

UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

İşıl FİDANOĞLU

İSTATİSTİKSEL DARALTICI (SHRINKAGE) MODEL VE
UYGULAMALARI

İSTATİSTİK ANABİLİM DALI

ADANA, 2009

ÇUKUROVA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİKSEL DARALTICI (SHRINKAGE) MODEL VE UYGULAMALARI

Işıl FİDANOĞLU

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

Bu tez/..... Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/ Oyçokluğu İle Kabul Edilmiştir.

İmza.....
Prof. Dr.Fikri AKDENİZ
DANIŞMAN

İmza.....
Prof. Dr.Olcay ARSLAN
ÜYE

İmza.....
Prof. Dr. Altan ÇABUK
ÜYE

Bu tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.

Kod No:

Prof. Dr. Aziz ERTUNÇ

Enstitü Müdürü

İmza ve Mühür

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS

İSTATİSTİKSEL DARALTICI (SHRINKAGE) MODEL VE UYGULAMALARI

Işıl FİDANOĞLU

ÇUKUROVA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

Danışman: Prof. Dr. Fikri AKDENİZ

Yıl: 2009, **Sayfa:** 83

Jüri: Prof. Dr. Fikri AKDENİZ

: Prof. Dr. Olcay ARSLAN

: Prof. Dr. Altan ÇABUK

Bridge Regresyon, $\gamma \geq 1$ koşulu ile $\sum |\beta_j|^\gamma$ penalty (ceza) fonksiyonu, cezalı regresyonun özel türüdür. Bridge tahmin edicisi; $\gamma = 1$ için atış metodu ya da $\gamma > 1$ için düzeltilmiş Newton-Raphson metodu ile cezalı skor denklemlerinin çözülmesi sonucu elde edilir. Bridge tahmin edici, yanlılığı biraz gözden çıkarma ile küçük varyanslar verir ve böylece doğrusal regresyon modelinde mevcut açıklayıcı değişkenler arasında ilişki olduğunda, küçük hata kareler ortalaması ve küçük hata tahmini elde edilir.

Cezalandırma kavramı; ortak likelihood fonksiyonlarının oluşuna rağmen, cezalandırmanın uygulanmasını sağlarken, cezalı skor denklemleri ile genelleştirilir. Cezalandırma, genelleştirilmiş lineer modeller (GLM) ve genelleştirilmiş tahmin denklemlerinden (GEE) sonra uygulanır.

Ceza parametresi γ ve düzen (ayar) parametresi λ ; genelleştirilmiş çapraz geçerlilik testi (GCV) ile seçilir. Yarı-GCV; cezalı genelleştirilmiş tahmin denklemleri için parametre seçmeye geliştirilir.

Anahtar Kelimeler: Bridge Regresyon, Cezalı Genelleştirilmiş Tahmin Denklemleri, LASSO, Ridge regresyon

ABSTRACT

MSc THESIS

A STATISTICAL SHRINKAGE MODEL AND ITS APPLICATIONS

Işıl FİDANOĞLU

DEPARTMENT OF STATISTICS

INSTITUTE OF NATURAL AND APPLIED SCIENCES

UNIVERSITY OF ÇUKUROVA

Supervisor: Prof. Dr. Fikri AKDENİZ

Year: 2009 **Pages:** 83

Jury : Prof. Dr. Fikri AKDENİZ

: Prof. Dr. Olcay ARSLAN

: Prof. Dr. Altan ÇABUK

Bridge regression, a special type of penalized regression of a penalty function $\sum |\beta_j|^\gamma$ with $\gamma \geq 1$ is considered. The Bridge estimator is obtained by solving the penalized score equations via the modified Newton-Raphson method for $\gamma > 1$ or the Shooting method for $\gamma = 1$. The Bridge estimator yields small variance with a little sacrifice of bias. And thus achieves small mean squared error and small prediction error when collinearity is present among regressors in a linear regression model.

The concept of penalization is generalized via the penalized score equations, which allow the implementation of penalization regardless of the existence of joint likelihood functions. Penalization is then applied to generalized linear models and generalized estimating equations (GEE).

The penalty parameter γ and the tuning parameter λ are selected via the generalized cross-validation (GCV). A quasi-GCV is developed to select the parameters for the penalized GEE.

Key Words: Bridge regression, LASSO, Ridge regression, Penalized Generalized Estimating Equations

TEŞEKKÜR

Çukurova Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Anabilim dalında Yüksek Lisans tezi olarak hazırlanan bu çalışma Prof. Dr. Fikri Akdeniz danışmanlığında gerçekleştirilmiştir.

Bu tezin başlangıcından bitişine kadar sahip olduğu engin bilgi ve deneyimini benden esirgemeyen, değerli zamanını bana ayıran sevgili hocam ve danışmanım Prof. Dr. Fikri Akdeniz' e sonsuz teşekkür ederim.

Ayrıca eğitim ve öğretim hayatım boyunca benden maddi ve manevi katkılarını esirgemeyen aileme teşekkürlerimi bir borç bilirim.

İÇİNDEKİLER	SAYFA
ÖZ.....	I
ABSTRACT	II
TEŞEKKÜR.....	III
İÇİNDEKİLER	IV
TABLolar LİSTESİ	VI
ŞEKİLLER LİSTESİ	VII
KISALTMALAR	VIII
1. GİRİŞ.....	1
1.1. Giriş.....	1
1.2. Daraltıcı Modellerde Bazı Temel Bilgiler.....	1
1.3. Büzücü Regresyon Tahminleri.....	4
1.3.1. LASSO ve Ridge Regresyon Tahminleri.....	4
1.3.2. Bridge Tahmin Edicileri.....	8
1.4. γ 'nın Seçimi.....	10
2. BRİDGE REGRESYON.....	12
2.1. Giriş.....	12
2.2. Bridge Tahmin Edicinin Yapısı.....	12
2.3. Bridge ve LASSO Tahmin Ediciler için Algoritmalar.....	14
2.4. Bridge Tahmin Edicinin Varyansı.....	20
2.5. Büzülme Etkisinin Gösterimi.....	22
2.6. Ortonormal Matris için Bridge Regresyon.....	27
2.7. Bayeşçi Prior Olarak Bridge Ceza.....	29
2.8. Ayar Parametresi λ ve t Arasındaki İlişki.....	33
3. CEZALI SKOR DENKLEMLERİ.....	36
3.1. Giriş.....	36
3.2. Genelleştirilmiş Tahmin Denklemleri ve Likelihood.....	36
3.3. Yarı-Likelihood ve Yarı-Skor Fonksiyonları.....	41
3.4. Cezalı Skor Denklemleri.....	44
3.5. Cezalı Skor Denklemleri için Algoritmalar.....	46

4. CEZALI GEE.....	48
4.1. Giriş.....	48
4.2. Genelleştirilmiş Tahmin Denklemleri.....	48
4.3. Cezalı GEE.....	55
5. BÜZÜLME PARAMETRESİNİN SEÇİMİ.....	57
5.1. Giriş.....	57
5.2. Çapraz Geçerlilik ve Genelleştirilmiş Çapraz Geçerlilik Testi.....	57
5.3. GCV ile λ ve γ Parametrelerinin Seçimi.....	58
5.4. Cezalı GEE için Yarı GCV.....	60
6. SİMÜLASYON ÇALIŞMALARI.....	66
6.1. Lineer Regresyon Modeli.....	66
6.2. Karmaşık Lineer Regresyon Modeli.....	69
7. SONUÇ VE ÖNERİLER.....	73
7.1. Sonuçlar.....	73
7.2. Öneriler.....	73
KAYNAKLAR.....	74
ÖZGEÇMİŞ.....	78
EKLER (Tanım ve Matlab Kodları).....	79

TABLolar LİSTESİ**SAYFA**

Tablo 1.1. Korelasyon katsayıları ile artan varyans.....	3
Tablo 1.2. Ridge tahmin edicinin varyansı, yanlılık ve MSE değerleri	7
Tablo 2.1. Ortonormal X için Bridge tahmin ediciler ve standart hatalar.....	24
Tablo 2.2. Ortonormal olmayan X için Bridge tahmin ediciler ve standart hatalar.....	26
Tablo 6.1. 200 kez tekrarlayan simülasyonla model karşılaştırılması.....	68
Tablo 6.2. Farklı γ değerleri için MSE_r ve PSE_r 'nin ortalama ve standart hataları.....	72

ŞEKİLLER LİSTESİ

SAYFA

Şekil 1.1. İki boyutlu parametre uzayında $t=1$ için Bridge regresyonun sınırlandırıldığı bölge.....	9
Şekil 2.1. γ 'nın farklı değerleri için RHS fonksiyonunun farklı şekilleri...	17
Şekil 2.2. Algoritmalar.....	19
Şekil 2.3. Sabit $\lambda \geq 0$ için Bridge regresyonların büzülme etkisi.....	28
Şekil 2.4. $\lambda = 0.5$ ile Bayesçi prior olarak Bridge ceza.....	31
Şekil 2.5. $\lambda = 10$ ile Bayesçi prior olarak Bridge ceza.....	32
Şekil 2.6. $p = 2$ ve $c_j = 1$ özel durumu için $t(\lambda) - \lambda$ grafiği.....	35
Şekil 5.1. GCV ile λ ve γ parametrelerinin seçimi.....	60
Şekil 5.2. Yarı- GCV ile λ ve γ parametrelerinin seçimi.....	62

KISALTMALAR

CV	: Çapraz Geçerlilik Testi (Cross-Validation)
EKK	: En Küçük Kareler
GCV	: Genelleştirilmiş Çapraz Geçerlilik Testi (Generalized Cross-Validation)
GEE	: Genelleştirilmiş Tahmin Denklemleri (Generalized Estimation Equations)
GLM	: Genelleştirilmiş Lineer Modeller
IRLS	: Yeniden Ağırlıklandırılmış En Küçük Kareler
LHS	: Sol kısımdaki fonksiyon (Left Hand Side)
ML	: Maksimum likelihood
MLE	: Maksimum likelihood tahmincisi (Maximum Likelihood Estimator)
MNR	: Düzeltilmiş Newton-Raphson (Modified Newton-Raphson)
MSE	: Hata Kareler Ortalaması (Mean Squares Error)
OLS	: En Küçük Kareler (Ordinary Least Squares)
PCR	: Temel Bileşenler Regresyonu (Principal Component Regression)
PSE	: Öntahmin Hata Kareler Ortalaması (Predicted Squares Error)
RHS	: Sağ kısımdaki fonksiyon (Right Hand Side)
RSS	: Hata Kareler Toplamı (Residual Sum of Squares)

1. GİRİŞ

1.1. Giriş

Halk sağlığı çalışmalarında ya da birçok uygulamalı bilim dalında araştırmacılar, yanıt (response) değişkenler ve açıklayıcı değişkenler arasındaki ilişkilerle ilgilenirler. Örneğin göğüs kanseri çalışmasında, nüfusta kanser artışıdaki olasılığın hastanın diyeti, yaşı, boyu ve ağırlığı gibi bazı potansiyel risk faktörlerine bağlı olup olmadığını bilmek isterler.

İstatistiksel analizlerin amacı, meydana gelen olaya önemli katkıda bulunması nedeniyle risk faktörlerini belirlemektir. Çok nadir, olasılık teorisi ve istatistiksel modelleme merkezli olan analiz, regresyon dediğimiz istatistiksel süreç boyunca yürütülür. Regresyon analizi, risk faktörlerinin bulunması için bilgi sağlar ve böylece bilimsel kararlar olarak araştırmacılara yardımcı olur.

Bazı çalışmalarda açıklayıcı değişkenlerde doğrusal ilişki mevcuttur, bazıları diğerlerine doğrusal olarak bağlıdır. Buna içilişki denir. Açıklayıcı değişkenler arasındaki içilişkinin varlığı regresyon modellerinde geniş varyasyona ve belirsizliğe neden olduğundan, model parametrelerinin tahmini büyük varyansa sahip olur ve model üzerindeki öntahmin çok zayıf olur. Bu yüzden modeller araştırmacıların ihtiyaçlarına uygun olmaz.

Bu tezde içilişki problemi araştırılmış ve istatistiksel metot olan Bridge cezalandırma tekniği önerilmiştir. Ayrıca istatistiksel simülasyonlar gösterilmiş ve bu metot tahmin ve öngörü bakımından iyi çalışmıştır.

1.2. Daraltıcı Modellerde Bazı Temel Bilgiler

$$y = X\beta + \varepsilon \quad (1.1)$$

lineer regresyon modelini düşünelim. Burada, y ; $n \times 1$ tipinde rasgele yanıtlar vektörü, X ; $n \times p$ tipinde tasarım matrisi, β ; $p \times 1$ tipinde regresyon

parametrelerinin vektörü, ε ; $n \times 1$ tipinde rasgele hatalar vektörü $\varepsilon \sim iid(N(0; \sigma^2))$ dir. Amacımız; (1.1) nolu modeldeki β 'yı tahmin etmektir. Bunun için

$$RSS = (y - X\beta)^T (y - X\beta) = \varepsilon^T \varepsilon$$

kareler toplamı kullanılır ve minimum yapan β bulunur. β 'nın en küçük kareler tahmin edicisi

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$$

ve

$$Var(\hat{\beta}_{OLS}) = (X^T X)^{-1} \sigma^2$$

dir. Yani herhangi bir lineer yansız tahmin edici $\hat{\beta}$ için;

$$\hat{\beta} = Ay, \quad E(\hat{\beta}) = \beta \quad \text{ve} \quad Var(\hat{\beta}_{OLS}) \leq Var(\hat{\beta})$$

dir. Bu nedenle $\hat{\beta}_{OLS}$; Gauss-Markov koşulları altında en iyi lineer yansız tahmin edicidir (BLUE). Ancak yansızlık ve minimum varyans $\hat{\beta}_{OLS}$ için her zaman yeterli değildir.

1. Regresyon matrisi X tam ranklı değilse, tahmin edici tek değildir. Aslında hata kareler toplamını minimum yapan pek çok tahmin edici vardır.
2. Regresyon matrisi X'de yaklaşık içilişki problemi varsa EKK tahmin edici yine yansızdır fakat $Var(\hat{\beta}_{OLS}) = (X^T X)^{-1} \sigma^2$ varyansı büyük olur.

Hata kareler ortalaması (MSE)

$$MSE = E[(\hat{\beta} - \beta)^T (\hat{\beta} - \beta)] = \sum_j \left\{ [bias(\beta_j)]^2 + Var(\hat{\beta}_j) \right\} = \sum_j \{ Var(\hat{\beta}_j) \}$$

dir. Örneğin iki açıklayıcı değişkenli basit regresyon problemini düşünelim.

