\boldsymbol{\alpha}} \\ &= \left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^{T}\boldsymbol{Y}\end{aligned} \tag{3.77}$$

olarak yazılır. Liu tahmin edicisi $\hat{\boldsymbol{\alpha}}_{L}$ nin beklenen değeri,

$$\begin{aligned}E\left(\hat{\boldsymbol{\alpha}}_{L}\right) &= E\left[\left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\hat{\boldsymbol{\alpha}}\right] \\ &= \left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\hat{\boldsymbol{\alpha}}\end{aligned} \tag{3.78}$$

olarak elde edilir. Eğer, $Cov\left(\hat{\boldsymbol{\alpha}}_{L}\right)$,

$$\mathrm{Cov}\left(\hat{\boldsymbol{\alpha}}_{L}\right) = \sigma^{2}\left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\left(\boldsymbol{\Lambda}+d\mathbf{I}_{k}\right)\left(\boldsymbol{\Lambda}+\mathbf{I}_{k}\right)^{-1}, \tag{3.79}$$

değeri ortalama hata kareleri,

$$\mathrm{MSE}(\hat{\boldsymbol{\alpha}}_{L}) = iz\mathrm{Cov}\left(\hat{\boldsymbol{\alpha}}_{L}\right) + \left\|E\left(\hat{\boldsymbol{\alpha}}_{L}-\boldsymbol{\alpha}\right)\right\|^{2}, \tag{3.80}$$

ifadesinde yerine yazılırsa, Liu tahmin edicisi $\hat{\boldsymbol{\alpha}}_{L}$ için ortalama hata kareleri,

$$\mathrm{MSE}\left(\hat{\boldsymbol{\alpha}}_{L}\right) = \sigma^{2}\sum_{i=1}^{k}\frac{\left(\lambda_{i}+d\right)^{2}}{\lambda_{i}\left(\lambda_{i}+1\right)^{2}} + \left(d-1\right)^{2}\sum_{i=1}^{k}\frac{\alpha_{i}^{2}}{\left(\lambda_{i}+1\right)^{2}} \tag{3.81}$$

olarak elde edilir. Ceza parametresi d nin seçiminde, Ridge küçültme yönteminde olduğu gibi amaç MSE değerinin minimum yapılmasıdır (Liu 1993). (3.81) Eşitliğinin d sabitine göre türevi,

$$\frac{\partial \mathrm{MSE}\left(\hat{\boldsymbol{\alpha}}_{L}\right)}{\partial d} = \mathrm{g}'\left(d\right)$$

ile gösterilirse, bu türev

$$g'(d) = 2\sigma^{2}\sum_{i=1}^{k}\frac{\left(\lambda_{i}+d\right)}{\lambda_{i}\left(\lambda_{i}+1\right)^{2}} + 2\left(d-1\right)\sum_{i=1}^{k}\frac{\alpha_{i}^{2}}{\left(\lambda_{i}+1\right)^{2}}, \tag{3.82}$$

olarak elde edilir ve

$$g'(1) = 2\sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i (\lambda_i + 1)} > 0, \tag{3.83}$$

bulunur. Böylece $g(d) < g(1)$ olacak şekilde bir $0 < d < 1$ sayısı vardır veya eşdeğer olarak $\text{MSE}(\hat{\boldsymbol{\alpha}}_L) < \text{MSE}(\hat{\boldsymbol{\alpha}})$ olarak elde edilir (Liu 1993).

Liu (1993), $\text{MSE}(\hat{\boldsymbol{\alpha}}_L)$ değerini minimum yapan ceza parametresi d, (3.82) eşitliğinden faydalanarak,

$$d = \sum_{i=1}^{k} \frac{\alpha_i^2 - \sigma^2}{(\lambda_i + 1)^2} \Bigg/ \sum_{i=1}^{k} \frac{\sigma^2 + \lambda_i \alpha_i^2}{\lambda_i (\lambda_i + 1)^2} \ . \tag{3.84}$$

olarak belirlemiştir. (3.84) İfadesinde α_i^2 ve σ^2 nin yerlerine sırası ile yansız tahmini değerleri $\hat{\alpha}_i^2 - \dfrac{\hat{\sigma}^2}{\lambda_i}$ ve $\hat{\sigma}^2$ kullanılırsa, d nin tahmini $\hat{d}$,

$$\hat{d} = 1 - \hat{\sigma}^2 \left[\sum_{i=1}^{k} \frac{1}{\lambda_i (\lambda_i + 1)} \Bigg/ \sum_{i=1}^{k} \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2} \right], \tag{3.85}$$

olarak bulunur. Liu (1993), $\hat{d}$ tahmin edicisinin genelleştirilmiş halini $h > 0$ olmak üzere,

$$\hat{d} = 1 - h\hat{\sigma}^2 \left[\sum_{i=1}^{k} \frac{1}{\lambda_i (\lambda_i + 1)} \Bigg/ \sum_{i=1}^{k} \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2} \right], \tag{3.86}$$

olarak vermiştir.

Mallows (1973)' un ridge ceza parametresi λ nın seçimi için verdiği C_p kriterini Liu(1993), d nin seçimi için uyarlamıştır. Bu amaçla,

$\boldsymbol{H}_d = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{I}_k\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{X} + d\boldsymbol{I}_k\right)\boldsymbol{X}^T$ ve RSS_L, $\hat{\boldsymbol{\beta}}_L$'ye ait kalan kareler toplamı olmak üzere,

$$C_L = \frac{RSS_L}{\hat{\sigma}^2} + 2\text{iz}(\boldsymbol{H}_L) - (n-2) \tag{3.87}$$

eşitliğini tanımlayarak, C_L 'yi minimum yapan ,

$$\hat{d}_{C_L} = 1 - h\hat{\sigma}^2 \left[\sum_{i=1}^{p} \frac{1}{(\lambda_i + 1)} \Big/ \sum_{i=1}^{p} \frac{\lambda_i \hat{\alpha}_i^{\,2}}{(\lambda_i + 1)^2} \right] \tag{3.88}$$

değerini ceza parametresinin ideal değeri olarak kullanmıştır. Liu ceza parametresi *d* nin ideal değeri, Hoerl ve Kennard (1976) ın geliştirdiği yöntem kullanılarak da elde edilebilir. (3.85) eşitliği ile verilen *d*'nin iteratif tahmin edicisi $\hat{\boldsymbol{\alpha}}$ 'nın bazı değerlerine aşağıdaki şekilde elde edilir

(Liu 1993):

$$\hat{\boldsymbol{\alpha}} \qquad d_0 = 1 - \hat{\sigma}^2 \left[\sum_{i=1}^{k} \frac{1}{\lambda_i (\lambda_i + 1)} \Big/ \sum_{i=1}^{k} \frac{\hat{\alpha}_i^{\,2}}{(\lambda_i + 1)^2} \right],$$

$$\hat{\boldsymbol{\alpha}}(d_0) \qquad d_1 = 1 - \hat{\sigma}^2 \left[\sum_{i=1}^{k} \frac{1}{\lambda_i (\lambda_i + 1)} \Big/ \sum_{i=1}^{k} \frac{\hat{\alpha}_{di}^{\,2}(d_0)}{(\lambda_i + 1)^2} \right],$$

$$\hat{\boldsymbol{\alpha}}(d_1) \qquad d_2 = 1 - \hat{\sigma}^2 \left[\sum_{i=1}^{k} \frac{1}{\lambda_i (\lambda_i + 1)} \Big/ \sum_{i=1}^{k} \frac{\hat{\alpha}_{di}^{\,2}(d_1)}{(\lambda_i + 1)^2} \right]$$

$\vdots$

$T = iz\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} / k$ olup, işlem $\left(d_j - d_{j+1}\right)/d_j < 20T^{-1.3}$ elde edilene kadar devam ettirilerek $\boldsymbol{\alpha}$ nın tahmini olarak $\hat{\boldsymbol{\alpha}}_d\left(d_j\right)$ kullanılır.

4. ARAŞTIRMA BULGULARI

Bu bölümde kullanılan veriler https://dasl.datadescription.com/datafile/cost-of-living-2017/?_sfm_methods=Multiple+Regression&_sfm_cases=4+59943 adresinden alınmıştır. Bu veri kümesi, 2017 yılına ait 511 şehrin yiyecek, konaklama masrafları ve tüketim malları fiyatlarına göre bu şehirlerin pahalılık durumlarına ilişkin gözlemlerden oluşmuştur. Bu veri kümesi için yaşam maliyeti bağımlı değişkeni, yaşam maliyeti artı kira indeksi, kira endeksi, bakkaliye endeksi, restoran fiyat endeksi ve yerel satın alma gücü endeksi ise yaşam maliyetinin bağımlı olduğu açıklayıcı değişkenleri göstermektedir. Bu değişkenler arasında iç ilişkinin varlığı araştırılıp, iç ilişki tespit edilmiştir. İç ilişki çözüm yöntemlerinden küçültme yöntemleri (Ridge küçültme yöntemi, Liu küçültme yöntemi, LASSO ve Elastik net) ile problem çözülmeye çalışılmıştır. Bu yöntemleri karşılaştırmak için hata kareler toplamı (RSS), hata kareler ortalaması (MSE) ve R^2 belirleme katsayısı değerleri göz önüne alınmıştır.