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Buradaki $\varepsilon \sim N(0; \sigma^2)$ dağılımına sahiptir. Regresörler arası ilişilşkinin etkilerini örnekte gösterelim. Kolaylık için $\bar{x}_j = 0$, $\|x_j\| = 1$, $j = 1, 2$ ve $\sigma^2 = 1$ ile regresyon vektörleri x_1 ve x_2 'yi standartlaştırıyoruz. Örneklem korelasyon katsayısı $r = x_1^T x_2$ ve

$$X^T X = \begin{pmatrix} x_1^T x_1 & x_1^T x_2 \\ x_2^T x_1 & x_2^T x_2 \end{pmatrix} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

dir. Böylelikle $\hat{\beta}_{OLS} = (\hat{\beta}_1 \quad \hat{\beta}_2)^T$ tahmin edicisinin varyans-kovaryans matrisi

$$Var(\hat{\beta}_{OLS}) = (X^T X)^{-1} = \frac{1}{1-r^2} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}$$

ve

$$Var(\hat{\beta}_j) = \frac{1}{1-r^2}, \quad j = 1, 2$$

dir. x_1 ve x_2 açıklayıcı değişkenleri ilişkisiz ise, yani $r = 0$ ise, $Var(\hat{\beta}_j) = 1$ fakat x_1 ve x_2 ilişkili ise $Var(\hat{\beta}_j)$ çok büyük olacaktır. Örneğin $r = 0.95$ için $Var(\hat{\beta}_j) = 10.26$ dır.

Tablo 1.1. Korelasyon katsayıları ile artan varyans

r	0	0.3	0.5	0.7	0.95
$Var(\hat{\beta}_j)$	1.00	1.10	1.33	1.96	10.26

Hata kareler ortalaması tahminin doğruluğunu yansıttığından, büyük MSE kötü tahmin anlamına gelir. X 'te ilişilşki varsa, $\hat{\beta}_{OLS}$ 'ye dayalı öntahmin çok kötü performans verecektir. Örneğin 2 açıklayıcı değişkenli öntahmin hata kareler

ortalamasını (PSE) düşünelim. OLS tahmin edicisi $\hat{\beta}$ ile (x^{*T}, y^*) gibi keyfi bir noktada öntahmin hatasının beklenen değeri

$$\begin{aligned} E(PSE) &= E\left(y^* - x^{*T}\hat{\beta}\right)^2 \\ &= E\left(x^{*T}\beta + \varepsilon^* - x^{*T}\hat{\beta}\right)^2 \\ &= \sigma^2 + Var\left(x^{*T}\hat{\beta}\right) + \left[bias\left(x^{*T}\hat{\beta}\right)\right]^2 \\ &= \sigma^2 \left[1 + x^{*T}\left(X^T X\right)^{-1} x^*\right] \end{aligned}$$

olur. Buradaki ε^* ; öntahmin noktasındaki rasgele hata, σ^2 ; rasgele hataların varyansıdır. Böylece PSE değeri x^* vektörünün konumuna bağlıdır.

Özel olarak yüksek içilişkili $X^T X = diag(1 \ 0.001)$ alalım.

$E(PSE) = \sigma^2 \left[1 + x_1^{*2} + 1000x_2^{*2}\right]$ 'dir. $|x_2^*| \ll \max\{1, |x_1^*|\}$ ise hatanın öntahmini yeterlidir. Aksi halde yüksek içilişki yüzünden x_2^* 'nin faktörü büyük ölçüde şişirilmiştir. Çoklu içilişki, Hoerl ve Kennard (1970a,b), Lawson ve Hansen (1974), Seber (1977), Sen ve Srivastava (1990), Frank ve Friedman (1993), Hocking (1996)'de detaylı olarak incelenmiştir.

1.3. Büzücü Regresyon Tahminleri

1.3.1. LASSO ve Ridge Regresyon Tahminleri

LASSO ve Ridge regresyon, tahminleri dengelediği için “Büzülme Modelleri” olarak adlandırılırlar. LASSO ve Ridge regresyon farklı cezalandırma kullanırlar.

$y = X\beta + \varepsilon$ lineer regresyon modeli için her bir bileşen

$$y_i = \beta_0 + \sum_{j=1}^m x_{ij}\beta_j + \varepsilon_i, \quad i=1,2,\dots,n$$

olsun. Tibshirani (1996) En Küçük Mutlak Büzülme ve Operatör Seçimini (LASSO, Least Absolute Shrinkage And Selection Operator);

$$\sum |\beta_j| \leq t \quad \text{koşulu ile} \quad \min_{\beta} (y - X\beta)^T (y - X\beta)$$

olarak vermiştir. LASSO, EKK tahmin edicisi $\hat{\beta}_{OLS}$ 'yi sıfıra büzebilir ve böylece bazı j değerleri için $\hat{\beta}_j = 0$ olabilir.

LASSO regresyonun amaç fonksiyonu

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m x_{ij}\beta_j \right)^2 + \lambda_L \sum_{j=1}^m |\beta_j| \rightarrow \min$$

$C(t)$, t 'ye bağlı ve j 'den bağımsız pozitif bir sabit olmak üzere, Tibshirani (1996) ortonormal X matrisi için LASSO tahmin edicisini

$$\hat{\beta}_{lasso}^{(j)} = \text{sign}\left(\hat{\beta}_{OLS}^{(j)}\right) \left(\left| \hat{\beta}_{OLS}^{(j)} \right| - C(t) \right)^+, \quad j=1,2,\dots,p$$

olarak vermiştir. t parametresi genelleştirilmiş çapraz geçerlilik testi (GCV) ile optimize edilir. LASSO sadece 0'a büzen değil, aynı zamanda iyi tahmin sağladığı da simülasyon çalışmalarından görülmektedir (Fu, 1998).

Ridge regresyonun amaç fonksiyonu

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m x_{ij}\beta_j \right)^2 + \lambda_R \sum_{j=1}^m \beta_j^2 \rightarrow \min$$

dir. Problemin çözümü

$$\hat{\beta}_{rdg} = \left(X^T X + \lambda_R I \right)^{-1} X^T y$$

dir. $\hat{\beta}_{rdg}$ yanlı ridge tahmin edicisi, $\lambda_R = 0$ olduğunda $\hat{\beta}_{OLS}$ tahmin edicisine büzülür. Buradaki λ , yanlılık/ayar parametresidir. Varyansı $\lambda_R > 0$ için $\hat{\beta}_{OLS}$ 'den daha küçüktür.

$$\begin{aligned} Var(\hat{\beta}_{rdg}) &= (X^T X + \lambda_R I)^{-1} X^T X (X^T X + \lambda_R I)^{-1} \sigma^2 \\ &< Var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1} \end{aligned}$$

λ_L ve λ_R parametreleri büzülme miktarını kontrol eder ve sıfır veya daha büyük değerler seçilmelidir. Eğer parametre 0 ise; EKK gibi LASSO ve Ridge regresyonun sonuçları da aynıdır. Parametre seçimi CV (çapraz geçerlilik testi) veya bootstrap ile olur. Öntahmin hatasını minimize edecek şekilde seçilmelidir. Ridge ve LASSO regresyon PCR (Temel Bileşenler Regresyonu) ve değişken seçimine alternatiftir. Ridge regresyon, regresyon modelindeki tüm x değişkenlerini kullanırken, LASSO regresyon sadece x değişkenlerinin bir alt kümesini kullanır. Her iki model de, en iyi modeli bulmak için parametre seçimine bağlıdır (Hoerl ve Kennard, 1970a; Tibshirani, 1996; Varmuza ve Filzmoser, 2009).

Daraltıcı modeller ile ilgili çok sayıda çalışma bulunmaktadır. Knight ve Fu (2000), LASSO tipli tahmin edicilerin asimptotik özelliklerini çalışmış; Fan ve Li (2001), lineer regresyonda cezalı likelihood metotlarını çalışmıştır ve LASSO, Ridge ve Bridge bunun özel durumlarıdır. Huang (2003), LASSO tahmin edicisinin öntahmin hatasını vermiş ve LASSO tahmin ediciyi genelleştirmiştir. Yuan ve Lin (2004), Grup LASSO 'yu; Tibshirani ve ark. (2005), Birleştirilmiş LASSO 'yu (Fused-LASSO) önermişlerdir. Zou ve Hastie (2005), Bridge tahmin edicileri çalışmış ve buradan yola çıkarak Elastic-Net tahmin edicilerini önermişlerdir. Wang, Li ve Tsai (2007), Düzeltilmiş LASSO önermiş ve bunun için algoritma vermiştir. Huang ve ark. (2008), $0 < \gamma < 1$ ile Bridge regresyonun asimptotik özelliklerini çalışmıştır.

Ridge ve LASSO regresyon için amaç fonksiyonlarındaki tek fark, regresyon katsayılarını farklı cezalandırma yoludur. Ridge regresyon L_2 normu (regresyon

katsayılarının kareleri toplamı) ile cezalandırılırken, LASSO regresyon L_1 normu (regresyon katsayılarının mutlak değer toplamı) ile cezalandırılır.

Ridge regresyonun büzülme etkisini örneklendirelim. İki açıklayıcı değişkenli bir lineer regresyon problemini düşünelim. Ridge tahmin edicinin varyansı

$$Var(\hat{\beta}_{rdg}) = \begin{pmatrix} 1+\lambda & r \\ r & 1+\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} 1+\lambda & r \\ r & 1+\lambda \end{pmatrix}^{-1}$$

yanlılık

$$Bias(\hat{\beta}_{rdg}) = \begin{pmatrix} 1+\lambda & r \\ r & 1+\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

ve hata kareler ortalaması

$$MSE = Var(\hat{\beta}_j) + [Bias(\hat{\beta}_j)]^2$$

ile verilir.

Tablo 1.2. Ridge tahmin edicinin varyansı, yanlılık ve MSE değerleri

	$\lambda = 0$			$\lambda = 1$			$\lambda = 5$			$\lambda = 10$		
r	Var	bias ²	MSE	Var	bias ²	MSE	Var	bias ²	MSE	Var	bias ²	MSE
0	1.000	0	1.000	0.250	0.250	0.500	0.028	0.694	0.722	0.008	0.826	0.834
0.5	1.333	0	1.333	0.231	0.160	0.391	0.026	0.592	0.618	0.008	0.756	0.764
0.9	5.263	0	5.263	0.154	0.119	0.273	0.022	0.525	0.547	0.007	0.706	0.713

bias² ve MSE değerleri gerçek $\beta = (1 \ 1)^T$ ile hesaplanmıştır.

x_1 ve x_2 ilişkisiz ise, yani $r = 0$ ise, $\lambda = 1$ için $Var(\hat{\beta}_{j\ rdg}) = 1/(1 + \lambda)^2 = 0.25$
 $\lambda = 0$ için $Var(\hat{\beta}_{j\ OLS}) = 1$ 'den daha küçüktür. Eğer x_1 ve x_2 ilişkili ise, örneğin
 $r = 0.9$ ise, $\lambda = 1$ için $Var(\hat{\beta}_{j\ rdg}) = 0.15$, $\lambda = 0$ için $Var(\hat{\beta}_{j\ OLS}) = 5.26$ 'dan çok
daha küçüktür.

Yukarıdaki tabloda λ 'ya göre $bias^2$ 'nin artışı görülmektedir. $(Bias)^2$,
 $bias(\hat{\beta}_j) = -\lambda\beta/(1 + \lambda + r)$ olarak hesaplanmıştır ($\beta_1 = \beta_2 = \beta = 1$ özel durumu için).
Varyans değerlerinde λ arttıkça azalma görülmektedir. MSE değeri ise; $\lambda = 0$ 'dan
1'e giderken azalmakta; $\lambda = 1$ 'den 5 veya 10'a giderken artmaktadır.

LASSO regresyon katsayıları y 'nin bir fonksiyonu olarak yazılamaz.
Kuadratik programlama ile optimize edilerek çözümü bulunabilir (Tibshirani, 1996).

1.3.2. Bridge Tahmin Edicileri

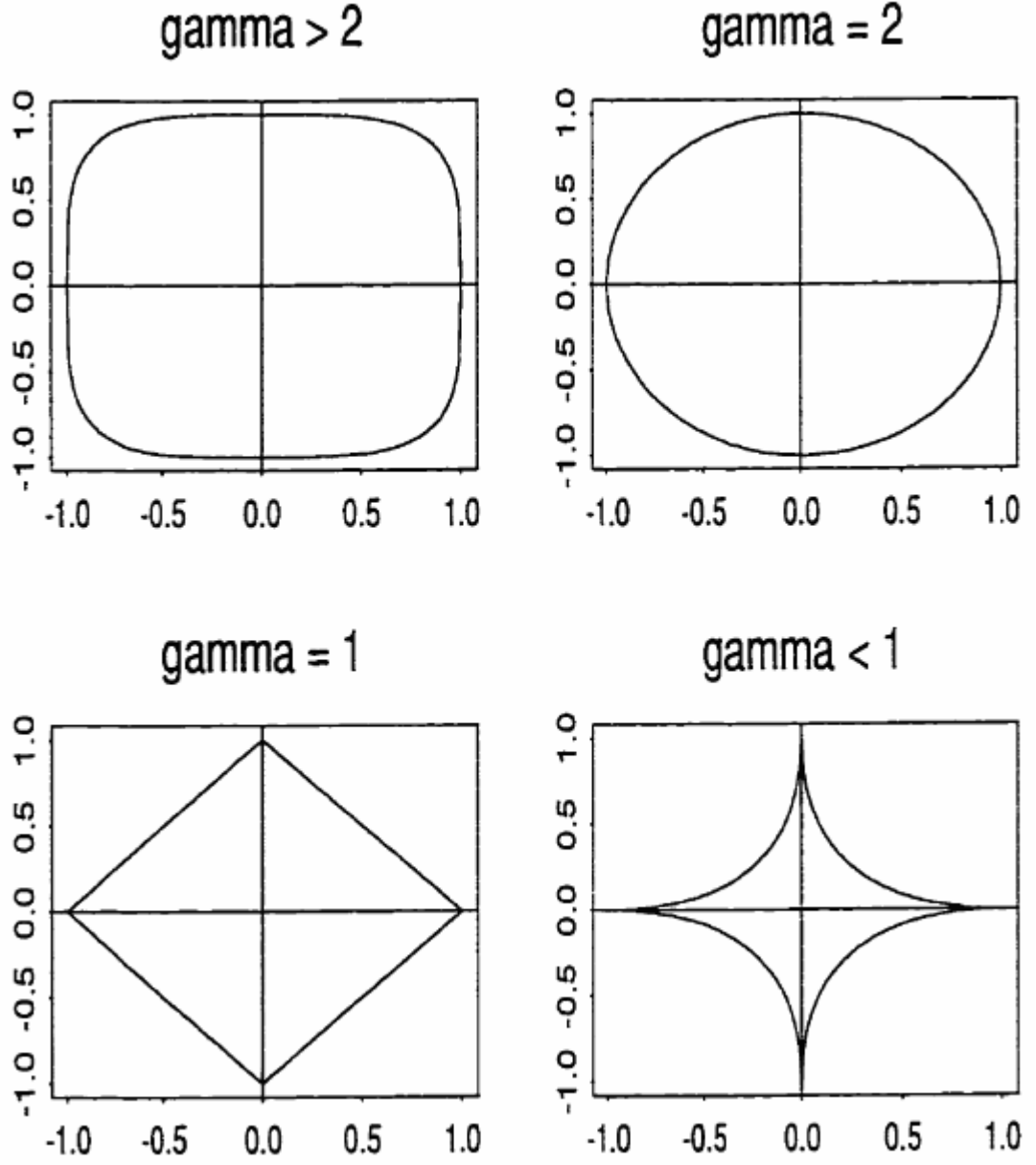
Aşağıdaki cezalı hata kareleri toplamını

$$L_n(\beta) = \sum_{i=1}^n (y_i - x_i'\beta)^2 + \lambda \sum_{j=1}^p |\beta_j|^\gamma$$

ya da eşdeğer olarak

$$\lambda > 0 \text{ ile } \sum |\beta_j|^\gamma \leq t \text{ kısıtı ile } \min_{\beta} (y - X\beta)^T (y - X\beta)$$

yı minimum yapan $\hat{\beta}_n$ değerine Bridge tahmin edici denir (Frank ve Friedman, 1993; Fu, 1998). $\gamma = 2$ iken çok iyi bilinen Ridge tahmin edici, $\gamma = 1$ iken LASSO tahmin edici adını alır (Tibshirani, 1996). Şekil 1.1 'de $t = 1$ için iki boyutlu parametre uzayında orijin çevresinde farklı tahminler gösterilmiştir (Frank ve Friedman, 1993)



Şekil 1.1. İki boyutlu parametre uzayında Bridge regresyonun sınırlandırıldığı bölge
(Fu,1998)

1.4. γ 'nın Seçimi

Ridge ve LASSO'nun performansı OLS'den daha iyi olmasına rağmen X 'de içilişki olduğunda Frank-Friedman (1993) ve Tibshirani (1996) bazı durumlarda LASSO'nun Ridge regresyondan, bazı durumlarda da Ridge regresyonun LASSO'dan daha iyi performans gösterdiğine dikkat çekmişlerdir. Bunu bulabilmek için bazı sorular sormuşlardır: En iyi performans için γ ne olmalıdır? Optimal γ değeri nasıl seçilmeli?

Bu sorulara cevap için, $\gamma=1$ LASSO veya $\gamma=2$ Ridge seçimi gibi, veriye dayanarak seçilen optimal γ değerine tercih edilen bazı teknikler geliştirilmelidir.

Bu tezde özellikle

$$\gamma \geq 1 \text{ ve } \sum |\beta_j|^\gamma \leq t \quad \text{kısıtı ile} \quad \min_{\beta} (y - X\beta)^T (y - X\beta)$$

üzerinde çalışılmıştır.

Bölüm 2'de herhangi sabit $\gamma \geq 1$ olan Bridge tahmin ediciyi çözmek için yeni algoritmalar geliştirilmesi ve Bridge tahmin edicinin yapısı çalışılmıştır. LASSO tahmin ediciyi daha basit ve daha kolay hesaplayabilmek için yeni algoritmalar geliştirilmiştir. Bridge tahmin edicinin varyansı elde edilmiştir. Lineer regresyonun basit bir örneği ile Bridge regresyonun büzülme etkisi tanımlanmıştır ve ortonormal regresyon matris durumu için teorik olarak sınanmıştır. Bridge ceza fonksiyonu Bayesçi prior gibi çalışır. Bölüm 3'te genelleştirilmiş lineer modeller (GLM), likelihood fonksiyonları, yarı likelihood yeniden gözden geçirilmiştir. Bridge regresyon, genelleştirilmiş lineer modellere genişletilmiştir. Cezalı skor denklemlerini açarak ortak likelihood fonksiyonlarının bağımsızlaşması, cezalandırmayı genelleştirmiştir. Cezalı skor denklemlerini çözen algoritmalar geliştirilmiştir. Bölüm 4'te longitudinal (boylamasına) çalışmalarda genelleştirilmiş tahmin denklemleri (GEE) ve cezalı skor denklemleri tarafından genelleştirilmiş tahmin denklemlerine cezalandırma uygulanmıştır. Bölüm 5'te çapraz geçerlilik testi (CV) ve genelleştirilmiş çapraz geçerlilik testi (GCV) metotları incelenmiştir. Büzülme parametresi γ ve ayar parametresi λ GLM için GCV ile seçilmiştir.

Yarı-GCV, cezalı GEE için γ ve λ seçmek için elde edilmiştir. Bölüm 6’da simülasyon çalışmalarından Bridge model ile diğer büzülme modelleri: büzülmesiz, LASSO ve Ridge karşılaştırılmıştır. Bölüm 7’de sonuç ve öneriler verilmiştir. Ekte ise matematiksel ispatlar, tanım ve MATLAB kodları verilmiştir.

2. BRIDGE REGRESYONLAR

2.1. Giriş

Bölüm 1’de kısaca regresyonlar ve büzülme modelleri tanıtılmış, özellikle de Bridge regresyonlar ele alınmıştır. Bridge regresyonlar önermeye rağmen, henüz tahmin edicileri üzerinde çalışılmamıştır. Frank ve Friedman (1993)’ın belirttiği gibi elde edilen en iyi sonuç için optimal γ değeri seçim yöntemi çalışması önemlidir.

Bu bölümde Bridge regresyon ve tahmin edicileri çalışılmıştır. Herhangi $\gamma > 1$ verilmişken Bridge tahmin edicisi çözümü için düzeltilmiş Newton-Raphson metodu (MNR) verilmiş ve bir algoritma tasarlanmıştır. Bir de LASSO tahmin edici çözen Atış metodu ve bunun için yeni bir algoritma tasarlanmıştır.

Bridge tahmin edicinin varyansı delta metodu ile elde edilmiştir. Büzülme etkisi ortonormal regresyon matris tipi için teorik olarak ispatlanmış ve basit bir örnekle gösterilmiştir.

2.2. Bridge Tahmin Edicinin Yapısı

$\gamma \geq 1$ verilmişken Bridge regresyonun çözümünde aşağıdaki iki problem dikkate alınır (Fu, 1998).

$$\gamma \geq 1 \text{ ve } t \geq 0 \text{ verilmişken } \sum |\beta_j|^\gamma \leq t \text{ koşulu ile } \min_{\beta} RSS \quad (2.1)$$

$$\gamma \geq 1 \text{ ve } \lambda \geq 0 \text{ verilmişken } \min_{\beta} \left(RSS + \lambda \sum |\beta_j|^\gamma \right) \quad (2.2)$$

Bu problemin ikisi de denktir. Yani $\lambda \geq 0$ için $t \geq 0$ ’ın var olması veya tersi olarak $t \geq 0$ için $\lambda \geq 0$ ’ın var olmasıdır. İki problem de aynı çözümü paylaşır.

(2.2) problemi $\sum |\beta_j|^\gamma$ ceza ve λ ayar parametresi ile cezalı regresyon gibi yorumlanabilir.

(2.2) 'yi düşünürsek, $G(\beta, X, y, \lambda, \gamma) = RSS + \lambda \sum |\beta_j|^\gamma$ olsun. Öklid normu $\|\beta\| \rightarrow +\infty$, $G \rightarrow +\infty$ 'dir. Böylece G fonksiyonu minimize edilir. Yani

$$\hat{\beta} = \arg \min_{\beta} G(\beta, X, y, \lambda, \gamma)$$

dır.

$\beta_j = 0$ 'da $\sum |\beta_j|$ fonksiyonu diferansiyellenebilir olmadığından, $\beta_j \neq 0$ 'da G 'nin β_j 'ye göre kısmi türevini alabiliriz.

$$S_j(\beta, X, y) = \frac{\partial RSS}{\partial \beta_j} \quad \text{ve} \quad d(\beta_j, \lambda, \gamma) = \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j)$$

olsun. $\frac{\partial G}{\partial \beta_j} = 0$ alındığında

$$\begin{cases} S_1(\beta, X, y) + d(\beta_1, \lambda, \gamma) = 0 \\ \dots \\ S_p(\beta, X, y) + d(\beta_p, \lambda, \gamma) = 0 \end{cases} \quad (2.3)$$

yazılabilir (Fu, 1998). Bir sonraki bölümde görülebileceği gibi (2.2) 'nin çözümü bizi (2.3) 'e götürecektir.

(2.3) 'ün nasıl çözüldüğünü görebilmek için,

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

doğrusal regresyon modelini düşünelim. Hata kareler toplamı

$RSS = \sum_i (y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2$ 'dir. G fonksiyonunun β_j 'ye göre kısmi türevlerini

aldığımızda, (2.3) 'teki

$$\begin{cases} -2 \sum_i x_{i1} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2}) + \lambda \gamma |\beta_1|^{\gamma-1} \text{sign}(\beta_1) = 0 \\ -2 \sum_i x_{i2} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2}) + \lambda \gamma |\beta_2|^{\gamma-1} \text{sign}(\beta_2) = 0 \end{cases}$$

denklemleri olur.

2.3. Bridge ve LASSO Tahmin Ediciler için Algoritmalar

$\gamma \geq 1$ ve $\lambda > 0$ verilmişken Bridge regresyonun çözümü için (2.3) problemi ile başlayalım. Aşağıda sadece Gauss yanıt değişkenleri için metot göstermemize rağmen, “İteratif Yeniden Ağırlıklandırılmış EKK” (IRLS) yöntemi ile yanıt değişkeninin diğer birçok türü için algoritmalar uygulanabilir.

β^{-j} , β_j hariç diğer β_l 'leri içeren $p-1$ vektör olmak üzere, $(\beta_j, \beta^{-j})^T$ ile β olsun.

(2.3) 'ün j. denklemini düşünelim.

$$S_j(\beta_j, \beta^{-j}, X, y) = -d(\beta_j, \lambda, \gamma) \quad (2.4)$$

(2.4) denkleminin solundaki fonksiyon

$$LHS = 2x_j^T x_j \beta_j + \sum_{i \neq j} 2x_j^T x_i \beta_i - 2x_j^T y$$

sabit β^{-j} için $2x_j^T x_j$ pozitif eğimi ile β_j 'nin lineer bir fonksiyonudur.

(2.4) denkleminin sağındaki fonksiyon

$$RHS = -\lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j)$$

β_j 'de lineer değildir. Şekil 2.1 farklı γ değerleri için farklı RHS fonksiyonunu göstermektedir. RHS fonksiyonu $\beta_j = 0$ dışında $\gamma > 1$ için sürekli diferansiyellenebilir ve monoton azalandır. $1 < \gamma < 2$ için $\beta_j = 0$ 'da diferansiyellenebilir değildir. $\gamma = 1$ için $\beta_j = 0$ 'da 2λ boy sıçraması ile bir heavy-side fonksiyondur. Bu nedenle (2.4) denklemi $\gamma > 1$ için tek çözümdür ya da çözümü yoktur.

$\gamma > 1$ verilmişken Bridge tahmin ediciyi hesaplamak için Newton-Raphson metodu kullanılmıştır. Ancak $\gamma < 2$ verilmişken $\beta_j = 0$ 'da d fonksiyonu

diferansiyellenebilir olmadığından, çözüm yakınsama ile elde edilmeye çalışılmıştır. (2.3)'ün j . denkleminin tek çözümü için iteratif çözüm ile genelde $\gamma > 1$ için aşağıdaki düzeltilmiş Newton-Raphson metodu geliştirilmiştir (Fu, 1998).

$\gamma > 1$ Bridge için Düzeltilmiş Newton-Raphson (MNR) Algoritması

- (1). $\hat{\beta}_0 = \hat{\beta}_{OLS} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$ ile başla.
- (2). m . adımda, her bir $j = 1, \dots, p$ için $S_0 = S_j(0, \hat{\beta}^{-j}, X, y)$ olsun. $S_0 = 0$ ise $\hat{\beta}_j = 0$.

Diğer durumlarda $\gamma \geq 2$ ise, (2.4) denkleminin $\hat{\beta}_j$ tek çözümü için Newton-Raphson metodu uygula. $\gamma < 2$ ise, Şekil 2.2'de gösterildiği gibi tam orijin ve çözüm arasındaki noktada $(S_j, -d)$ kesişimi teğet çizgisini biraz değiştirerek $-d$ fonksiyonunu düzelt. Bu nokta ikiye bölme metodu ile bulunabilir.

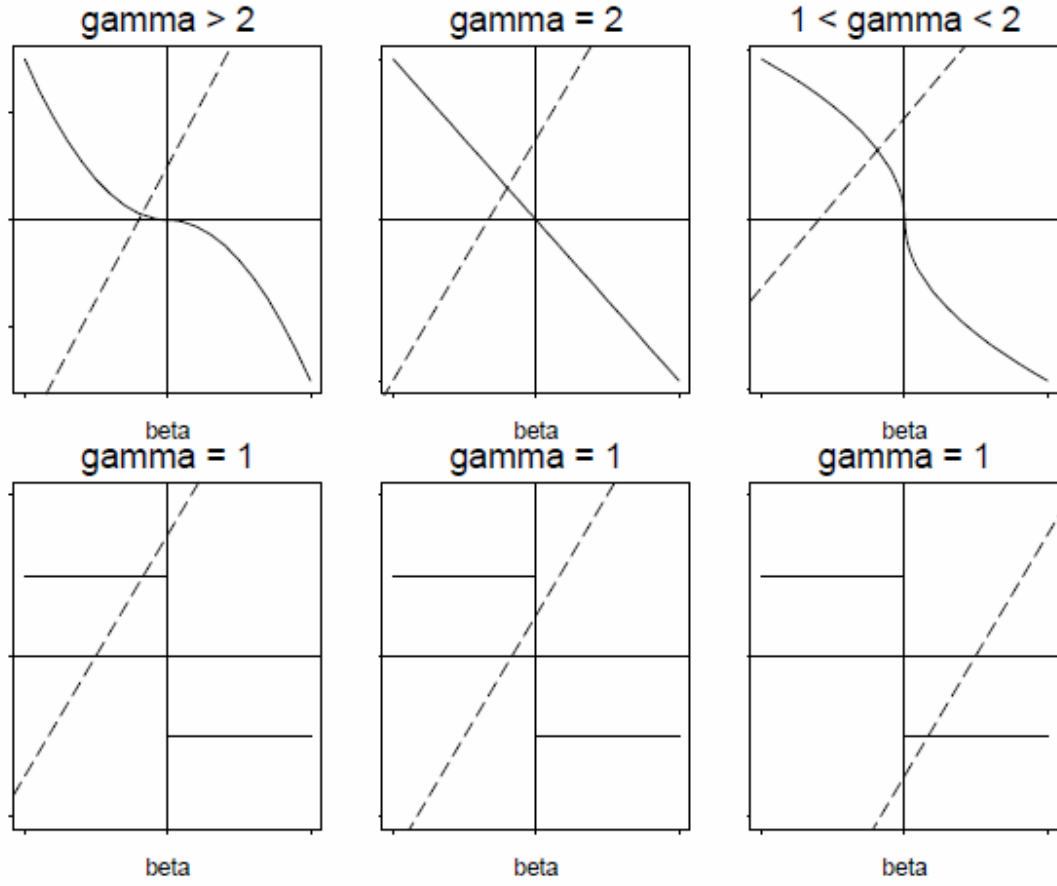
$\hat{\beta}_j$ tek çözümü için, değiştirilmiş $-d$ fonksiyonu ile (2.4) denkleminde Newton-Raphson metodu uygulanır. Tüm $\hat{\beta}_j$ 'ler bulunduğundan sonra yeni tahmin edici $\hat{\beta}_m = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$ şeklindedir.