Çizelge 4.1 de verilen verilerde yaşam maliyeti artı kira indeksi X_1 bağımsız değişkeni, kira endeksi X_2 bağımsız değişkeni, bakkaliye endeksi X_3 bağımsız değişkeni, restoran fiyat endeksi X_4 bağımsız değişkeni ve yerel satın alma gücü endeksi X_4 bağımsız değişkeni, yaşam maliyeti endeksi ise Y bağımlı değişkeni ile gösterilmek üzere çoklu regresyon modeli ,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \qquad i = 1, 2, ..., n$$

şeklinde yazılır. Bu model için yapılan bütün hesaplamalar için RStudio paket programı kullanılmıştır.

Çizelge 4.1. 2017 Yılında yaşam maliyetini etkileyen etkenler

Y Yaşam Maliyeti Endeksi	X_1 Yaşam Maliyeti Artı Kira Endeksi	X_2 Kira Endeksi	X_3 Bakkaliye Endeksi	X_4 Restoran Fiyat Endeksi	X_5 Yerel Satın Alma Gücü Endeksi
149.53	108.77	66.76	163.71	140.58	126.3
148.13	133.2	117.8	146.2	152.54	93.42
142.54	106.54	67.39	148.36	143.18	106.62

Çizelge 4.1. 2017 Yılında yaşam maliyetini etkileyen etkenler (devamı)

142.12	106.73	70.25	147.13	138.96	116.2
142.02	97.47	51.54	149.54	131.72	113.13
136.51	91.11	46.33	146.72	122.48	101.65
131.38	93.56	54.56	136.95	127.87	112.33
…	…	…	…	…	…
26.4	16.93	6.18	28.26	14.41	60.04
26.24	15.06	4.57	28.87	14.71	52.38
24.83	16.37	6.61	26.31	16.29	47.01
24.54	14.96	5.09	23.5	17.66	21.42
23.17	14.56	6.68	23.88	12.61	96.97
21.75	13.59	6.17	23.72	12.12	62.76

4.1. İç İlişkinin Saptanması

Çoklu lineer regresyonda, küçültme yöntemleri ile tahmin yapabilmek için önceki bölümlerde teorik olarak verdiğimiz yöntemleri kullanarak modelimizde iç ilişki probleminin varlığını inceleyeceğiz. İlk olarak korelasyon matrisini inceleyelim. X=cbind(x1,x2,x3,x4,x5) # X matrisi ve A=cor(X) #korelasyon matrisi R-kodu olmak üzere korelasyon matrisi,

$$X^{\mathrm{T}}X = \begin{bmatrix} 1.0000000 & \mathbf{0.9288298} & \mathbf{0.9217869} & \mathbf{0.9141099} & 0.6595403 \\ 0.9288298 & 1.0000000 & 0.7674361 & 0.7523090 & 0.6000432 \\ 0.9217869 & 0.7674361 & 1.0000000 & \mathbf{0.8518550} & 0.6458339 \\ 0.9141099 & 0.7523090 & 0.8518550 & 1.0000000 & 0.6436926 \\ 0.6595403 & 0.6000432 & 0.6458339 & 0.6436926 & 1.0000000 \end{bmatrix}$$

şeklindedir. Korelasyon matrisinin elemanlarından 1'e yakın olan değerler iç ilişki problemi olduğunu göstermektedir. Dolayısıyla X_1 ve X_2, X_1 ve X_3, X_1 ve X_4, X_3 ve X_4 değişkenleri arasında iç ilişki olduğu görülmektedir. Ayrıca korelasyon matrisinin determinantının 0'a yakın olması da iç ilişki probleminin bir diğer göstergesidir. det(A) #korelasyon matrisinin determinantı olmak üzere,

$$\det\left(\boldsymbol{X}^T\boldsymbol{X}\right) = 0.0004336268$$

dir. İç ilişki problemi konusunda en çok başvurulan yöntem VIF değerleri, korelasyon matrisinin tersi $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ matrisinin köşegen elemanlarına eşittir.

C=solve(A) #korelasyon matrisinin tersi için R-kodu

$$\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = \begin{bmatrix} \mathbf{127.775607} & -60.137405 & -39.976551 & -39.452024 & 3.025067 \\ -60.137405 & \mathbf{31.031856} & 17.680712 & 17.707361 & -1.774313 \\ -39.976551 & 17.680712 & \mathbf{16.853082} & 9.815803 & -1.445698 \\ -39.452024 & 17.707361 & 9.815803 & \mathbf{16.308765} & -1.442192 \\ 3.025067 & -1.774313 & -1.445698 & -1.442192 & \mathbf{1.931520} \end{bmatrix}$$

şeklindedir. Bağımsız değişkenlere ait VIF değerleri Çizelge 4.2.'de verilmiştir.

Çizelge 4.2. VIF değerleri

Değişkenler	VIF
X_1	127.775607
X_2	31.031856
X_3	16.853082
X_4	16.308765
X_5	1.931520

Bağımsız değişkenlerden X_5 hariç diğer değişkenlere ait VIF değerlerinin 10'dan büyük olması iç ilişki probleminin varlığını gösterir.

İç ilişkiyi korelasyon matrisine ait özdeğer ve bu özdeğerlerin özvektörlerinin kullanılmasıyla da saptanabilir.

E=eigen(A) #korelasyon matrisinin özdeğer ve özvektörleri için R-kodu

$\lambda_1 = 4.097061249$, $\lambda_2 = 0.478113005$, $\lambda_3 = 0.271805002$, $\lambda_4 = 0.147499128$, $\lambda_5 = 0.005521616$

ve bu özdeğerlere ait öz vektörler,

$$\boldsymbol{T} = \begin{pmatrix} -0.4863439 & 0.2309168 & 0.07504214 & 0.0061795934 & 0.83933156 \\ -0.4447006 & 0.2636646 & 0.75559153 & 0.0569799783 & -0.39819228 \\ -0.4601036 & 0.1365118 & -0.40156810 & -0.7343853672 & -0.26285052 \\ -0.4572482 & 0.1250099 & -0.50052713 & 0.6763080448 & -0.25957018 \\ -0.3807095 & -0.9180934 & 0.10800818 & 0.0008098218 & 0.02232403 \end{pmatrix}$$

olarak elde edilir. Özdeğerlerin küçük olması yani sıfıra yakın değer alması iç ilişki problemi olduğunu göstermektedir. Bunun yanında koşul sayısının 100'den büyük olması ve koşul endeksinin de 10'dan büyük değer alması iç ilişki probleminin varlığını gösterir.

$$\mathrm{KS} = \frac{\lambda_{\max}}{\lambda_{\min}} = 742.0040$$

$$\mathrm{KI} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = 27.2397$$

Yukarıda birçok yöntemle iç ilişki problemi tespit edilmiştir. Lineer Regresyon modelimiz sırasıyla RR, LASSO, EN ve LIU küçültme yöntemleri kullanılarak tahmin edilmeye çalışılacaktır.

4.2. Lineer Regresyon Modelinin Tahmini

Tüm küçültme yöntemleri, veri kümesinin %70' i eğitim (train), %30' u deneme (test) kümesi olmak üzere rastgele iki parçaya ayrılması ile hesaplanmıştır. Eğitim veri kümesinde çalıştıktan sonra modelimizin güvenirliliğini test etmek için sistemin hesaplamaları yaparken görmediği deneme veri kümesi sonuçları kontrol edilmiştir.

İç ilişki probleminin EKK yöntemi üzerinde ciddi etkileri olmaktadır. Çizelge 4.3.'de incelendiğinde bağımsız değişkenlerin t anlamlılık değerlerinin düşük olduğu görülmektedir. Bu da X_1 değişkeni hariç modelimizdeki diğer değişkenlerin önemli belirleyiciler olmadığını göstermektedir. Lineer regresyon modeli için R programda aşağıdaki kod girilerek EKK sonuçları bulunur.

model=lm(Y~x1+x2+x3+x4+x5), summary(model) #Model özet sonuçları R-kodu.

Çizelge 4.3. Modelin EKK sonuçları

Parametreler	Tahminler	Standart Hata	t değeri
Sabit	-0.08478	0.001366	-47720.129
$\hat{\beta}_1$	0.96065	0.0001824	10797.959
$\hat{\beta}_2$	0.96913	0.00009657	-10042.016
$\hat{\beta}_3$	1.02907	0.00005152	2.291
$\hat{\beta}_4$	1.01826	0.00004633	1.093
$\hat{\beta}_5$	0.93277	0.0000126	-0.865

EKK yöntemiyle elden edilen parametre tahminleri sonucunda regresyon modeli;

$$Y = -0.08478 + 0.96065X_1 + 0.96913\ X_2 + 1.02907X_3 + 1.01826X_4 + 0.93277X_5$$

olarak elde edilir. anova(model) #Varyans analiz tablosu R-kodu.