- (3). $\hat{\beta}_m$ bir noktaya yakınsayıncaya kadar 2. adımı tekrarla.

Uyarılar

1. $\hat{\beta}_0$ başlangıç değeri için, her zaman OLS tahmin edici $\hat{\beta}_{OLS}$ kullanılabilir.
2. $p > n$ olduğunda, X tam ranklı değilse herhangi bir genel tahmin $\hat{\beta}_0$ 'nın başlangıç değeri olarak kullanılabilir.
3. Düzeltilmiş Newton-Raphson (MNR) algoritmasından, Bridge tahmin edici, bazı j 'ler için $\hat{\beta}_j = 0$ yapıyorsa, $\hat{\beta}_{brg}^{-j}$ de $S_j(0, \hat{\beta}_{brg}^{-j}, X, y)$ fonksiyonunu 0 yapar. Bu,

$\hat{\beta}_{brg}^{-j}$ ($p-1$) boyutlu vektörün ($p-2$) boyutlunun yerine geçmesi anlamına gelir ki bu da 0 ölçümlüdür. Bu nedenle $\hat{\beta}_j$ 'nin yaklaşık olarak sıfır değerini almadığı sonucuna varılır.



Şekil 2.1. γ 'nın farklı değerleri için RHS fonksiyonunun farklı şekilleri. Kesik çizgili S_j , düz çizgili ise $-d$ fonksiyonudur. Her bir grafikteki düşey eksen λ uzunlukludur (Fu, 1998)

LASSO için yeni bir algoritma olan Atış metoduna geçelim (Fu, 1998).

(1). $p = 1$. (2.3) problemini

$$2x^T x \beta - 2x^T y + \lambda \text{sign}(\beta) = 0 \quad (2.5)$$

olarak yazalım.

$\hat{\beta}_0$ 'nın ilk tahmin OLS tahmin edici ile başla. Şekil 2.2'de gösterildiği gibi yatay ekseninde $(\hat{\beta}_0, 0)$ noktasından $2x^T x$ eğimi yönünde at. Eğer sağ üstteki gibi bir noktaya denk geliyorsa $(-d = \lambda)$, ya da sağ alttaki gibi bir alt noktaya denk geliyorsa $(-d = -\lambda)$ (2.5) denklemini tek çözüme sahiptir, basit kapalı bir forma sahiptir ve LASSO tahmin edicisine eşittir. Eğer sol alt şekilde olduğu gibi herhangi bir noktaya denk gelmiyorsa, (2.5) denkleminin çözümü yoktur. LASSO tahmin edici için $\hat{\beta} = 0$ olur.

(2). $p > 1$. $\hat{\beta}_0$ 'nın ilk değeri OLS tahmin edici ile başla. m. adımda, (1)'i kullanarak $\hat{\beta}^{-j}$ sabiti için $\hat{\beta}_j$ 'ları güncelleyerek $\hat{\beta}_m$ 'yı hesapla. $\hat{\beta}_m$ bir noktaya yakınsayana kadar iterasyona devam et.

LASSO için Atış Algoritması

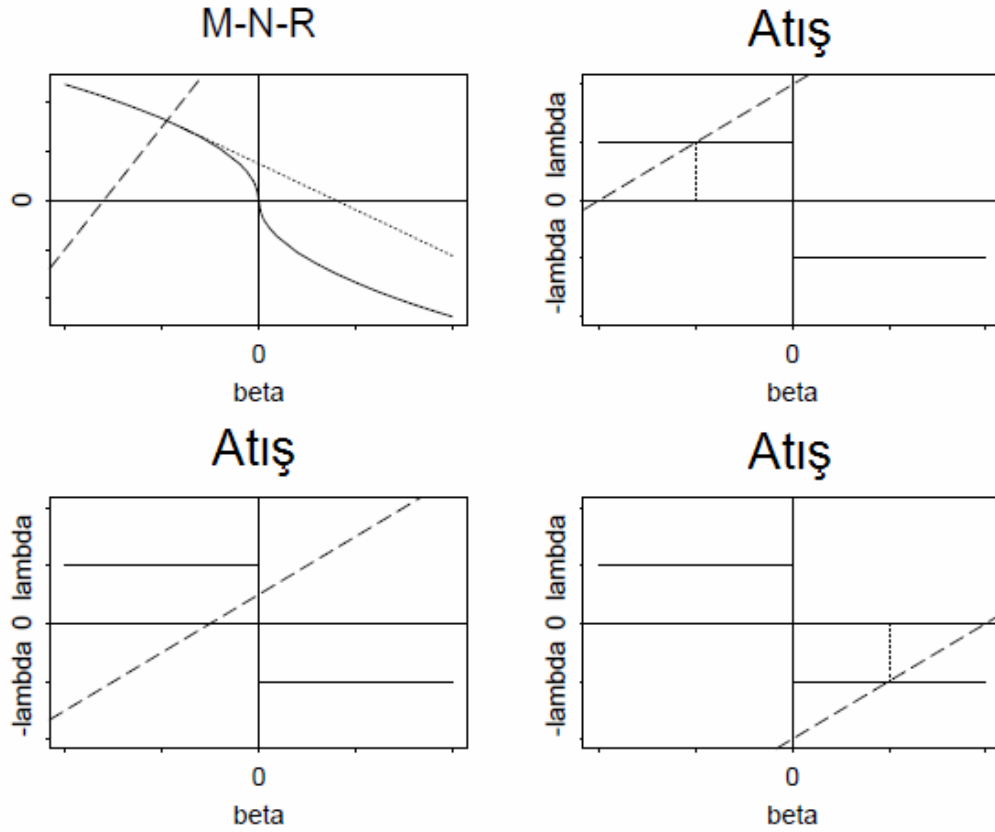
(1). $\hat{\beta}_0 = \hat{\beta}_{OLS} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$ ile başla.

(2). m. adımda, her bir $j = 1, \dots, p$ için $S_0 = S_j(0, \hat{\beta}^{-j}, X, y)$ olsun ve x_j , X 'in j. kolonu olmak üzere

$$\hat{\beta}_j = \begin{cases} \frac{\lambda - S_0}{2x_j^T x_j}, & S_0 > \lambda \\ \frac{-\lambda - S_0}{2x_j^T x_j}, & S_0 < -\lambda \\ 0, & |S_0| \leq \lambda \end{cases}$$

hesapla. Tüm $\hat{\beta}_j$ 'ları bulduktan sonra $\hat{\beta}_m = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$ bul.

(3). $\hat{\beta}_m$ bir noktaya yakınsayana kadar 2. adımı tekrarla.



Şekil 2.2. Algoritmalar. Kesik çizgili S_j , düz çizgili ise $-d$ fonksiyonudur. Sol üstteki şekil: noktalı çizgi ile teğete göre $-d$ 'nin değişimi; Sağ üstteki şekil: $S_0 > \lambda$; Sol alttaki şekil: $|S_0| \leq \lambda$; Sağ alttaki şekil: $S_0 < -\lambda$, çözüm noktalı çizgi ile gösterilmiştir (Fu, 1998)

2.4. Bridge Tahmin Edicinin Varyansı

$\gamma \geq 1$ Bridge tahmin edicinin varyansı

$$Var(\hat{\beta}) = \left(X^T X + D(\hat{\beta}) \Big|_{y_0} \right)^{-1} X^T Var(y) X \left(X^T X + D(\hat{\beta}) \Big|_{y_0} \right)^{-1} \quad (2.6)$$

Delta metodu kullanılarak (2.3) 'ten çıkarılır.

$$Var(f(y)) = \left(\frac{\partial f}{\partial y} \right) \Big|_{y_0} Var(y) \left(\frac{\partial f}{\partial y} \right)^T \Big|_{y_0}$$

Buradaki y_0 , örnek uzaydaki her bir noktadır. Varyans tahmini, tahmin edicisi ile $Var(y)$ yerine koyarak elde edilir.

$F_j = S_j(\hat{\beta}, X, y) + d(\hat{\beta}_j, \lambda, \gamma)$ olmak üzere $F = (F_1, F_2, \dots, F_p)^T$ olsun. Böylece (2.3) 'ten $F_j = 0$ olur. Gauss dağılımı için

$$\frac{\partial F}{\partial y} = -2X^T$$

ve

$$\frac{\partial F}{\partial \hat{\beta}} = 2X^T X + 2D(\hat{\beta})$$

Burada $D(\hat{\beta}) = \text{diag} \left(\frac{\lambda \gamma (\gamma - 1)}{2} |\hat{\beta}_j|^{\gamma-2} \right)$

ve böylece $2D(\hat{\beta}) = \text{diag} \left(\lambda \gamma (\gamma - 1) |\hat{\beta}_j|^{\gamma-2} \right)$. Bu da $d(\hat{\beta}_j, \lambda, \gamma)$ 'nın $\hat{\beta}$ 'ya göre türevinin köşegen elemanlarından oluşan matristir.

Kapalı fonksiyon türev teoremi ile

$$\begin{aligned}\frac{\partial \hat{\beta}}{\partial y} &= - \left(\frac{\partial F}{\partial \hat{\beta}} \right)^{-1} \frac{\partial F}{\partial y} \\ &= - \left(2X^T X + 2D(\hat{\beta}) \right)^{-1} (-2X^T) \\ &= \left(X^T X + D(\hat{\beta}) \right)^{-1} X^T\end{aligned}$$

Böylece, bulduğumuz sonuçları Delta metodunda yerine koyduğumuzda;

$$\begin{aligned}Var(\hat{\beta}) &= \left(\frac{\partial \hat{\beta}}{\partial y} \right) \Big|_{y_0} Var(y) \left(\frac{\partial \hat{\beta}}{\partial y} \right)^T \Big|_{y_0} \\ &= \left(X^T X + D(\hat{\beta}) \right)^{-1} X^T Var(y) X \left(X^T X + D(\hat{\beta}) \right)^{-1}\end{aligned}$$

dir. Bu şekilde (2.6) denklemi elde edilir.

Aşağıda 2 özel durum verilmiştir.

1. OLS regresyon, yani $\lambda = 0$. $D(\hat{\beta})$ fonksiyonu sıfır matrisi olur. Böylece

$$Var(\hat{\beta}) = \left(X^T X \right)^{-1} X^T Var(y) X \left(X^T X \right)^{-1}$$

2. Ridge regresyon, yani $\gamma = 2$. I birim matris ve $D(\hat{\beta}) = \lambda I$ olmak üzere

$$Var(\hat{\beta}) = \left(X^T X + \lambda I \right)^{-1} X^T Var(y) X \left(X^T X + \lambda I \right)^{-1}$$

Bu ridge tahmin edicinin varyansı, $Var(\hat{\beta}_{rdg})$ 'e eşittir.

LASSO bazı $\hat{\beta}_j = 0$ yaptığı için delta metodu uygulanamaz. Buna rağmen bootstrap ve jackknife metodu (Shao ve Tu, 1995) varyans hesaplamada kullanılabilir.

2.5. Büzülme Etkisinin Gösterimi

Bölüm 2.2 ve 2.3'te Bridge regresyon için tahmin edici ve algoritmalar, 2.4'te Bridge tahmin edicinin varyansı verilmiştir. Bu bölümde Bridge (LASSO) tahmin edicinin nasıl çözüleceği ile ilgileneceğiz. Basit örneklerle de Bridge tahmin edicinin büzülme etkisi gösterilecektir.

Ortonormal X matrisi ile örnek

40 gözlemlili basit doğrusal regresyon modelini

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

olarak düşünelim. Rasgele hatalar $\varepsilon \sim N(0, \sigma^2)$ olarak dağılsın. Ortonormal X matrisinin kolonları x_j vektörleri $\sum_i x_{ij} = 0$, $j = 1, 2, \dots, p$ ile standartlaştırılsın ve

$$x_l^T x_j = \begin{cases} 1, & l = j \\ 0, & l \neq j \end{cases}$$

olsun. Kolaylık için $\beta_0 = 0$ ve $\sigma^2 = 1$ diyelim. Y yanıtının 40 gözlemi, $\beta_1 = 1$, $\beta_2 = -2$, $\beta_3 = 5$ gerçek değerleri ile üretilsin. Sabitte büzülme etkisi olmadığından, sabit $\sum_i y_i = 0$ merkezileştirme ile silinir. $\lambda > 0$ ve $\gamma \geq 1$ için (2.3)'ün her bir denklemi, $j = 1, 2, \dots, p$ için

$$-2 \sum_i x_{ij} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) + \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0$$

$$2\beta_j - 2 \sum_i x_{ij} y_i + \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0$$

olur. Sonra çözüm $\gamma > 1$ için MNR, $\gamma = 1$ için Atış metodu ile hesaplanır. $\gamma > 1$ için (2.6) varyans formülü ile standart hatalar hesaplanır. Bootstrap metodu (Efron ve Tibshirani, 1993) $\gamma = 1$ için standart hataları hesaplamak için kullanılır.

Tablo 2.1’de farklı büzölme fonksiyonları için tahmin ve standart hatalar verilmiştir. $\gamma = 1$ LASSO için, parametre tahmini ve standart hata sabit γ için artan λ ile monoton büzölmeyi gösterir. Ancak $\gamma = 1$ için $\hat{\beta}_3$ ’nin standart hatası, λ ile monoton azalan trend göstermez. $\lambda = 0$ ’da 0.163, $\lambda = 10$ ’da 0.157, $\lambda = 100$ ’de 0.354’e eşittir. $\lambda > 0$ için LASSO standart hataları nedeniyle, yarı parametrik bootstrap metodu ile hesaplanır.

Tablo 2.1. Ortonormal X için Bridge tahmin ediciler ve standart hatalar (Fu, 1998)

$\lambda = 0$		$\lambda = 10$			
		$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
$\beta_1 = 1$	0.917(0.163)	0.792(0.157)	0.754(0.143)	0.734(0.126)	0.722(0.103)
$\beta_2 = -2$	-2.030(0.163)	-1.905(0.158)	-1.780(0.148)	-1.624(0.126)	-1.348(0.079)
$\beta_3 = 5$	5.071(0.163)	4.946(0.157)	4.666(0.152)	4.057(0.126)	2.578(0.054)
$\lambda = 0$		$\lambda = 100$			
		$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
$\beta_1 = 1$	0.917(0.163)	0.000(0.002)	0.162(0.048)	0.262(0.045)	0.379(0.041)
$\beta_2 = -2$	-2.030(0.163)	-0.780(0.059)	-0.590(0.071)	-0.580(0.045)	-0.614(0.028)
$\beta_3 = 5$	5.071(0.163)	3.821(0.354)	2.256(0.097)	1.449(0.045)	1.037(0.018)

Ortonormal olmayan X matrisi ile örnek

40 gözlemlili, benzer

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

modelini düşünelim. X regresyon matrisi ortonormal değil ve korelasyon matrisi

$$\text{Corr}(X) = \begin{bmatrix} 1.000 & 0.707 & 0.579 \\ 0.707 & 1.000 & 0.819 \\ 0.579 & 0.819 & 1.000 \end{bmatrix}$$

dir. $\sum_i x_{ij} = 0$, $\sum_i x_{ij}^2 = 1$ ile X 'in x_j kolonları standartlaştırılsın. Kolaylık için $\beta_0 = 0$ ve $\sigma^2 = 1$ diyelim. Y yanıtının 40 gözlemi, $\beta_1 = 2$, $\beta_2 = 3$, $\beta_3 = -1$ gerçek değerleri ile üretilsin. Sabit terimde büzülme etkisi olmadığından, sabit $\sum_i y_i = 0$ merkezileştirme ile silinir. $\lambda > 0$ ve $\gamma \geq 1$ için (2.3)'ün her bir denklemi $j = 1, 2, \dots, p$ için

$$-2 \sum_i x_{ij} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) + \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0$$

olur. Sonra çözüm $\gamma > 1$ için M-N-R, $\gamma = 1$ için Atış metodu ile hesaplanır. $\gamma = 1$ için standart hatalar bootstrapping metodu ile hesaplanır.

Tablo 2.2'de farklı büzülme fonksiyonları için tahmin ve standart hatalar verilmiştir. $\gamma = 1$ LASSO için standart hataları 10000 bootstrap örneği ile hesaplanır (Fu, 1998).

Tablo 2.2. Ortonormal olmayan X için Bridge tahmin ediciler ve standart hatalar (Fu, 1998)

	$\lambda = 0$	$\lambda = 10$			
		$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
$\beta_1 = 2$	0.947(0.231)	1.874(0.239)	1.819(0.187)	1.715(0.137)	1.424(0.069)
$\beta_2 = 3$	2.877(0.329)	2.239(0.249)	2.027(0.200)	1.769(0.134)	1.371(0.059)
$\beta_3 = -1$	-0.902(0.285)	-0.212(0.086)	-0.079(0.130)	0.112(0.136)	0.530(0.106)
	$\lambda = 0$	$\lambda = 100$			
		$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
$\beta_1 = 2$	1.947(0.231)	1.215(0.344)	0.891(0.072)	0.765(0.040)	0.718(0.022)
$\beta_2 = 3$	2.877(0.329)	1.406(0.360)	0.877(0.064)	0.748(0.036)	0.706(0.021)
$\beta_3 = -1$	-0.902(0.285)	0.000(0.070)	0.307(0.047)	0.436(0.038)	0.531(0.028)

2.6. Ortonormal Matris için Bridge Regresyon

Bir önceki bölümde ortonormal matris X için Bridge regresyon örneğinde büzülme etkisi verilmiştir. Bu bölümde ortonormal matris X için Bridge regresyon teorik olarak çalışılacak ve farklı γ değerleri için farklı büzülme etkileri gösterilecektir.

$X = (x_{ij})$ ortonormal matrisi için,

$$\sum_i x_{ij}x_{il} = \begin{cases} 1, & j = l \\ 0, & j \neq l \end{cases}$$

olsun. $j = 1, 2, \dots, p$ için p bağımsız denklem

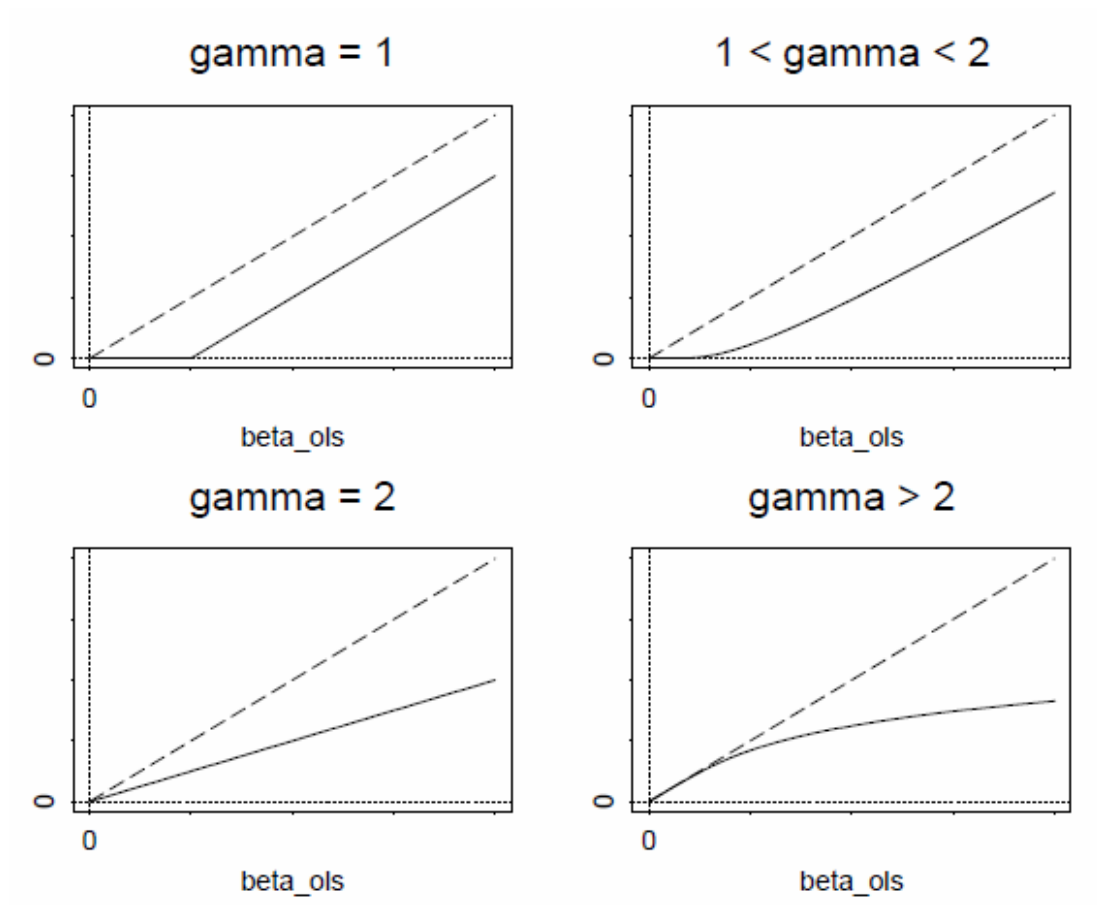
$$2\beta_j - \sum_i x_{ij}y_i + \lambda\gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0 \quad (2.7)$$

(2.3) probleminden görülebilir. Çözüm, $\gamma > 1$ için MNR, $\gamma = 1$ için Atış metodu ile hesaplanır. γ 'nın farklı değerlerinin büzülme etkisi için, Bridge ve OLS tahmin edici karşılaştırılır. Hiçbir koşul olmadan, kolaylık için x_{ij} ve β_j 'nin alt indisi j 'yi kaldıralım. (2.7) denklemi

$$\beta = \sum_i x_i y_i - \frac{\lambda\gamma}{2} |\beta|^{\gamma-1} \text{sign}(\beta)$$

olarak yazılabilir. Sağ taraftaki ilk terim OLS tahmin ediciye eşittir. İkinci terim büzülmeye neden olandır. Böylece büzülme etkisini yansıtır.

$$\hat{\beta}_{brg} = \hat{\beta}_{OLS} - \frac{\lambda\gamma}{2} |\hat{\beta}_{brg}|^{\gamma-1} \text{sign}(\hat{\beta}_{brg})$$



Şekil 2.3. Sabit $\lambda > 0$ için Bridge regresyonların büzülme etkisi. Tam çizgi Bridge tahmin edici, kesikli çizgi OLS tahmin edicisi (Fu, 1998)

Bridge regresyonun büzülme etkisini göstermek için, Fu (1998) makalesinden alınan Şekil 2.3'te; Bridge tahmin edici $\hat{\beta}_{brg}$ 'nin tam değerinin grafiği ve OLS tahmin edici karşılaştırılmış, tam değeri grafiklendirilmiştir. γ 'nın küçük değerleri, küçük parametreleri sıfıra büzmeye meyilli iken, büyük γ değerli Bridge regresyon, küçük parametreleri tutma eğilimindedir. Bu yüzden doğru model küçük fakat sıfır olmayan parametreler içeriyorsa, büyük γ değeri ile Bridge iyi performans gösteriyorken, LASSO kötü performans gösterir. Doğru model sıfır parametrelerini içeriyorsa LASSO iyi performans gösterirken, büyük γ değerli Bridge kötü performans verir.

2.7. Bayesçi Prior Olarak Bridge Ceza

Bu bölümde $\beta = (\beta_1, \dots, \beta_p)^T$ parametresinin Bayes prior dağılımı olarak Bridge ceza fonksiyonu $\sum |\beta_j|^\gamma$ çalışılacaktır.

Bayes bakış açısından, Bridge ceza fonksiyon $\min_{\beta} \left(RSS + \lambda \sum |\beta_j|^\gamma \right)$, $(\beta/Y) \sim C \exp \left\{ -\frac{1}{2} \left(RSS + \lambda \sum |\beta_j|^\gamma \right) \right\}$ 'nin log-posterior dağılımını maksimize etmek gibi kabul edilir (C bir sabit). Böylece Bridge ceza $\sum |\beta_j|^\gamma$, $\beta = (\beta_1, \dots, \beta_p)^T$ parametresinin prior dağılımı $C_0 \exp \left\{ -\frac{1}{2} \lambda \sum |\beta_j|^\gamma \right\}$ 'nin logaritması gibi kabul edilir. Buradaki $C_0 > 0$ olan normalleştirilmiş bir sabittir. $\beta_1, \beta_2, \dots, \beta_p$ parametreleri karşılıklı bağımsız ve özdeş dağılmışlardır. Alt indis j'yi çıkarır ve sadece β 'nin prioru $C_0 \exp \left(-\lambda |\beta|^\gamma / 2 \right)$ ile çalışırız.

Basit cebir ile

$$\int_{-\infty}^{+\infty} \exp \left(-\frac{\lambda}{2} |\theta|^\gamma \right) d\theta = \frac{1}{\gamma} 2^{1+1/\gamma} \lambda^{-1/\gamma} \Gamma \left(\frac{1}{\gamma} \right)$$

$\Gamma(\cdot)$, gama fonksiyonudur.

Böylece β 'nın olasılık yoğunluk fonksiyonu

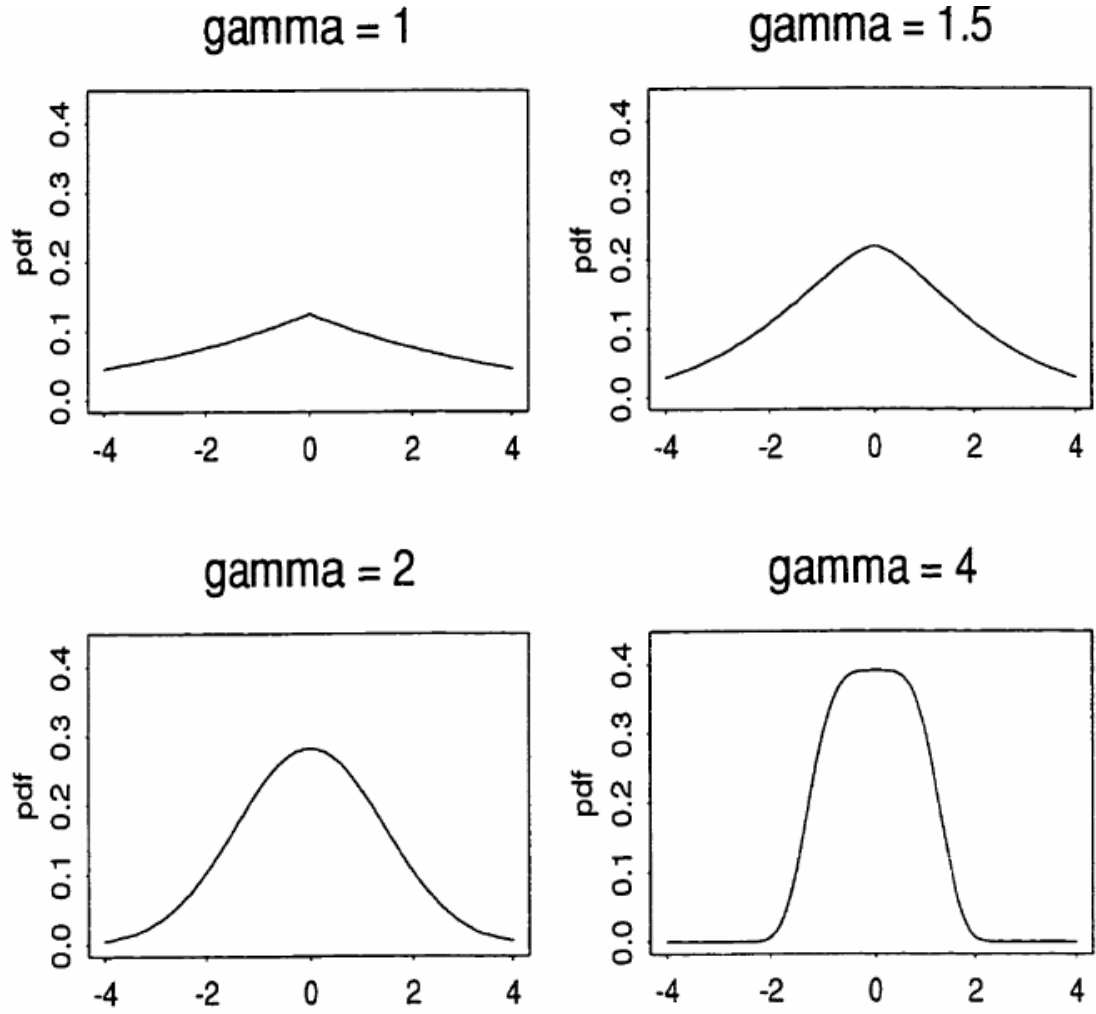
$$\Pi_{\lambda,\gamma}(\beta) = \frac{\gamma 2^{-(1+1/\gamma)} \lambda^{1/\gamma}}{\Gamma(1/\gamma)} \text{Exp}\left(-\frac{1}{2} \left| \frac{\beta}{\lambda^{-1/\gamma}} \right|^\gamma\right)$$

dir. Buradaki $\lambda^{-1/\gamma}$, yoğunluğun pencere boyutu kontrolleridir. Özellikle $\gamma = 2$ olduğunda β , Gauss dağılımına sahipse, (β/Y) 'nin posterior dağılımı da Gauss dağılımıdır. Bu, Ridge regresyon için Ridge tahmin edicinin özel bir durumudur.