Çizelge 4.4. Modelin varyans analiz tablosu

Parametreler	F	p
X_1	4.9155e+09	< 2e-16 ***
X_2	4.2972e+08	< 2e-16 ***
X_3	3.9127e+00	0.04847 *
X_4	8.1250e-01	0.36781
X_5	7.4860e-01	0.38733

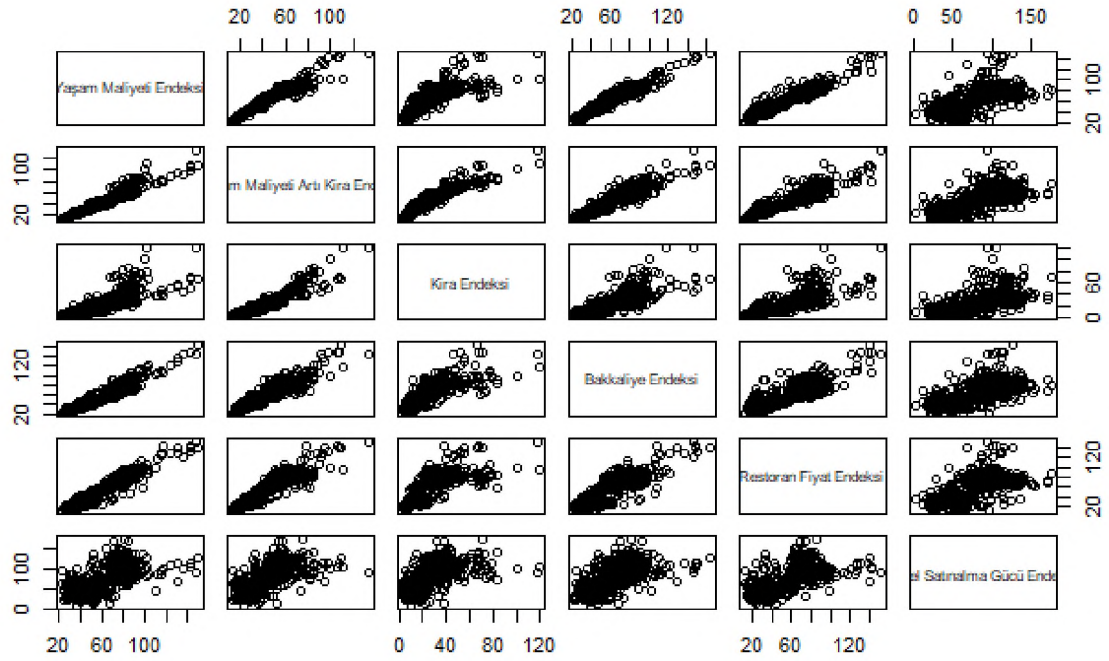
Çizelge 4.4.'de p-değerinin ilk üç değişken için sıfıra yakın olduğu görülmektedir. Bu da bağımsız değişkenlerden X_1, X_2 ve X_3'ün istatistiksek olarak anlamlı olduğu anlamına gelmektedir. Modele ait standart hatanın düşük olması ve parametrelere ait F değerlerinin istatistiksel olarak anlamlı olması elde edilen tahminlerin güvenilir

olduğunun işaretidir ancak parametrelere ait t değerlerinin düşük olması iç ilişki probleminin sonuçlarındadır.

Şekil 4.1'de bağımlı ve bağımsız değişkenlerin matris diyagramları verilmiştir.

plot(model) #Değişkenlerin matris diyagramları kodu ile elde edilir.

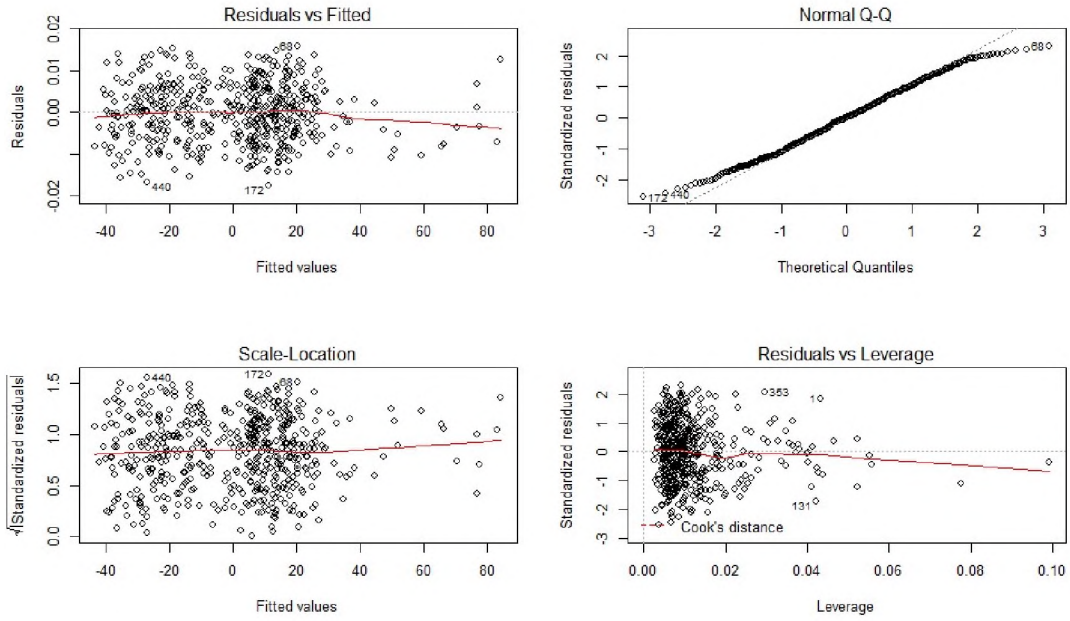
Grafik incelendiğinde X_1 ve X_2, X_1 ve X_3, X_1 ve X_4, X_3 ve X_4 değişkenleri arasında iç ilişki olduğu görülmektedir. Ayrıca grafiğin sağ üst köşesinde sabit olmayan varyans değerleri, X_5 bağımsız değişkeninin anlamsız olduğunu göstermektedir.



Şekil 4.1. Değişkenlerin matris diyagramları

Şekil 4.2.'de verilen ilk grafikte Y gözlem değerlerine karşın verilen artıkların sıfır etrafında toplandığı görülmektedir. Bu model varsayımlarının sağlandığını göstermektedir. Normal Q-Q grafiği incelendiğinde alt ve üst noktalardaki bazı artıkların normallikten saptığını bunlar dışında diğer artıklar doğru boyunca yer almaktadır. Sapan noktaların normal bir dağılımdan gelmemesi mümkündür, ancak noktalarımızın çoğu normal bir dağılımdan geliyor gibi göründüğünden artıklar normal dağılım göstermektedir denilebilir. Üçüncü grafik, Y gözlem değerlerine karşın verilen artık grafiklerine benzer, ancak artıkların yerine standartlaştırılmış artıkların mutlak değerinin kareköklerini kullanır. İlk grafikte olduğu gibi artıkların sıfır etrafında

toplandığı görülmektedir. Son grafikte ise standartlaştırılmış hatalara ait etkili (leverage) noktalar verilmiştir. Burada verilen Cook's distance, sınır dışında yer alan bir noktanın aykırı değer kabul edileceği anlamına gelmektedir. Grafiğimizde bazı aykırı değerlerin olduğu görülmektedir. Şekil 4.2'deki artıklar grafikleri için R-kodu; par(mfrow=c(2,2)), plot(model) şeklindedir.



Şekil 4.2. Artıkların grafikleri

Çizege 4.5. Eğitim (train) veri kümesi için EKK sonuçları

RSS_{train}	MSE_{train}	R^2	R_d^2
409.774	1.170827	0.8193	0.8167

Çizege 4.6. Deneme (test) veri kümesi için EKK sonuçları

RSS_{test}	MSE_{test}	R^2	R_d^2
195.8573	1.271801	0.8359	0.8347

4.3. Ridge Regresyon Yöntemi İle Modelin Çözümü

Ridge regresyonda $\hat{\beta}_R = \left(X^T X + KI_k\right)^{-1} X^T Y$ değerini hesaplamak için en uygun K değerini seçmemiz gerekmektedir. RR, LASSO ve EN yöntemlerinde hesaplamalar için RStudio programında glmnet paketi kullanılmış, Bunun için öncelikle glmnet paketini yüklenmiştir.

installpackages("glmnet") #glmnet paketini yükler

library("glmnet") #glmnet paketi için kütüphaneye bağlanır.

Tüm veri %70 train (eğitim) ve %30 test (deneme) kümesi olmak üzere iki kümeye parçalanmıştır. Tüm hesaplamalar eğitim kümesinde yapıldıktan sonra deneme kümesinde modelin performansı değerlendirilmiştir. Train ve test kümesi için R-kodları aşağıda verilmiştir:

```
CLI.pred.names=names(CLI)[2:length(names(CLI))]
CLI.pred=CLI[,CLI.pred.names]
set.seed(1)
train=sample(1:nrow(CLI.pred), .70*nrow(CLI.pred))
test=(-train)
x.train=model.matrix(~.,CLI.pred[train,])[,-1]
y.train=CLI$`Yaşam Maliyeti Endeksi`[train]
x.test=model.matrix(~.,CLI.pred[test,])[,-1]
y.test=CLI$`Yaşam Maliyeti Endeksi`[test]
lambda.grid=10^seq(5,-2,length=100)
ridge.mod=glmnet(x.train,y.train,alpha=0,lambda=lambda.grid)
```

Yukarıda verilen R-kodlarında lambda ceza parametresini ifade etmektedir. Veri adı olarak da "CLI" kullanılmıştır.