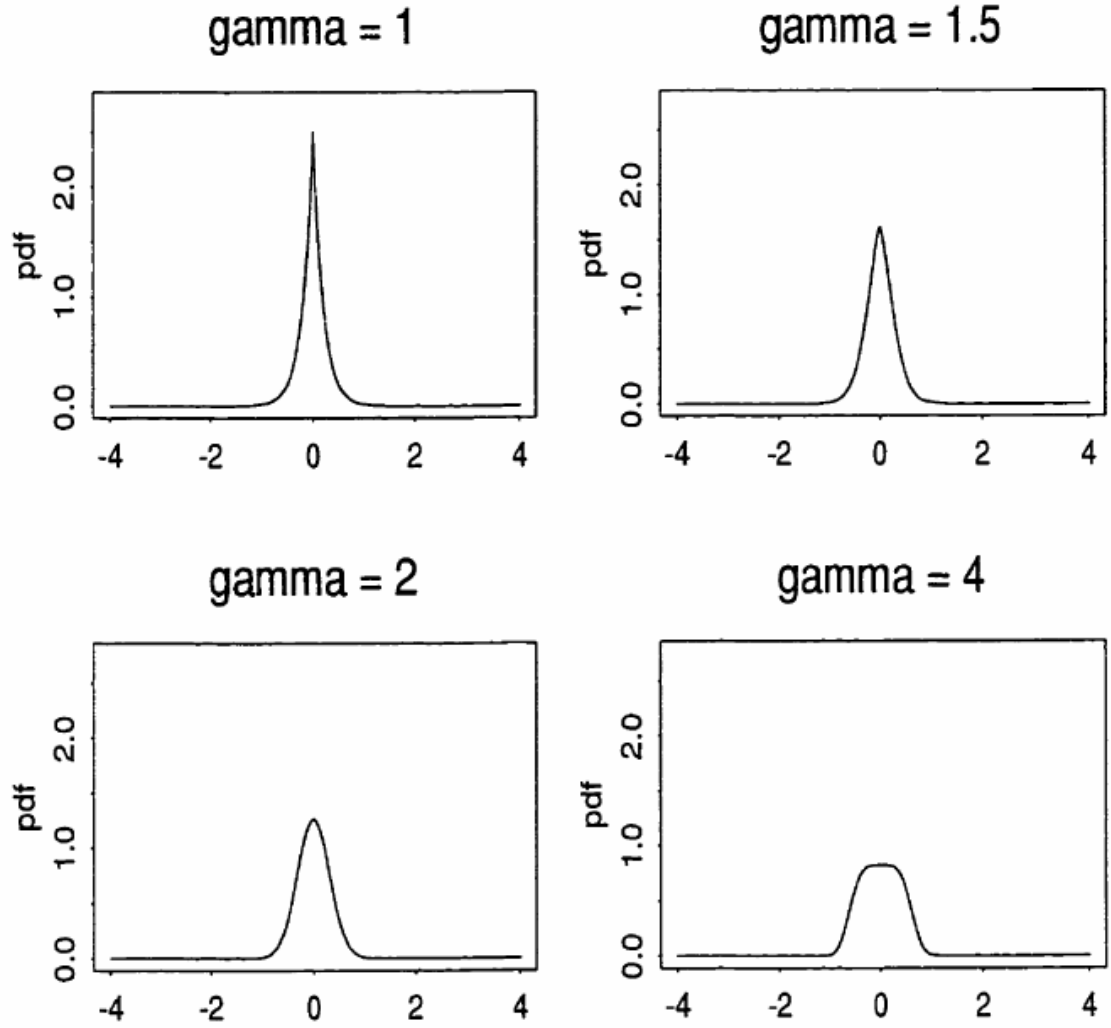
λ ve γ 'nın farklı değerlerinin ceza fonksiyonlarını karşılaştırmak için, Şekil 2.4 ve 2.5 'de gösterildiği gibi yoğunluk fonksiyonu $\Pi_{\lambda,\gamma}(\beta)$ grafiği çizilir. γ 'nın büyük değerleri, merkezde $\beta = 0$ çevresinde toplanır ve böylece yoğunluk daha az yayılırken Şekil 2.4 'teki gibi γ 'nın küçük değerleri daha çok kuyruk kısmında toplanır ve böylece yoğunluk geniş pencere boyutundadır yani yoğunluk daha fazla yayılır. Şekil 2.5 'teki gibi λ 'nın geniş değerleri için $\lambda^{-1/\gamma}$ olduğundan, değişmeyen pencere boyutu 1'den küçük olur ve γ artarken çok hızlı bir şekilde 1'e yaklaşır. γ 'nın küçük değerleri $\beta = 0$ 'da en üst nokta ile $\beta = 0$ 'a çok yakın birçok küme koyarken, γ 'nın geniş değerlerinde kümeler düz bir şekilde dağılma eğilimindedirler. $\gamma = 2$ iken yoğunluk Gauss dağılımıdır.

λ 'nın küçük değerleri için; büyük regresyon parametrelili küçük γ değerli Bridge ceza ya da sıfır olmayan fakat küçük parametrelili büyük γ değerli Bridge ceza modeli tercih edilir.

λ 'nın büyük değerleri için; sıfır içeren regresyon parametrelili küçük γ değerli Bridge ceza ya da sıfır olmayan fakat küçük parametrelili büyük γ değerli Bridge ceza modeli tercih edilir. Bu sonuç ortonormal regresyon matrisi için söylediğimiz sonuç ile aynı anlamdadır.



Şekil 2.4. $\lambda = 0.5$ ile Bayeşçi prior olarak Bridge ceza (Fu, 1998)



Şekil 2.5. $\lambda = 10$ ile Bayeşçi prior olarak Bridge ceza (Fu, 1998)

2.8. Ayar Parametresi λ ve t Arasındaki İlişki

Bölüm 2.1 'de problem 1 ve 2'nin denkliğinden bahsetmiştik. Yani $\lambda \geq 0$ verilmişken bir $t \geq 0$ varken (2.1) ve (2.2)'nin aynı sonuçları paylaştığı söylenmişti. Bu bölümde ortonormal X matrisinin özel durumu için λ ve t arasındaki ilişkiye bakacağız.

$\gamma \geq 1$ verilmişken (2.1)'in sınırlı alanı Şekil 1.1 'deki gibi kapalı bir formdadır. Böylece Bridge tahmin edici sınırından elde edilir. Bu, sabit $\lambda \geq 0$ için $t(\lambda) = \sum |\beta_j(\lambda, \gamma)|^\gamma$ anlamına gelir.

Ortonormal X matrisi ile, (2.3)'ün p bağımsız denklemi

$$2\beta_j - 2\sum_i x_i y_i + \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0 \quad (2.8)$$

$\sum_i x_i y_i = \hat{\beta}_{OLS,j}$ olduğundan OLS tahmin edicinin j. koordinatıdır. Bridge tahmin edici $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$,

$$\hat{\beta}_j - \hat{\beta}_{OLS,j} + \frac{\lambda \gamma}{2} |\hat{\beta}_j|^{\gamma-1} \text{sign}(\hat{\beta}_j) = 0$$

eşitliğini sağlar. $c_j = \hat{\beta}_{OLS,j}$ ve Bridge tahmin edicinin OLS tahminine oranı $s_j = \hat{\beta}_j / c_j$ olsun.

$$\frac{\lambda \gamma}{2} \sum |\hat{\beta}_j|^\gamma = \sum c_j \hat{\beta}_j - \sum \hat{\beta}_j^2$$

Böylece

$$t(\lambda) = \frac{2}{\lambda \gamma} \sum c_j^2 s_j (1 - s_j).$$

Buradaki s_j , (2.8) denkleminde

$$s_j^2 - s_j + \frac{\lambda\gamma}{2} s_j^\gamma c_j^{\gamma-2} = 0$$

denkleminin çözümü ile hesaplanır. Böylece $t(\lambda)$, yukarıdaki formülde s_j 'nin yerine koyulmasıyla hesaplanır. Özel durum olarak $c_j = c$ ve $s_j = s$ j 'den bağımsız sabitler olmak üzere

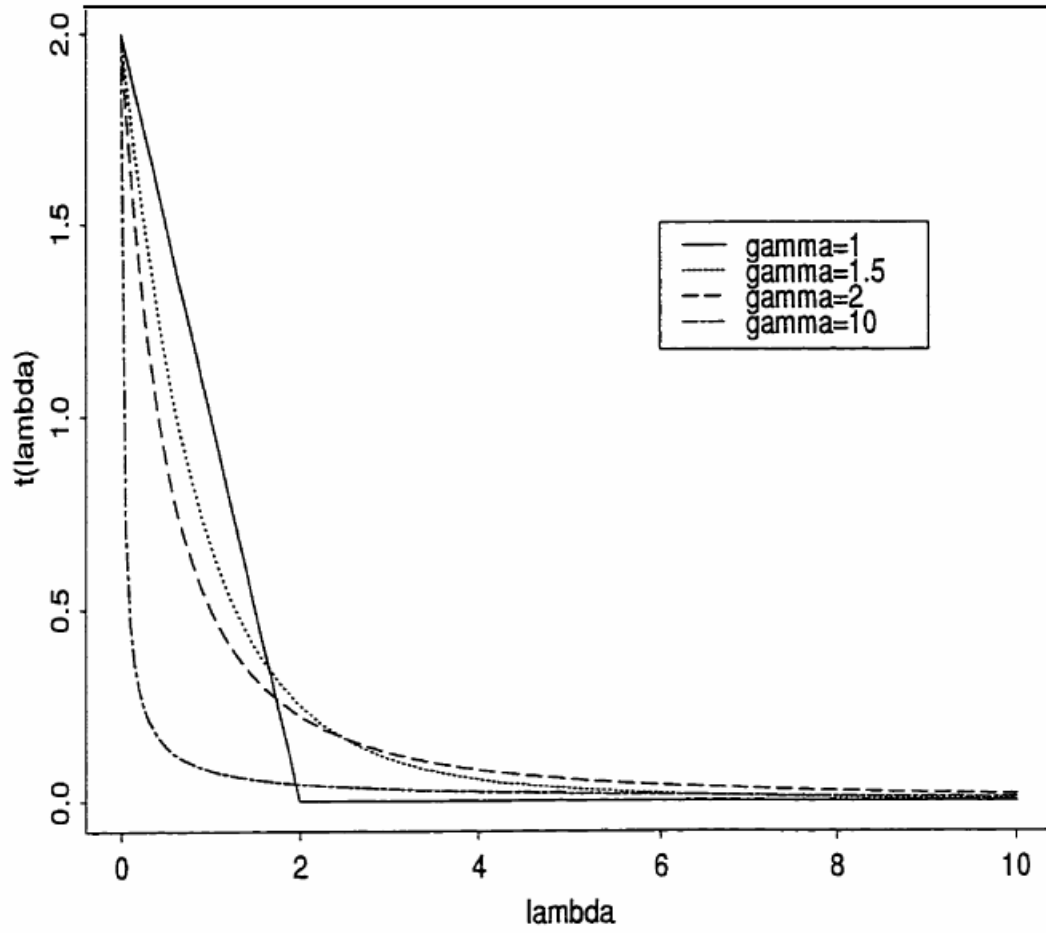
$$t(\lambda) = \frac{2p}{\lambda\gamma} c^2 s(1-s).$$

Şekil 2.6; $\gamma = 1, 1.5, 2, 10$ farklı sabitleri için $p = 2$ ile $c_j = 1$ özel durumu için, hesaplanmış $t(\lambda)$ fonksiyonunu göstermektedir. Bu, t ve λ arasındaki birebir uyumu gösterir. Bu durum için, LASSO $\hat{\beta}_j = 0$ için λ 'nın başlangıç değeri $\lambda_0 = 2$ 'dir. Herhangi $\lambda \geq \lambda_0$, $\hat{\beta}_j(\lambda) = 0$ 'ı verir. Şekil 2.6'dan sabit $\gamma \geq 1$ için $t(\lambda)$ 'nın monoton azalan bir fonksiyon olduğu görülür.

$\gamma > 1$ için λ tüm $\hat{\beta}_j = 0$ 'ları, $\gamma = 1$ için $\lambda \geq \lambda_0 = 2$ tüm $\hat{\beta}_j = 0$ 'ları bürer.

Bu nedenle $t(\lambda) = 0$ 'dır.

t (lambda) fonksiyonunun grafiği

Şekil 2.6. $p = 2$ ve $c_j = 1$ özel durumu için $t(\lambda) - \lambda$ grafiği (Fu, 1998)

3. CEZALI SKOR DENKLEMLERİ

3.1. Giriş

Bölüm 2’de Bridge tahmin edicilerin bazı teorik sonuçları elde edilmiştir ve (2.3) ile Bridge tahmin edicileri için çözüme genel yaklaşım, yani $\gamma > 1$ için MNR metodu ve $\gamma = 1$ için Atış metodu geliştirilmiştir. Bu bölümde teorik olarak cezalı skor denklemleri tanıtıldı ve böylece cezalandırma kavramı genelleştirildi. Cezalı skor denklemleri için algoritmalar IRLS ile MNR ve Atış metodu ile verilmiştir. İlk olarak GLM, likelihood fonksiyonlar ve yarı likelihood incelenmiştir.

3.2. Genelleştirilmiş Lineer Modeller ve Likelihood

Birçok uygulamalı bilimde, yanıt sürekli olmayabilir. Yanıt; ikili veri, sayım verisi ve derecelendirilmiş veri (acının derecesi: yok, hafif, orta, şiddetli) olabilir ve bu durumda

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

gibi bir lineer model uygun olmayabilir. Nelder ve Wedderburn (1979), yanıt değişkenin birçok sınıfı için lineer regresyon modellerinin doğal yayılımı olan GLM’i tanıtmıştır. GLM’de hatalar için normal dağılım varsayımı gerektirmez. GLM, açıklayıcı değişkenlerin doğrusal yapısını cevap değişkenin beklenen değerine bağlayan bir bağıntı (link) fonksiyonu kullanır (McCullagh ve Nelder, 1989; Dobson, 1990).

GLM yönteminde verilerin orijinal dağılımı üstel (exponential) formda yazılır, parametre tahminleri de maksimum likelihood (ML) veya yarı likelihood (quasi-likelihood) yöntemleriyle elde edilmektedir. Bazı durumlarda gözlem değerleri (y_i) normal dağılımlı olmayabilir. GLM, standart doğrusal modellerle verilerin orijinal dağılımını esas alarak ML yöntemi ile parametre tahmini yapar. GLM’de, gözlem değerlerinin olasılık yoğunluk fonksiyonu

$$f(y) = l(\theta, \Phi, y) = \exp\{(y\theta - b(\theta))/a(\Phi) + c(y, \Phi)\}$$

şeklinde olmakta ve a , b ve c dağılımın ne olduğunu belirlemektedir. Burada θ doğal (natural) ve Φ ölçek (scale) parametresi olmaktadır. Ayrıca y ve μ , $y > 0$ ve $\mu > 0$ şeklinde sınırlandırılır. Örneğin Poisson dağılımını ele alalım. Bu dağılıma ait herhangi bir gözlem değerinin olasılık fonksiyonu,

$$P(Y = y / \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}$$

biçiminde verilmektedir ve bu olasılık fonksiyonunun log-likelihood fonksiyonu, GLM kullanılarak,

$$l(\lambda; y) = y \log(\lambda) - \lambda - \log(y!)$$

biçiminde verilir. Burada, terimleri karşılaştırdığımızda; $\log(\lambda) = \theta$, $\lambda = b(\theta)$, $\log(y!) = c(y, \Phi)$ ve $a(\Phi) = 1$ olmaktadır (Dobson, 1990; Littell ve ark., 1996).

Bir GLM'in 3 bileşeni vardır.

1) Rasgele Bileşen: $Y = (Y_1, Y_2, \dots, Y_n)^T$ 'nin bileşeni karşılıklı bağımsız ve $E(Y) = \mu$ ortalama, $Var(\mu)$ varyanslı üstel ailenin özdeş dağılımına sahiptirler.

2) Sistemik Bileşen: $x_1, x_2, \dots, x_p$ değişkenleri,

$$\eta = \sum_{j=1}^p x_j \beta_j$$

lineer prediktörünü gösterir.

3) Rasgele ve Sistemik bileşenler arasındaki bağıntı:

$$g(\mu_i) = \eta_i$$

olur. Burada $g(\cdot)$ monoton diferansiyellenebilir fonksiyondur ve link fonksiyon olarak adlandırılır. Böylece GLM;

$$g(E(Y)) = x_1\beta_1 + \dots + x_p\beta_p$$

olarak yazılabilir.

Yanıtların ve kanonik link fonksiyonlarının en çok kullanılanları; $g(\mu) = \mu$ birim

linkli Gauss yanıt, $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ logit linkli Binom yanıt ve $g(\mu) = \log(\mu)$

log linkli Poisson sayılarıdır. $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ parametrelili sonuç, likelihood fonksiyonuna dayalı

$$L(\beta) = L(\theta(\beta), \Phi; y) = \prod_i f_{Y_i}(y_i; \theta(\beta), \Phi)$$

fonksiyonudur ve maksimum likelihood tahmin edicisi (MLE) $\hat{\beta}_{mle}$,

$$\hat{\beta}_{mle} = \arg \max_{\beta} L(\beta)$$

olarak tanımlanmıştır. MLE tahmin edicisi $\hat{\beta}_{mle}$, aşağıdaki Newton-Raphson metodu, Fisher metodu veya IRLS metodu ile hesaplanır.

Büyük örneklem testi ile, MLE $\hat{\beta}_{mle}$ düzen koşulları altında asimptotik olarak tutarlıdır.

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, [I(\beta)]^{-1})$$

Burada $I(\beta)$;

$$I(\beta) = -E\left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T}\right)$$

ile tanımlı Fisher bilgi matrisi ve $l(\beta) = \log(L(\beta))$ log-likelihood fonksiyonudur.

MLE $\hat{\beta}_{mle}$ çözümü için $l(\beta)$ log-likelihood fonksiyonunun β_j 'ye göre kısmi türevini alırız. $\hat{\beta}_{mle}$;

$$\begin{cases} \partial l / \partial \beta_1 = 0 \\ \dots \\ \partial l / \partial \beta_p = 0 \end{cases} \quad (3.1)$$

denklemlerini sağlamalıdır. $\partial l / \partial \beta_j$ $l(\beta)$ likelihoodun skor denklemleri olarak adlandırılır.

Newton-Raphson Metodu

$\partial l(\beta) / \partial \beta$ skor denklemlerinin Taylor serisini alır ve karesel terimden sonrasını önemsemezsek;

$$0 = \frac{\partial l(\beta)}{\partial \beta} \Big|_{\hat{\beta}_{mle}} = \frac{\partial l}{\partial \beta} \Big|_{\beta} + \frac{\partial^2 l}{\partial \beta \partial \beta^T} \Big|_{\beta} (\hat{\beta}_{mle} - \beta)$$

ve

$$\hat{\beta}_{mle} = \beta + \left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial l}{\partial \beta} \quad (3.2)$$

Böylece;

$$\hat{\beta}_{m+1} = \hat{\beta}_m + \left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right)^{-1} \Big|_{\hat{\beta}_m} \frac{\partial l}{\partial \beta} \Big|_{\hat{\beta}_m} \quad (3.3)$$

iteratif çözümü ile $\hat{\beta}_{mle}$ hesaplanır. İterasyon, $\hat{\beta}$ tahmini ya da sapma

$$D(y; \hat{\beta}_m) = 2l(y; \mu_{\max}) - 2l(y; \mu(\hat{\beta}_m))$$

yakınsayınca kadar devam ettirilir. $\mu_{\max}$; doymuş modelin yanıtının ortalamasıdır ve genellikle y 'ye eşittir.

Fisher Skor Metodu

Newton-Raphson metodunda (3.2) denklemindeki gözlenen bilgi matrisi $\left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right)$ yerine, beklenen bilgi matrisi $E_{\theta} \left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right)$ koyalım. Buradaki θ ; parametrenin gerçek değeridir. MLE $\hat{\beta}_{mle}$ için çözüm, aşağıdaki Fisher skor metodu ile elde edilir.

$$\hat{\beta}_{mle} = \beta + \left[E_{\theta} \left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right) \right]^{-1} \frac{\partial l}{\partial \beta} \quad (3.4)$$

$E_{\theta} \left(\frac{-\partial^2 l}{\partial \beta \partial \beta^T} \right)$; θ yardımıyla β 'ya bağlıdır. Bu, hesaplamaları kolaylaştırır. Eğer Y , kanonik link fonksiyonlu üstel ailenin dağılımının bir sonucu ise, gözlenen ve beklenen Fisher bilgi matrisleri özdeştir. Böylelikle Fisher skor metodu Newton-Raphson metodu ile benzerdir (McCullagh ve Nelder, 1989; Hastie ve Tibshirani, 1990).

İteratif Yeniden Ağırlıklandırılmış En Küçük Kareler (IRLS) Metodu

Green (1984), link fonksiyonun lineer yayılmasının alınmasıyla MLE hesaplamak için aşağıdaki IRLS metodunu önermiştir.

$$\begin{aligned} g(y) &= g(\mu) + (y - \mu)g'(\mu) \\ &= \eta + (y - \mu)\partial\eta/\partial\mu \end{aligned}$$

η ; lineer prediktör, $V(\mu)$; μ ortalamalı Y 'nin varyansı olmak üzere, düzeltilmiş bağımlı değişken $z = \mu + (y - \mu)/V(\mu)$ kanonik linkler için tanımlanmıştır. MLE

tahmin edici, $V(\mu)$ ağırlıklı X matrisinde z 'nin regress edilmesiyle hesaplanabilir. IRLS yöntemi aşağıdaki gibi özetlenebilir.

IRLS Yöntemi

1. $\hat{\beta}_0$ başlangıç tahmini ile başla.
2. $\eta = X\hat{\beta}$ ve $V(\mu) = \text{diag}(V_1(\mu_1), \dots, V_n(\mu_n))$ ağırlıklarını hesapla.
3. Düzeltilmiş bağımlı değişken $z = \eta + [V(\mu)]^{-1}(y - \mu)$ tanımla.
4. Yeni tahmin $\hat{\beta}$ oluşturmak için $V(\mu)$ ağırlıklı X matrisinde z 'yi regress et.
5. Yakınsama elde edilinceye kadar 2-4 adımlarını tekrarla.

IRLS yönteminin Newton-Raphson veya Fisher Skor metoduna göre avantajı; Ağırlıklandırılmış EKK yöntemi kullanmasıdır. Ağırlıklandırılmış EKK standart yöntemdir ve birçok istatistiksel yazılımlarda tanımlanması kolaydır.

3.3. Yarı-Likelihood ve Yarı-Skor Fonksiyonları

Son bölümde kısaca GLM ve üstel ailenin dağılımlarını gözden geçirmiştik. Olasılık fonksiyonu açıkça belirtildiğinde, likelihood fonksiyonu kurulabilir ve MLE kolayca hesaplanabilir. Ancak bazı durumlarda tam olasılık dağılımını ve böylece ortak likelihood fonksiyonunu açıkça belirtmek gerekli değildir, veya ortak likelihood fonksiyonunu açıkça belirtmek mümkün değildir.

Wedderburn (1974) olasılık dağılımında GLM'e genişletilen yarı-likelihoodu önermiştir. Rasgele değişkenin $V(\mu)\sigma^2$ varyansına ihtiyacı olan yarı-likelihood, ortalamanın bilinen fonksiyonudur. Üstel aileden açıkça belirtilmeksizin dağılımıdır. Tek boyutlunun yarı-skoru

$$U(\mu, y) = \frac{y - \mu}{\sigma^2 V(\mu)} \quad (3.5)$$

olarak tanımlanmıştır.

$U(\mu, y)$, likelihood fonksiyonun skor fonksiyonlarının 3 temel özelliğini sağlar.

$$E(U(\mu, y)) = 0$$

$$Var(U(\mu, y)) = \frac{1}{\sigma^2 V(\mu)}$$

$$-E\left(\frac{\partial U}{\partial \mu}\right) = \frac{1}{\sigma^2 V(\mu)}$$

Böylelikle (3.5) denkleminin integrali

$$Q(\mu, y) = \sum_i \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 V_i(t)} dt \quad (3.6)$$

varsa, log-likelihood fonksiyonunun benzer özelliklerine sahiptir.

Yarı-likelihood için aşağıdaki 2 duruma bakacağız.

1. Bağımsız Gözlemler: Gözlemler bağımsız olduğundan, varyans-kovaryans matrisi diyagonaldir.

$$V(\mu) = \text{diag}(V_1(\mu_1), \dots, V_n(\mu_n))$$

Buradaki $V_1, V_2, \dots, V_n$ fonksiyonları özdeştir. (3.5) ile gösterilen yarı-skor, (3.6)'daki yarı-likelihood fonksiyonudur. Yarı-likelihood fonksiyonu $Q(\mu, y)$, GLM'deki log-likelihood fonksiyonu olarak rol oynar. Yarı-likelihood tahmin ediciye dayanan sonuç,

$$\begin{cases} U_1(\mu, y) = 0 \\ \dots \\ U_p(\mu, y) = 0 \end{cases} \quad (3.7)$$

yarı-skor denklemlerini sağlar.

GLM'in MLE'sine benzer yarı-likelihoodun tahmin edicisi Fisher Skor metodu,

$$\hat{\beta}_{m+1} = \hat{\beta}_m + \left[E_{\theta} \left(-\frac{\partial U}{\partial \beta} \right) \right]^{-1} U$$

yardımıyla hesaplanabilir. Ayrıca bu tahmin edici asimptotik olarak da tutarlıdır. Yani, düzen koşulları altında

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N\left(0, [I(\hat{\beta})]^{-1}\right)$$

dir.

2. Bağımlı Gözlemler: Gözlemler bağımlı olduğundan, varyans-kovaryans matrisi $V(\mu)$ diyagonal değildir. Genelde yarı-skor $U = (U_1, \dots, U_p)^T$,

$$\frac{\partial U_r(\beta)}{\partial \beta_s} \neq \frac{\partial U_s(\beta)}{\partial \beta_r}$$

eşitsizliğine sahiptir. Bu, yarı-skor $U(\mu, y)$ ile tanımlı vektörün path bağımlı olduğu anlamına gelir. Böylece kısmi türevi varsa yarı-skor olan $Q(\mu, y)$ skaler fonksiyonu yoktur. Böylelikle (3.6)'daki $Q(\mu, y)$ integrali path bağımlıdır ve iyi tanımlanmamıştır. Sonuç, $Q(\mu, y)$ 'ye dayalı olamaz. Daha önce bahsedildiği gibi log-likelihood fonksiyonlarının 3 ana özelliğini sağlayan $U(\mu, y)$ yarı-skor fonksiyonu tercih edilir. Asimptotik yakınsama, oldukça karmaşık durumlar altında da sağlanır (McCullagh 1991).

$U(\mu, y)$ yarı-skor fonksiyonunun kısmi türevinin beklenen değeri simetrik olduğundan ve kısmi türevinin olmadığından, McCullagh (1991), simetrik kısmi türevleri ile bir ana terim ve asimetrik kısmi türev ile bir küçük "gürültü" terimi olarak U 'nun 2 terime ayrışımının olasılığına dikkat çekmiştir.

Ayrışım, bilgi kaybı olmadan ilk terimin yarı-likelihoodu aracılığıyla, yarı-skor $U(\mu, y)$ 'nin çalışmasına izin verir. Li ve McCullagh (1994) potansiyel

fonksiyonlar ve tutucu tahmin fonksiyonları çalışmıştır. Tahmin fonksiyonları simetrik kısmi türeve sahip olan tutucu tahmin fonksiyonlarının alt uzayında tahmin fonksiyonları tasarladılar, böylece bu tahmin fonksiyonları bir yarı-likelihood fonksiyonuna sahiptir. Yarı-likelihood, tahmin fonksiyonunun potansiyel fonksiyonu gibi adlandırılır.

Denklemleri parametre tahmin edicilerini sağlayan tahmin fonksiyonları, fonksiyonların geniş kapsamlı bir sınıfıdır. Yarı-skor fonksiyonlar, tahmin fonksiyonlarının özel bir sınıfıdır. y 'de doğrusaldır ve asimptotik olarak tutarlı tahmin edici sağlarlar. Potansiyel fonksiyonlar, sıradan log-likelihood fonksiyonlar gibi asimptotik olarak benzer özelliklere sahiplerdir. Böylece, yarı-skor denklemlerinin mümkün çoklu çözümlerinden, istenilen hesaplamaya yardımcı olur.