Ridge küçültme yönteminde en uygun K değerini seçerken işlem kolaylılığından dolayı standartlaştırılmış veriler kullanılmaktadır. Farklı K değerlerine ait parametre tahminleri Çizelge 4.7'de verilmiştir. K değeri arttıkça parametre tahminlerinin küçüldüğü görülmektedir.

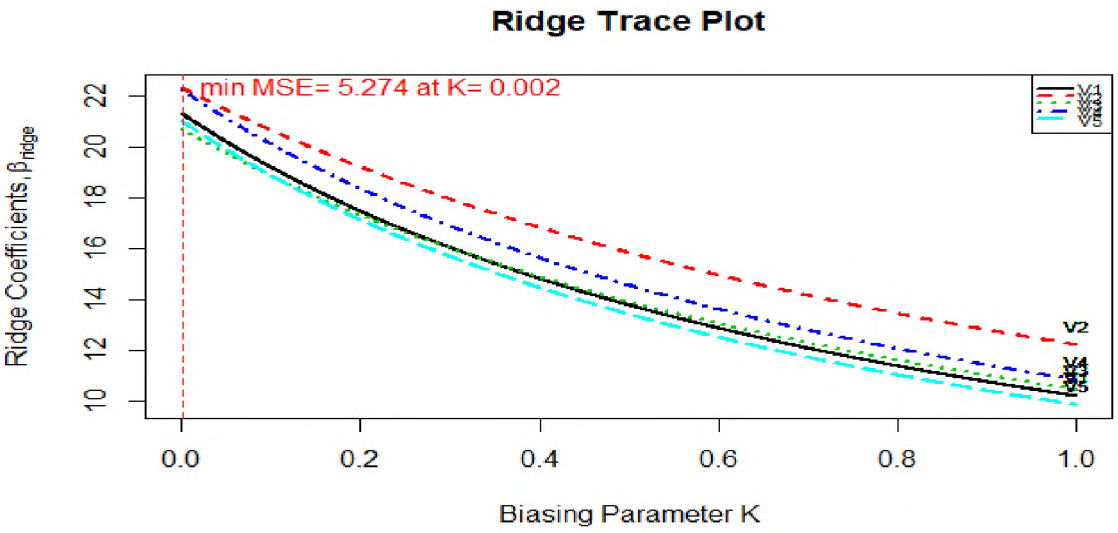
coef(ridge.mod) #RR yöntemine ait parametre tahminlerini verir.

Çizelge 4.7. Bazı K değerlerine ait Ridge regresyon katsayıları

K	$\hat{\beta}_{R_1}$	$\hat{\beta}_{R_2}$	$\hat{\beta}_{R_3}$	$\hat{\beta}_{R_4}$	$\hat{\beta}_{R_5}$
K=0	0.96065199	0.96913175	1.02907282	1.01825659	0.93277039
K =0.001	0.96022162	0.96880571	1.02871415	1.01786224	0.93240373
K =0.002	0.95979163	0.96847988	1.02835574	1.01746820	0.93203735
K =0.003	0.95936202	0.96815428	1.02799757	1.01707446	0.93167127
K =0.004	0.95893279	0.96782889	1.02763965	1.01668104	0.93130548
K =0.005	0.95850393	0.96750372	1.02728198	1.01628792	0.93093997
K =0.006	0.95807544	0.96717877	1.02692456	1.01589511	0.93057476
K =0.007	0.95764734	0.96685403	1.02656739	1.01550261	0.93020984
…	…	…	…	…	…
K =0.1	0.91940868	0.93757515	0.99440257	0.98029711	0.89749237
K =0.2	0.88150083	0.90800327	0.96199267	0.94510035	0.86481196
K =0.3	0.8465443	0.8802353	0.9316292	0.9123721	0.8344500
K =0.4	0.81421119	0.85411164	0.90312423	0.88185878	0.80616620
K =0.5	0.78422011	0.82949087	0.87631234	0.85334085	0.77975265
K =0.6	0.75632819	0.80624728	0.85104703	0.82662699	0.75502841
…	…	…	…	…	…
K =1	0.6619201	0.7249705	0.7630522	0.7347506	0.6701339

4.3.1. Ridge İzi

Ridge ceza parametresi bulunurken çeşitli yöntemlerden yararlandığımızı önceki bölümlerde söylemiştik. Kullanılan yöntemlerden biri de Şekil.4.3 de verilen ridge izi yöntemidir. Amacımız en uygun K değerine karşılık gelen parametreleri tahmin etmektir. K değeri 0'dan başlatılarak belli aralıklarla artırılarak ilerlemiştir. Ridge izi grafiğine göre minimum MSE'yi veren nokta $K = 0.002$ olarak verilmiştir.

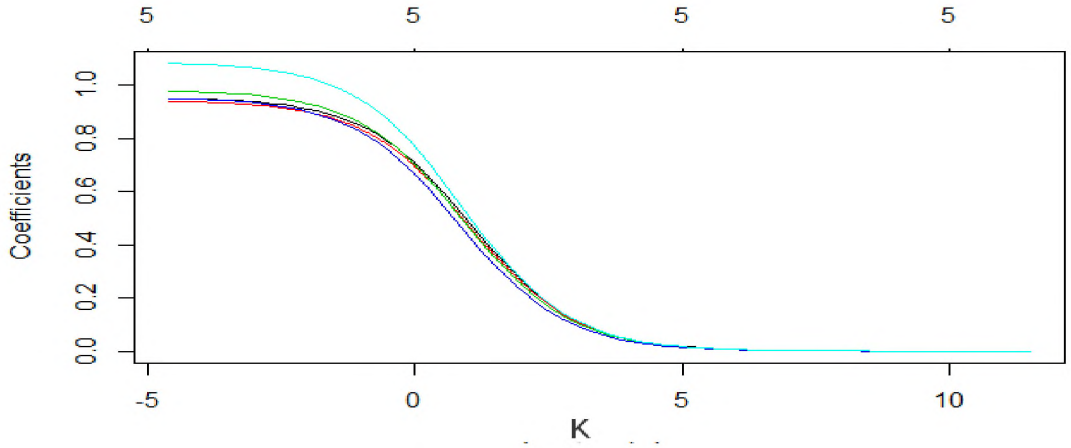


Şekil 4.3. Ridge izi

4.3.2. Çapraz Değerlendirme Yöntemi

Çapraz değerlendirme yöntemi yardımıyla yanıt değerlerini en yüksek doğrulukla tahmin edebilmek için belli bir aralıkda seçilen her *K* değeri için regresyon modeli oluşturulup en iyi *K* ceza parametresi bulunmuştur. Bunun için Rstudio programında "glmnet" paketi kullanılmıştır. Hesaplamalar sonucunda *K* değerinin artmasıyla Ridge parametre tahminlerinde meydana gelen değişiklik Şekil 4.4.'de gösterilmiştir. *K* değerinin artmasıyla Ridge parametre tahminlerinin de sıfıra yaklaştığı görülmektedir.

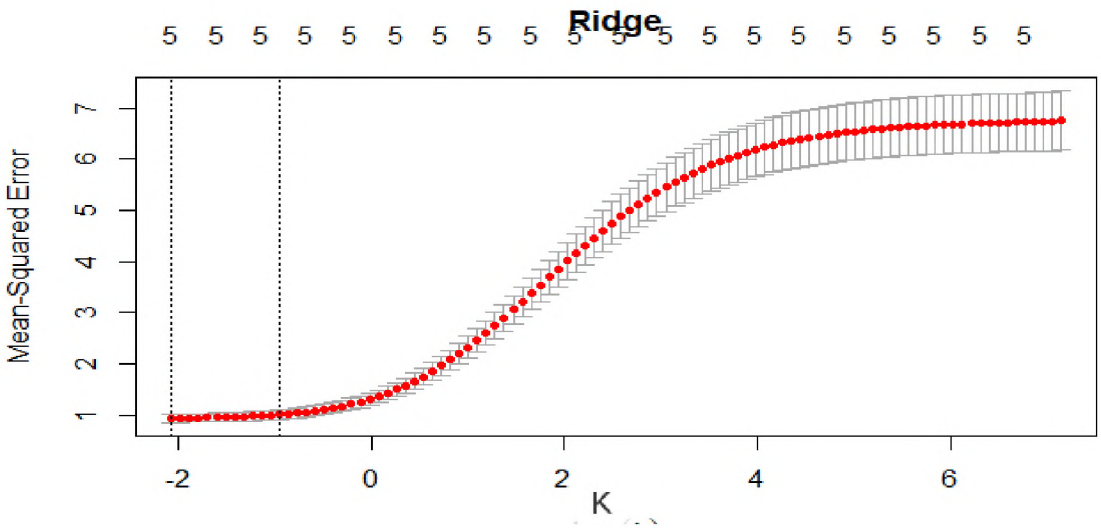
```
plot(ridge.mod,xvar='lambda') # şekil 4.4 için R-kodu
```



Şekil 4.4. K ceza parametresine göre RR parametre tahminlerinin değişimi

Şekil 4.5.'de artan K değerlerine karşılık gelen MSE değerleri gösterilmiştir. Dikey çizgilerden ilki minimum MSE'nin yerini gösterirken ikinci kesikli çizgi 1 standart hatalık değişime karşılık gelen minimum hatayı gösterir. Tahminde bulunurken minimum MSE'yi veren K ceza parametresi kullanılmıştır.

```
ridge.cv.out=cv.glmnet(x.train,y.train,alpha=0)
plot(ridge.cv.out)  # şekil 4.5 için R-kodu
```



Şekil 4.5. K ceza parametresine göre MSE değişimi

```
ridge.bestlam=ridge.cv.out$lambda.min #  K_best  için R-kodu
ridge.mod.best=glmnet(x.train,y.train,alpha=0,lambda=ridge.bestlam)
coef(ridge.mod.best) #Çizelge 4.8'de verilen parametre tahminleri için R-kodu
```

Çizelge 4.8.'de çapraz değerlendirme yöntemi ile bulunan en iyi $K_{best} = 0.35$ değerine göre Ridge regresyon katsayıları verilmiştir.

Çizelge 4.8. $K_{best} = 0.35$ değeri için RR parametre tahminleri

Sabit	$\hat{\beta}_{R_1}$	$\hat{\beta}_{R_2}$	$\hat{\beta}_{R_3}$	$\hat{\beta}_{R_4}$	$\hat{\beta}_{R_5}$
-0.8339738	0.83006838	0.86697717	0.91715520	0.89685300	0.82006202

Böylece elde edilen K_{best} değeri için regresyon modeli,

$$Y_{best} = -0.8339738 + 0.8306838X_1 + 0.86697717X_2 + 0.91715520X_3 + 0.89685300X_4 + 0.82006202X_5$$

şeklinde bulunur.

Ridge küçültme yöntemine ait sonuçlar ise çizelge 4.9 ve çizelge 4.10'da verilmiştir.

Çizelge 4.9. Eğitim (train) veri kümesi için RR sonuçları

$RSS_{R_{train}}$	$MSE_{R_{train}}$	$R^2_{R_{train}}$
344.4287	0.9647861	0.8561034

Çizelge 4.10. Deneme (test) veri kümesi için RR sonuçları

$RSS_{R_{test}}$	$MSE_{R_{test}}$	$R^2_{R_{test}}$
124.3933	0.8077485	0.8565431

y.ridge.best.train = predict(ridge.mod.best, newx=x.train)

rssridge=sum((y.ridge.best.train – y.train)^2) # $RSS_{R_{train}}$ için R-kodu

mean((y.ridge.best.train – y.train)^2) # $MSE_{R_{train}}$ için R-kodu

tssridge = sum((y.train - mean(y.train)) ^ 2)

rsq = 1 - rssridge/tssridge# $R^2_{R_{train}}$ için R-kodu

y.ridge.best.test= predict(ridge.mod.best, newx=x.test)

Rssridge=sum((y.ridge.best.test – y.test)^2) # $RSS_{R_{test}}$ için R-kodu

mean((y.ridge.best.test – y.test)^2) # $MSE_{R_{test}}$ için R-kodu

Tssridge= sum((y.test - mean(y.test)) ^ 2)

Rsq = 1 - Rssridge/Tssridge # $R^2_{R_{test}}$ için R-kodu

4.4. LASSO Küçültme Yöntemi İle Modelin Çözümü

Çizelge 4.11.'de bazı ψ ceza parametreleri için hesaplanan regresyon katsayıları verilmiştir. Görüldüğü üzere LASSO küçültme yöntemi tahminlerde bulunurken bağımsız değişkenlerden bazılarını sıfıra eşitlerken bazılarını da sıfıra yakın hesaplar. Ayrıca şekil 4.6. incelendiğinde de ψ değerinin artmasıyla LASSO parametre

tahminlerinin sıfıra yaklaştığı hatta bir noktadan sonra tahminlerin sıfıra eşitlendiği görülmektedir.

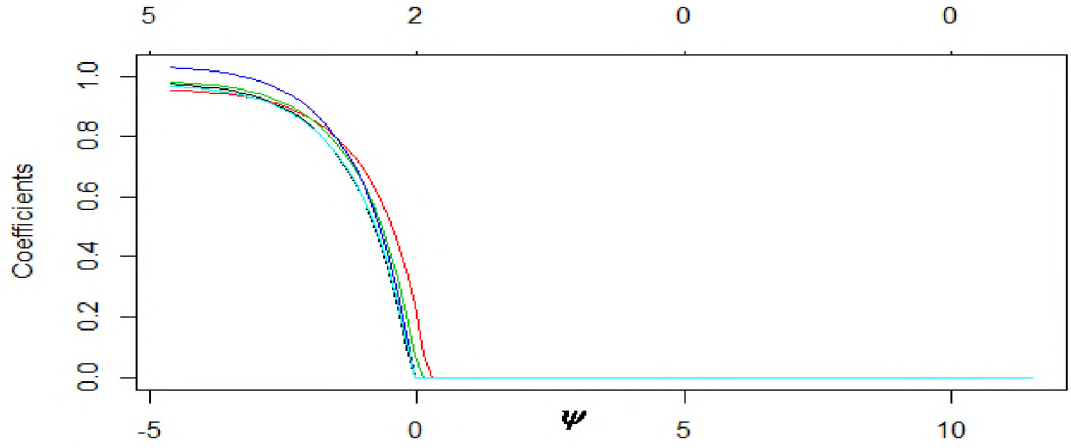
lasso.mod=glmnet(x.train,y.train,alpha=1,lambda=lambda.grid)

Yukarıda verilen R-kodlarında lambda, ceza paramtresini ifade etmektedir. Alpha değeri RR için 0 alınır, LASSO için ise 1 alınır.

coef(lasso.mod) #Çizelge 4.11'de verilen parametre tahminleri için R-kodu

Çizelge 4.11. ψ değerlerine karşılık gelen LASSO parametre tahminleri

ψ	$\hat{\beta}_{LASSO1}$	$\hat{\beta}_{LASSO2}$	$\hat{\beta}_{LASSO3}$	$\hat{\beta}_{LASSO4}$	$\hat{\beta}_{LASSO5}$
0.001	0.9518769	0.9407932	0.9801460	0.9524944	1.0847374
0.002	0.9510668	0.9398795	0.9791382	0.9514454	1.0836878
0.003	0.9502568	0.9389658	0.9781304	0.9503964	1.0826383
0.004	0.9494468	0.9380521	0.9771226	0.9493474	1.0815888
...	....	...	...	...	...
0.9	0.223660547	0.119382020	0.074138647	0.009431589	0.141216483
1	0.141185145	0.023710331	.	.	0.047156420
1.1	0.0515561	.	.	.	.
1.2	.	.	.	.	.

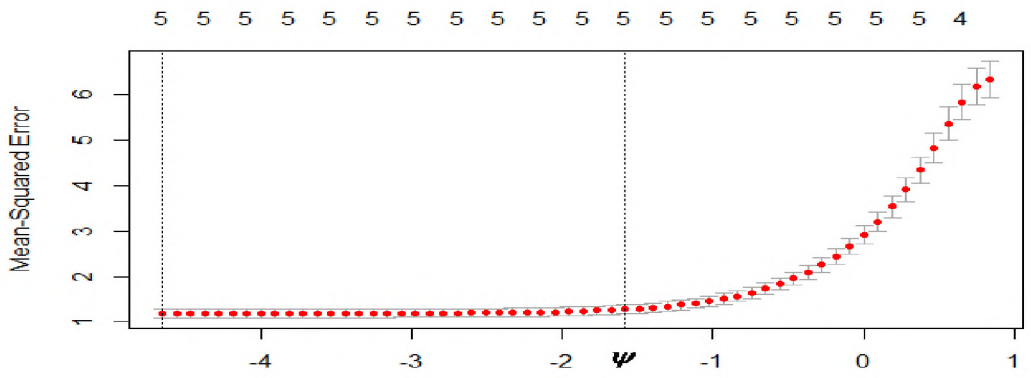


Şekil 4.6. ψ değerine göre LASSO parametre tahminlerinin değişimi

plot(lasso.mod,xvar='lambda') # şekil 4.6 için R-kodu

LASSO ceza parametresini bulmak için Ridge küçültme yönteminde de kullandığımız ceza değerlendirme yöntemini kullanacağız. En iyi ψ ceza parametresi MSE'yi minimum yapan değer olup, Rstudio programında "glmnet" paketi kullanılarak bulunmuştur.

Şekil 4.7. incelendiğinde ψ değerinin artmasıyla MSE'nin de arttığı görülmektedir. Ridge küçültme yönteminde dikey çizgilerden ilkinin minimum MSE'nin yerini gösterirken ikinci kesikli çizginin 1 standart hatalık değişime karşılık gelen noktayı gösterdiğinin belirtmiştik. Tahminde bulunurken minimum MSE'yi veren ψ_{best} ceza parametresi kullanılmıştır



Şekil 4.7. ψ ceza parametresine göre MSE değişimi

lasso.cv.out=cv.glmnet(x.train,y.train,alpha=1)

plot(lasso.cv.out) # şekil 4.7 için R-kodu

Çapraz değerlendirme yöntemi ile minimum MSE'yi veren ceza parametresi $\psi_{\text{best}} = 0.10$ olarak bulunmuştur. Çizelge 4.12.'de LASSO parametre tahminleri verilmiştir.

lasso.bestlam=lasso.cv.out$lambda.min # ψ_{best} için R-kodu

lasso.mod.best=glmnet(x.train,y.train,alpha=1,lambda=lasso.bestlam)

coef(lasso.mod.best) #Çizelge 4.12'de verilen parametre tahminleri için R-kodu

Çizelge 4.12. LASSO parametre tahminleri

ψ	Sabit	$\hat{\beta}_{LASSO1}$	$\hat{\beta}_{LASSO2}$	$\hat{\beta}_{LASSO3}$	$\hat{\beta}_{LASSO4}$	$\hat{\beta}_{LASSO5}$
ψ_{best}	-0.08368559	0.85010571	0.88260014	0.93818985	0.92468375	0.84026801

Böylece elde edilen ψ_{best} için regresyon denklemi,

$\hat{Y} = -0.08368559 + 0.85010571X_1 + 0.88260014X_2 + 0.93818985X_3 + 0.92468375X_4 + 0.84026801X_5$

olarak bulunur.

Çizelge 4.13. Eğitim veri kümesi için LASSO sonuçları

$RSS_{LASSO_{train}}$	$MSE_{LASSO_{train}}$	$R^2_{LASSO_{train}}$
334.4765	0.9369091	0.8602612

Çizelge 4.14. Deneme veri kümesi için LASSO sonuçları

$RSS_{LASSO_{test}}$	$MSE_{LASSO_{test}}$	$R^2_{LASSO_{test}}$
121.6796	0.7901271	0.8596727

y.lasso.best.train = predict(lasso.mod.best, newx=x.train)

rsslasso=sum((y.lasso.best.train – y.train)^2) # $RSS_{LASSO_{train}}$ için R-kodu

mean((y.lasso.best.train – y.train)^2) # $MSE_{LASSO_{train}}$ için R-kodu

tsslasso= sum((y.train - mean(y.train)) ^ 2)

rsq = 1 - rsslasso/tsslasso# $R^2_{LASSO_{train}}$ için R-kodu

y.lasso.best.test= predict(lasso.mod.best, newx=x.test)

Rsslasso=sum((y.lasso.best.test – y.test)^2) # $RSS_{LASSO_{test}}$ için R-kodu

mean((y.lasso.best.test – y.test)^2) # $MSE_{LASSO_{test}}$ için R-kodu

Tsslasso = sum((y.test - mean(y.test)) ^ 2)

Rsq = 1 - Rsslasso/Tsslasso # $R^2_{LASSO_{test}}$ için R-kodu

4.5. Elastik Net İle Modelin Çözümü

Bir diğer küçültme yöntemi Elastik net'in LASSO ve Ridge küçültme yöntemlerinin birleşimi olduğunu söylemiştik. LASSO ve Ridge hesaplamalarında olduğu gibi bu yöntemde de RStudio glmnet paketi kullanılacaktır. Minimum hatayı elde etmemizi sağlayan en uygun lemma2 değeri çapraz değerlendirme yöntemiyle bulunmuştur.

bf.en=glmnet(x.train,y.train,alpha=0.5,lambda=lambda.grid)
Alpha değeri EN için 0 ile 1 arasında bir değer alınır.

coef(bf.en) #Çizelge 4.15'de verilen parametre tahminleri için R-kodu.

Çizelge 4.15'de farklı lemma2 değerlerine ait parametre tahminleri verilmiştir. δ_2 değeri arttıkça parametre tahminlerinin sıfıra yaklaştığı, bir değerden sonrada sıfıra eşitlendiği görülmektedir.

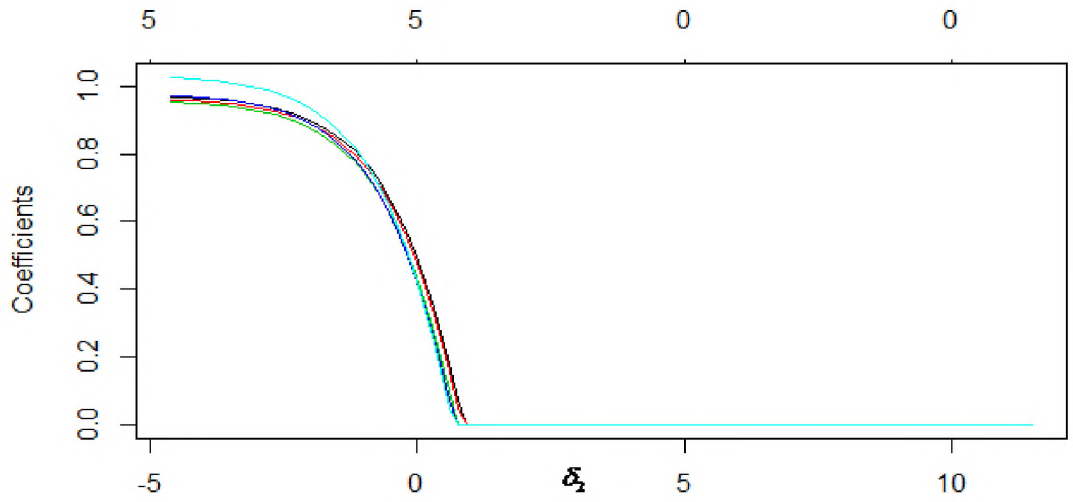
Çizelge 4.15. Farklı δ_2 değerleri için EN parametre tahminleri

δ_2	$\hat{\boldsymbol{\beta}}_{EN1}$	$\hat{\boldsymbol{\beta}}_{EN2}$	$\hat{\boldsymbol{\beta}}_{EN3}$	$\hat{\boldsymbol{\beta}}_{EN4}$	$\hat{\boldsymbol{\beta}}_{EN5}$
0.001	0.9521160	0.9410815	0.9804511	0.9528099	1.0850351
0.002	0.9515452	0.9404564	0.9797487	0.9520766	1.0842837
0.003	0.9509747	0.9398314	0.9790466	0.9513437	1.0835326
…	…	…	…	…	…
0.01	0.9469863	0.9354631	0.9741398	0.9462220	1.0782838
0.02	0.94130498	0.92924127	0.96715445	0.93893232	1.07081263
0.03	0.93564280	0.92304127	0.96019770	0.93167419	1.06337324
…	…	…	…	…	…
1.5	0.27087169	0.20083158	0.17460666	0.12394480	0.23091312

Çizelge 4.15. Farklı δ_2 değerleri için EN parametre tahminleri(devamı)

1.6	0.235239430	0.162422790	0.134059767	0.082926133	0.188347647
...	...	...	...	...	...
1.9	0.134098201	0.052430613	0.020069169	.	0.070547148
2	0.101410544	0.016859062	.	.	0.036101965

Şekil 4.8. incelendiğinde δ_2 değerinin artmasıyla parametre tahminlerinin sıfıra yaklaştığı hatta bazı tahminlerin sıfıra eşit bulunduğu görülmektedir.

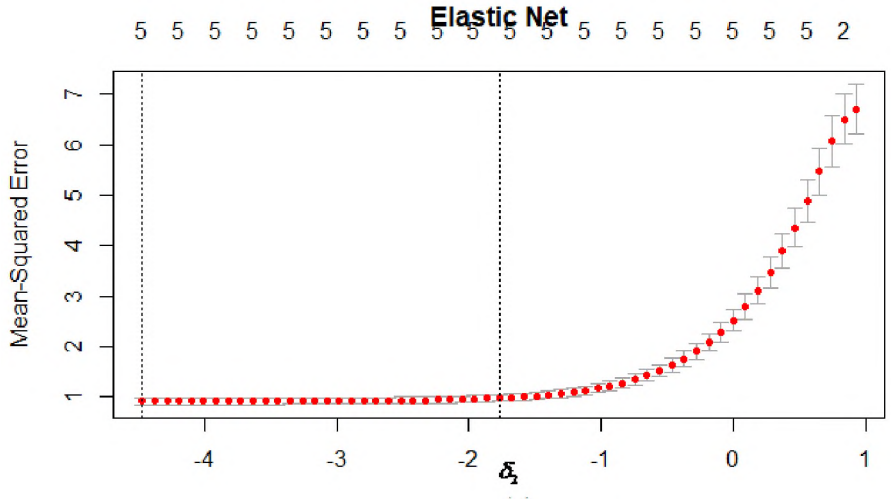


Şekil 4.8. δ_2 değerine göre EN parametre tahminlerinin değişimi

```
plot(bf.en,xvar='lambda') # şekil 4.8 için R-kodu
```

Şekil 4.9.'da δ_2 değerlerinin artmasıyla MSE değerlerinin de arttığı görülmektedir.Ridge ve LASSO'da olduğu gibi burada da sadece $\delta_{2.best}$ değeri için parametre tahminleri yapılmıştır.

```
cv.en=cv.glmnet(x.train,y.train,alpha=0.5)
plot(cv.en)  # şekil 4.9 için R-kodu
```



Şekil 4.9. δ_2 ceza parametresine göre MSE'nin değişimi

Çapraz değerlendirme yöntemi ile minimum MSE'yi veren ceza parametresi $\delta_{2.\text{best}} = 0.15$ olarak bulunmuştur. Çizelge 4.16'da Elastic Net için parametre tahminleri verilmiştir.

en.bestlam = cv.en$lambda.min # $\delta_{2.\text{best}}$ için R-kodu

bf.en.best = glmnet(x.train, y.train, alpha=0.5, lambda=en.bestlam)

coef(bf.en.best #Çizelge 4.16'da verilen parametre tahminleri için R-kodu

Çizelge 4.16. EN parametre tahminleri

Sabit	$\hat{\beta}_{\text{EN1}}$	$\hat{\beta}_{\text{EN2}}$	$\hat{\beta}_{\text{EN3}}$	$\hat{\beta}_{\text{EN4}}$	$\hat{\beta}_{\text{EN5}}$
-0.08365498	0.84912708	0.88199916	0.93641306	0.92133111	0.83868378

Böylece elde edilen regresyon denklemi,

$$\hat{Y} = -0.08365498 + 0.84912708X_1 + 0.88199916X_2 + 0.93641306X_3 + 0.92133111X_4 + 0.83868378X_5$$

şeklindedir.

Çizelge 4.17. Eğitim veri kümesi için EN sonuçları

$RSS_{en_{train}}$	$MSE_{en_{train}}$	$R^2_{en_{train}}$
336.3344	0.9393122	0.8599028

Çizelge 4.18. Deneme veri kümesi için EN sonuçları

$RSS_{en_{test}}$	$MSE_{en_{test}}$	$R^2_{en_{test}}$
121.9818	0.7920895	0.8593242

y.en.best.train = predict(bf.en.best, newx=x.train)

rssen=sum((y.en.best.train – y.train)^2) # $RSS_{en_{train}}$ için R-kodu

mean((y.en.best.train – y.train)^2) # $MSE_{en_{train}}$ için R-kodu

tssen= sum((y.train - mean(y.train)) ^ 2)

rsq = 1 - rssen/tssen # $R^2_{en_{train}}$ için R-kodu

y.en.best.test = predict(bf.en.best, newx=x.test)

Rssen=sum((y.en.best.test – y.test)^2) # $RSS_{en_{test}}$ için R-kodu

mean((y.en.best.test – y.test)^2) # MSE_{entest} için R-kodu

Tssen = sum((y.test - mean(y.test)) ^ 2)

Rsq = 1 - Rssen/Tssen # $R^2_{en_{test}}$ için R-kodu

4.6. Liu Küçültme Yöntemi İle Modelin Çözümü

Liu tahmin edicisi $\hat{\beta}_d$, d ceza parametresinin doğrusal fonksiyonudur. Bu nedenle d ceza parametresinin seçimi, ridge küçültme yöntemindeki K'nın seçilmesine göre daha kolaydır. Bu yöntemde d ceza parametresi belirlendikten sonra Liu tahmin edicisi $\hat{\beta}_d = (X^T X + I)^{-1}(X^T Y + d\hat{\beta})$ değeri hesaplanacaktır. Farklı d değerleri için Liu parametre tahminleri Çizelge 4.19'da verilmiştir. Hesaplamalar için Rstudio programı liureg paketi (Imdadullah ve ark. , 2017) kullanılmış ve *d* ceza parametresi

-0.22 olarak bulunmuştur. Öncelikle liureg paketini kullanmak için aşağıdaki R-kodu kullanılmıştır.

installpackages("liureg") # liureg paketini yükler.

library("liureg") # kütüphaneyi getirir.

model=liu(y.train ~ ., data = as.data.frame(x.tarin), scaling = "centered",

+d = seq(0, 1, 0.01)) # Regresyon modeli, Liu için tanımlanır.

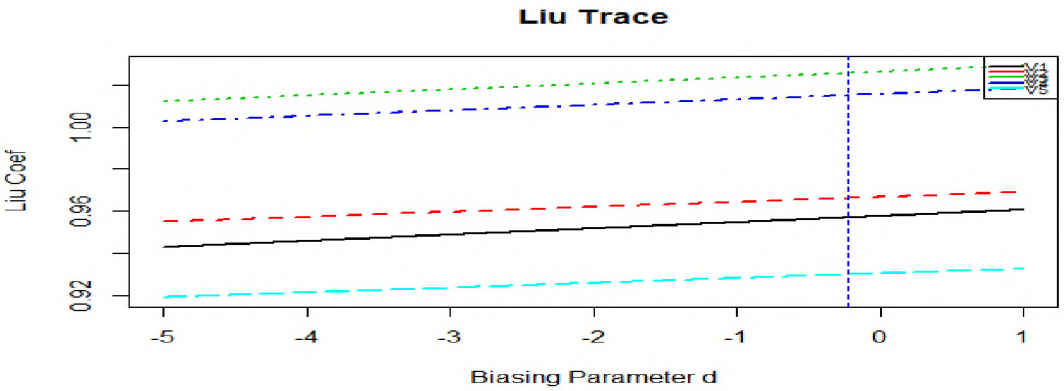
model$coef # çizelge 4.19 için R-kodu.

Çizelge 4.19. Bazı d değerleri için parametre tahminleri

d	$\hat{\beta}_{L1}$	$\hat{\beta}_{L2}$	$\hat{\beta}_{L3}$	$\hat{\beta}_{L4}$	$\hat{\beta}_{L5}$
d =-0.22	0.95704366	0.96625321	1.02563501	1.01507724	0.92999789
d =-0.2	0.95710281	0.96630040	1.02569136	1.01512936	0.93004334
d =-0.1	0.95739858	0.96653634	1.02597315	1.01538997	0.93027059
d =0	0.95769435	0.96677229	1.02625494	1.01565057	0.9304978
d =0.01	0.95772393	0.96679588	1.02628312	1.01567663	0.93052057
d =0.02	0.95775350	0.96681948	1.02631130	1.01570269	0.93054330
...	...	...	...	...	...
d =0.1	0.95799012	0.96700823	1.02653673	1.01591117	0.93072510
d =0.2	0.9582859	0.9672442	1.0268185	1.0161718	0.9309524
d =0.3	0.95858166	0.96748013	1.02710031	1.01643237	0.93117961
d =0.4	0.95887743	0.96771607	1.02738209	1.01669298	0.93140687
d =0.5	0.95917320	0.96795202	1.02766388	1.01695358	0.93163412
...	...	...	...	...	...
d =1	0.96065204	0.96913175	1.02907282	1.01825659	0.93277039

Şekil 4.10'da en uygun d ceza parametresinin yeri verilmiştir.

plot(model) # şekil 4.10 için R-kodu



Şekil 4.10. d ceza parametresine göre minimum MSE'nin yeri

d =-0.22 ceza parametresine göre elde edilen Liu parametre tahminleri Çizelge 4.20'de verilmiştir.

dest(model) # Farklı yöntemlere ait en iyi Liu ceza parametresi değerlerini verir.

coef(model) # Çizelge 4.20 için R-kodu

Çizelge 4.20. Liu parametre tahminleri

Sabit	$\hat{\beta}_{L1}$	$\hat{\beta}_{L2}$	$\hat{\beta}_{L3}$	$\hat{\beta}_{L4}$	$\hat{\beta}_{L5}$
-0.8476624	0.95704366	0.96625321	1.02563501	1.01507724	0.92999789

Son olarak Çizelge 4.21 incelendiğinde R^2 değerinin yüksek çıkması, bağımsız değişkenlerin Y bağımlı değişkenini büyük oranda açıkladığını, RSS ve MSE değerlerinin de küçük bulunması modelin EKK yöntemine göre daha az bir hata miktarı ile tahmin edildiğini göstermektedir.

summary(model) #Liu için MSE_L ve R_L^2 değerlerini hesaplar.

press(model) # Liu için hata kareler toplamını hesaplar.

mod=liu(y.test ~ ., data = as.data.frame(x.test), scaling = "centered",

+d = seq(0, 1, 0.01)) # Test kümesi için Liu regresyon modeli

Çizelge 4.22.'deki hesaplamalar için Liu modeli test kümesi için yeniden tanımlanmış, hesaplamalar ona göre elde edilmiştir.

Çizelge 4.21. Eğitim veri kümesi için Liu sonuçları

$RSS_{L_{train}}$	$MSE_{L_{train}}$	$R^2_{L_{train}}$
361.8329	1.013537	0.8349767

Çizelge 4.22. Deneme veri kümesi için Liu sonuçları

$RSS_{L_{test}}$	$MSE_{L_{test}}$	$R^2_{L_{test}}$
125.7756	0.8167246	0.853784

5. TARTIŞMA VE SONUÇ

Regresyon modelimizde iç ilişki olduğundan EKK yöntemi yerine kullanılan küçültme yöntemlerinden elde edilen sonuçlar Çizelge 5.1 ve Çizelge 5.2'de verilmiştir. Verimiz %70 eğitim (train) veri kümesi ve %30 deneme (test) veri kümesi olarak parçalanmış, hesaplamalar eğitim kümesinde yapıldıktan sonra deneme kümesinde modelin performansına bakılmıştır. EKK yöntemi ve küçültme yöntemlerinin karşılaştırılmasında RSS, MSE ve R^2 belirleme katsayıları kullanılmıştır.

Çizelge 5.1. Train veri kümesi için genel sonuçlar

Yöntem	Hata Kareler Top.(RSS)	Hata Kareler Ort.(MSE)	R^2_{train}
EKK (train)	409.774	1.170827	0.8192801
Ridge(train)	344.4287	0.9647861	0.8561034
Lasso(train)	334.4775	0.9379091	0.8602612
Elastik Net(train)	336.3344	0.9393122	0.8599028
Liu(train)	361.8329	1.013537	0.8349767

Çizelge 5.2. Test veri kümesi için genel sonuçlar

Yöntem	Hata Kareler Top.(RSS)	Hata Kareler Ort.(MSE)	R^2_{test}
EKK(test)	195.8573	1.271801	0.8358995
Ridge(test)	124.3933	0.8077485	0.8575431
Lasso(test)	121.7797	0.7901271	0.8597727
Elastik Net(test)	121.9818	0.7920895	0.8593242
Liu(test)	125.7756	0.8167246	0.853784

Eğitim ve deneme verilerindeki sonuçlar incelendiğinde R^2 belirleme katsayısına göre en iyi sonucu LASSO vermektedir. Diğer kriterler RSS ve MSE incelendiğinde de yine en iyi sonucun LASSO yönteminden elde edildiği görülmektedir. Karşılaştırma kriterlerine göre en etkili yöntemler sırasıyla LASSO, Elastik net, Ridge regresyon, Liu

ve EKK yöntemidir. Sonuç olarak değişkenler arasında iç ilişki olması durumunda küçültme yöntemlerinin EKK yöntemine alternatif olarak kullanılması önerilir.

6. KAYNAKLAR

Akdeniz, F. and Kaçıranlar, S. 1995. On the almost unbiased generalized Liu estimator and Unbiased estimation of the Bias and MSE, Communications in Statistics -Theory and Methods. 24(7) : 1789-1797.

Albayrak, S. A. 2005. Çoklu bağlantı halinde en küçük kareler teknikleri ve bir uygulama. Zonguldak Kara Elmas Üniversitesi, ***Sosyal Bilgiler Dergisi***, 1(1): 105-126.

Anderson, B. 2001. Scandinavian evidence on growth and age structure. ***Regional Studies***, 35(5): 377-390.

Bai, J. and S. Ng (2007), 'Determining the number of primitive shocks'. ***Journal of Business and Economic Statistics***, 25(1) : 52–60.

Belsley, D. A., Kuh, E. and Welsch, R. E. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons., New York.

Dettling, M. and Bühlmann, P. 2004. Finding predictive gene groups from microarray data. ***J. Multiv. Anal***., 90 : 106–131.

Diaz, U. R. 2003. A simple method for finding molecular signatures from gene expression data. Technical Report. Spanish National Cancer Center. Erisim: [http://www.arxiv.org/abs/q-bio.QM/0401043]. Erisim Tarihi :30.01.2004.

Draper, N.R. and Smith, H. 1998. Applied Regression Analysis. 3th Edition, Wiley, New York.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. 1995. Wavelet shrinkage: asymptopia (with discussion)? ***J. R. Statist. Soc. B,*** 57 : 301–369.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. 2004. Least angle regression. ***The Annals of Statistics***, 32(2) : 407-499.

Efron, B., Hastie, T. 2016. Computer Age Statistical Inference: Algorithms, Evidence and Data Science, Institute of Mathematical Statistics Monographs, 1st Ed. Cambridge University Press.

Farrar, D. E. and Glauber, R. R. 1967. Multicollinearity in regression analysis: the problem revisited. Rev. Econ. Statist. 49(1):92–107.

Golub, G. H., Heath, M. and Wahba, G.1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter.***Technometrics***, 21(2) : 215-223.

Gujarati, D. N. 1995. Basic econometrics, 3rd Ed. McGraw-Hill, 1002pp, New York.

Hastie, T., Tibshirani, R., Eisen, M., Brown, P., Ross, D., Scherf, U., Weinstein, J., Alizadeh, A., Staudt, L. and Botstein, D. 2000. 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. ***Genome Biol.***, 1 : 1–21.

Hastie, T., Tibshirani, R. and Friedman, J. 2001. The Elements of Statistical Learning; Data Mining, Inference and Prediction. Springer, New York.

Hastie, T., Tibshirani, R., Botstein, D. and Brown, P. 2003. Supervised harvesting of expression trees. Genome Biol., 2, 0003.1–0003.12.

Hastie, T., Tibshirani, R. and Wainwright, M. 2015. Statistical Learning with Sparsity. TheLasso and Generalizations. Chapman & Hall.

Hoerl, A. E. and Kennard, R.W. 1970a. Ridge regression: biased estimation for non-orthogonal problems. ***Technometrrics***, 12(1):55-67.

Hoerl, A. E. and Kennard, R. W. 1970b. Ridge regression: biased estimation for non-orthogonal problems. ***Technometrrics***, 12(1):69-82.

James, G., Witten, D., Hastie, T. and Tibshirani, R. 2013. An Introduction to Statistical Learning with Applications in R. Springer.

Johnston, J. 1972. Econometric Methods. McGraw-Hill Companies, Inc., New York.

Kunter, M. N., Nachtsheim, C. J., Neter, J. and William, L. 2005. Applied Linear Statistical Models, 5th Ed. McGraw-Hill Companies, Inc., New York.

Liu, K. 1993. A new class of biased estimate in linear regression', Communications in Statistics-Theory and Methods , 22(2) : 393–402.Liu, X. Q. 2011. Improved Liu Estimator in a Linear Regression Model. ***Journal of Statistical Planning and Inference***, 141(1):189–196.

Mallows, C.L. 1973. Some comments on Cp. ***Technometrics***, 15(4): 661-675.

Marquart, D.W. 1970. Generalized invers, ridge regression, biased linear estimation and nonlinear estimation. ***Techonometrics***, 12: 591-612.

Marquardt, D. W. and Snee, R. D. 1975. Ridge Regression in Pratice. ***The American Statistician***, 29: 3-20.

McDonald, G.C. and Galarneau, D. I. 1975. A Monte Carlo evaluation of some ridge-type estimators. ***J. Amer. Statist. Assoc***, 70: 407–416.

Montgomery, D. and Peck, E. 1982. Introductıon To Linear Regression Analysis. NewYork.

Neter, J., Wasserman, W., and Kunter, M. 1990. Applied Linear Statistical Models, 3rd Ed. 842pp, New Jersey.

Orhunbilge, N. 2002. Uygulamalı Regresyon ve Korelasyon Analizi . İ.Ü Basım ve Yayınevi, İstanbul.

RStudio, Erisim: [https://cran.r-project.org/web/packages/glmnet/glmnet.pdf].

Erisim Tarihi :26.06.2014

RStudio , Erisim: [https://cran.r-project.org/web/packages/liureg/liureg.pdf].

Erisim Tarihi :02.12.2017

Segal, M., Dahlquist, K. and Conklin, B. 2003. Regression approach for microarray data analysis. ***J. Computnl Biol.***, 10 : 961–980.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. ***Journal of the Royal Statistical Society: Series B (Statistical Methodology)***, 58(1):267–288.

Usai, M.G., Carta,A. and Casu, S. 2012. Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure, ***BMC Proceedings***, 6 (2):9

Webster, A. 1995. Applied Statist for Business and Economics. 683-684.

Zou, H. and Hastie, T. 2005. Regularization and variable selection via the elastic net. ***Statistical Society: Series B***, 67: 301–320.

ÖZGEÇMİŞ

1989 yılında Diyarbakır'da doğdum. İlkokul, ortaokul ve lise öğrenimimi Diyarbakır'da tamamladım. 2007 yılında Diyarbakır Anadolu Lisesi'nden, 2012 yılında da Dicle Üniversitesi İlköğretim Matematik Öğretmenliği Bölümünden mezun oldum. 2012 yılından bu yana Milli Eğitime bağlı farklı okullarda görev yapmaktayım. Çınar İmam Hatip Ortaokulunda görevime devam etmekteyim.

Erdem KALKAN

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
TEZ İNTİHAL FORMU

ÖĞRENCİ BİLGİLERİ	
ADI VE SOYADI	Erdem KALKAN
ÖĞRENCİ NO	13804007
EĞİTİM – ÖĞRETİM YILI	2019-2020
YARIYIL	☒ Güz ☐ Bahar
ANABİLİM DALI	MATEMATİK
PROGRAM	☒ Yüksek Lisans ☐ Doktora
TEZ KONUSU	Lineer Regresyonda Küçültme Yöntemlerinin Karşılaştırılması
İNTİHAL RAPORU BİLGİLERİ	
RAPOR TÜRÜ	Tez Savunma Sınavı Sonrası
SAYFA SAYISI	79
BENZERLİK ORANI	%17
RAPORLAMA TARİHİ	22/01/ 2020

Yukarıda başlığı/konusu gösterilen tez çalışmamın kapak sayfası, giriş, ana bölümler, sonuç ve tartışma kısımlarından oluşan toplam 79 sayfalık kısmına ilişkin, 22/01/2020 tarihinde şahsım/tez danışmanım tarafından TURNITIN adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan intihal raporuna göre, tezimin benzerlik oranı % 17'dir.

Uygulanan filtrelemeler:

☒Kabul/Onay sayfaları hariç,
☐Kaynakça hariç
☒Alıntılar dâhil
☐Diğer

Dicle Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Programlarda Tez Çalışması İntihal Raporu Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edilmesi durumunda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Erdem KALKAN
22.01.2020

Doç. Dr. Pakize TAYLAN **Tez Danışmanı** 22.01.2020	Prof. Dr. H. Özlem GÜNEY **Anabilim Dalı Başkanı** 22.01.2020

Formdaki bilgiler bilgisayar ortamında doldurulmalıdır. El yazısı ile doldurulan formlar geçersiz sayılmaktadır.

KGK-FRM-340/00