3.4. Cezalı Skor Denklemleri

Bir önceki bölümde GLM, likelihood fonksiyonlar, skor fonksiyonlar ve yarı-likelihoodu inceledik.

Uyarı: (2.3) problemi ve çözümü, ortak likelihood fonksiyonlarından bağımsızdır. Cezalandırma kavramı ve tahmin edicisi, ortak likelihood fonksiyonlarından bağımsız olarak genelleştirilir (Fu, 1998).

Cezalı skor denklemlerini

$$\begin{cases} S_1(\beta, X, y) + \lambda \gamma |\beta_1|^{\gamma-1} \text{sign}(\beta_1) = 0 \\ \dots \\ S_p(\beta, X, y) + \lambda \gamma |\beta_p|^{\gamma-1} \text{sign}(\beta_p) = 0 \end{cases} \quad (3.8)$$

kabul edelim.

Tanım 1: (Cezalı Skor Denklemleri)

Jakobian koşulu $\frac{\partial S}{\partial \beta}$ pozitif yarı tanımlı olan S fonksiyonlu (3.8) eşitliği, Bridge ceza $\sum |\beta_j|^\gamma$ ile cezalı skor denklemleri olarak adlandırılır (Fu, 1998).

Tanım 2: (Bridge Tahmin Edici)

$\lambda > 0$ ve $\gamma > 1$ olsun. Bridge tahmin edicisi $\hat{\beta}(\lambda, \gamma)$ olarak tanımlansın. (3.8) eşitliğinin tek çözümü, $\gamma \rightarrow 1+$ $\hat{\beta}(\lambda, \gamma)$ 'nin limiti LASSO tahmin edici $\hat{\beta}(\lambda, 1+)$ olarak tanımlanır (Fu, 1998).

Uyarılar

1. Cezalı skor denklemleri kavramı genelde $\sum g(\beta_j)$ 'nin bir cezası için sürdürülebilir. Buradaki g, düzgün konveks fonksiyondur.
2. Bridge (LASSO) tahmin edicisi, ortak likelihood fonksiyonlarından bağımsızdır. Ortak likelihood fonksiyonlarının bulunmadığı durumlarda uygulanabilir.

Sapmayı minimize eden cezalı skor denklemlerine yaklaşım, cezalandırmaya klasik yaklaşım ile karşılaştırılır. Yani $-2\text{Log}(lik)$, + ceza fonksiyonu gibi. Böyle bir genelleştirme, regresyon problemlerinde ortak likelihood fonksiyonlarının yokluğunda karşılaşılabilecek zorlukları önlemek için çok önemlidir, yüksek ilişkili açıklayıcı değişkenler nedeniyle cezalandırma istenilir. Çok önemli bir uygulama da, genelde ortak likelihood fonksiyonu olmadığında genelleştirilmiş tahmin denklemlerine bu yöntem uygulanır. Bridge (LASSO) tahmin edici için cezalı GEE çözümü ile açıklayıcı değişkenler arasında ilişki verildiğinde daha iyi öntahminler elde edileceği, algoritmalar için Bölüm 4'te, simülasyon sonuçları için Bölüm 6'da görülecektir.

3.5. Cezalı Skor Denklemleri için Algoritmalar

Bölüm 3.4'te cezalı skor denklemlerine teorik olarak giriş yapıldı. Bölüm 2.3'te Bridge çözümü için MNR ve Atış algoritmaları verildi. Bu algoritmalar Gauss yanıtları için uygulanabilir. Gauss olmayan yanıtlar için ise aşağıdaki IRLS yöntemi uygulanır.

IRLS yöntemiyle Bridge (LASSO) tahmin ediciler için Algoritma

1. $\hat{\beta}_0$ başlangıç değeri ile başla.
2. Genel tahmin $\hat{\beta}$ 'ya bağlı düzeltilmiş değişken z 'yi,

$$z = X\hat{\beta} + [V(\mu)]^{-1}(y - \mu)$$

olarak tanımla.

3. $W = V^{-1/2}$ olmak üzere WX üzerinde Wz 'nin lineer regresyonuna MNR (Atış) metodunu uygula, $\hat{\beta}$ 'yı güncelle.
4. $\hat{\beta}$ 'nın yaklaşımı elde edilinceye kadar 2. ve 3. adımlarını tekrarla.

Burada, eğer ortak likelihood fonksiyonu yoksa Jacobian koşulu sağlandığı sürece Bridge (LASSO) tahmin edici elde etmek için MNR metodu ya da Atış metodu uygulanabileceğine dikkat edilmelidir.

Algoritmaların Yakınsaması:

$\lambda > 0$ veriliyor. $\frac{\partial S}{\partial \beta}$ pozitif tanımlı ise;

1. MNR algoritması; $\gamma > 1$ için (2.3) 'ün Bridge tahmin edicisine yakınsar.
2. Atış algoritması; $\gamma = 1$ için (2.3) 'ün LASSO tahmin edicisine yakınsar.

Şimdiye kadarki yapılan çalışmalara göre; MNR ve Atış algoritmaları çok hızlı yakınsar ve IRLS yöntemi ile birleştirilebilir.

4. CEZALI GEE

4.1. Giriş

Halk sağlık çalışmalarında araştırmacılar, uzun zamanı kapsayan gözlemler serisini incelerler. Örneğin nefes darlığı ile ilgili çalışmalarda, çalışmadaki konuların her biri, bir yıl gibi belirli bir zaman dilimi için kontrol edildi. Konunun nefes darlığı ile ilgili durumu, her ziyarette havanın kalitesi, mevsim, sıcaklık, nem gibi bazı etkenlerle birlikte incelendi. Araştırmacıları asıl ilgilendiren, çok sık olarak nem ve sıcaklık gibi açıklayıcı değişkenlerle, astım durumu gibi yanıt değişken arasında ilişki bulundu. Bu tarz çalışmalar, longitudinal (boylamasına) çalışmalar olarak adlandırılan özel istatistiksel ayardır ve amaç, açıklayıcı değişkenlerde yanıtın zaman trendinin bağımlılığını teşhis etmektir.

Son 20 yıldır, longitudinal çalışmalar birçok sağlık araştırmacıları ve istatistikçiler tarafından ilgi çekmektedir ve longitudinal çalışmaların uygulamaları tıbbi, çevresel ve psikolojik çalışmalar gibi pek çok araştırmada bulunabilir (Laird ve Ware, 1982; Liang ve ark., 1992). Longitudinal çalışmalarda istatistiksel metotlar rasgele etki modelleri, koşullu Markov zinciri modelleri ve GEE metodunu içerir (Diggle ve ark., 1993).

Bu bölümde GEE metodu ve açıklayıcı değişkenler arasında ilişki olduğunda cezalı skor denklemlerine yaklaşım ile cezalandırma kullanma üzerinde durulmuştur.

4.2. Genelleştirilmiş Tahmin Denklemleri

Kategorik verilerin analizinde tekrarlamalı ölçümler olduğu zaman genellikle genelleştirilmiş tahmin denklemleri (GEE) kullanılarak analiz yapılmaktadır. GEE yaklaşımı (Liang ve Zeger, 1986) GLM'in bir açılımıdır. GEE, bireyler üzerinde uzun dönemli elde edilen gözlemler (longitudinal data) için bir yaklaşım sağlamaktadır. GEE ikili veya sayıma dayalı olarak elde edilen cevap değişkenlerinin analiz edilmesinde ideal bir yöntemdir. GEE'de uygun bir çalışılan (working) korelasyon matrisi tanımlanarak analiz yapılmaktadır. Çalışılan korelasyon matrisi

tüm bireyler için aynı olan a bilinmeyen parametre vektörüne bağlıdır (Davis, 2002).

GEE ortak dağılımın tam olarak tanımlanmasına gerek duyulmadan, tahmin denklemlerinin elde edildiği yarı parametrik bir yöntemdir. Bunun yerine marjinal dağılımlar için yalnızca olabilirliği ve her bir bireyden elde edilen tekrarlanmalı ölçümlerin vektörü için bir çalışılan kovaryans matrisi tanımlar (Davis, 2002; Liang ve Zeger, 1986). Çalışılan korelasyon matrisi her bir bireyin tekrarlanmalı ölçümleri için hesaplanır.

GEE yaklaşımında her bir denek bir küme olarak adlandırılır. Farklı kümeler için elde edilen gözlemlerin bağımsız, aynı küme için elde edilen gözlemlerin ise birbiriyle ilişkili olduğu düşünülür. GEE yöntemi, etkinliği arttırmak amacıyla bu ilişkiyi de dikkate alan tahmin tekniğidir. Söz konusu ilişki $R_i(a)$ ile gösterilen $n_i \times n_i$ boyutlu simetrik matrisle ifade edilir. Bu matrise aynı zamanda “üzerinde çalışılan ilişki matrisi” de denilir. Matrise bu ismin verilmesinin bir nedeni, bu ilişkinin yanlış tanımlanmış olma olasılığından kaynaklanmaktadır.

Yarı-likelihood fonksiyonunu GEE’ye uygulayabilmek amacıyla, yanıt vektörünün ortalama ve kovaryansı düşünülmelidir. Buna göre yarı-likelihood yaklaşımında üzerinde çalışılan kovaryans matrisi eşitlik (4.1)’de olduğu gibi hesaplanır:

$$V_i = A_i^{1/2} R_i(a) A_i^{1/2} \quad (4.1)$$

Burada; A_i ; $n_i \times n_i$ boyutlu köşegen matrisini ($A_i = \text{diag}(a_i''(\theta))$), $R_i(a)$; üzerinde çalışılan ilişki matrisini göstermektedir. (4.1) eşitliğinde denekler t zamanlarında t_{ij} kez gözlenmişlerdir. Buna göre $j = 1, \dots, n_j$ olmaktadır. Farklar vektörü aşağıdaki gibi tanımlanmış olsun:

$$S_i = y_i - \mu_i \quad (4.2)$$

(4.3) ile verilen denklem çalışılan kovaryans matrisidir.

$$\sum_{i=1}^K D_i^T V_i^{-1} S_i = 0 \quad (4.3)$$

β 'nın tahmin edilmesi için GEE, GLM'e benzer olarak,

$$\sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} (y_i - \mu_i) = 0$$

biçiminde yazılabilir. Y_i 'nin kovaryans matrisi,

$$V_i = A_i^{1/2} R(a) A_i^{1/2}$$

biçiminde tanımlanabilir.

$$\begin{aligned} D_i &= d \left\{ a_i'(\theta) \right\} / d\beta = \frac{\partial \mu_i}{\partial \beta} = A_i \Delta_i X_i \\ \Delta_i &= \text{diag} (d\theta_{it} / d\eta_{it}) \\ \eta_{it} &= x_{it}^T \beta \\ A &= \text{diag} (a_i''(\theta)) \\ S_i &= y_i - a_i'(\theta) \end{aligned}$$

Buradan $a_i'(\theta) = \mu_i$ olduğu anlaşılır.

Burada; $D_i = \frac{\partial \mu_i}{\partial \beta}$ olarak tanımlanan vektörü $(\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{in})^T)$, V_i ; (4.1)

eşitliğinde tanımlanan kovaryans matrisini, $i = 1, \dots, K$ olmak üzere denekleri göstermektedir. GEE yönteminde regresyon katsayıları tahminleri olasılıklar oranı yardımıyla hesaplanmakta ve yorumlar bu tahminlere göre yapılmaktadır.

GEE yaklaşımında, tutarlı ve asimptotik normal dağılım özelliklerini sağlayan regresyon katsayı tahminlerini ya da tutarlı varyans tahminlerini elde edebilmek için

çalışılan korelasyon matrisinin doğru belirlenmesi gerekmektedir. Birim sayısının çok fazla olduğu durumlarda asimptotik özellikler sağlanır, tutarlı ve etkin tahminler elde edilebilir. Bu gibi durumlarda bile korelasyon yapısının doğru tahmin edilmesi etkinlikte artan bir kazanca neden olur (Yazıcı, 2001).

GEE yönteminde bütün denekler için aynı korelasyon yapısının benimsenmiş olması şart değildir. Sabit bir korelasyon yapısının benimsenebilmesi, sadece eksik gözlemlerin tamamen rassal olması durumunda gerçekleşecektir. Dolayısıyla uygulamada her bir denek için gözlem sayısı eşit olmadığından, eksik gözlemlerde rassallığın sağlanması koşulu ile bu ilişkinin tüm denekler için sabit olduğu varsayılır.

Eşitlik (4.1) dikkate alınarak $R(a)$ aşağıdaki gibi yazılabilir (Lipsitz ve ark. 1974; Lipsitz ve Fitzmaurice, 1996):

$$R(a) = \begin{pmatrix} A_{i1}^{-1/2} V_{i1} A_{i1}^{-1/2} & \rho_{i12} & \cdots & \rho_{i1t} \\ \rho_{i12} & A_{i2}^{-1/2} V_{i2} A_{i2}^{-1/2} & \cdots & \rho_{i1t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i1t} & \rho_{i12} & \cdots & A_{it}^{-1/2} V_{it} A_{it}^{-1/2} \end{pmatrix} \quad (4.4)$$

Burada; $R(a)$ matrisi $(K-1) \times (K-1)$ boyutludur. Matrisin köşegen elemanları gözlemler arasındaki ilişkiyi göstermektedir.

GEE’nde yer alan farklı korelasyon yapıları şunlardır:

Yapılandırılmamış korelasyon: Genel olarak kümelerdeki gözlem sayıları çok azken, eksik gözlem bulunmaması durumunda uygun korelasyon yapısı yapılandırılmamış (unstructured) korelasyon yapısıdır. Örneklem yeterince büyük olduğunda yapılandırılmamış korelasyon yapısı kullanıldıysa tahminler tutarlı olacaktır.

Diğer taraftan aynı denemelerde yanıt değişkeni çok farklı değerler alıyorken yine yapılandırılmamış korelasyon kullanılmasında fayda vardır.

Bağımsız korelasyon yapısı: Örneklemede yer alan birim sayısı çok fazla olduğunda değişkenler arasında bağımsızlık varsayımı kullanılabilir. Bu durumda dikkat edilmesi gereken; eksik gözlemlerin mümkün olduğunca az sayıda ve tamamen rassal olmasıdır.

Birim sayısı az ancak her bir kümedeki tekrarlı gözlem sayısı fazla olduğu durumlarda da doğru korelasyon yapısının belirlenmesi ile etkinlikte biraz kazanç sağlanabilir. Bu durumda yine uygun korelasyon yapısı bağımsız korelasyon yapısıdır.

Yaş gibi zamana bağlı eş değişkenler varken bağımsız korelasyon yapısı kullanılarak elde edilen tahminler, diğer korelasyon yapısı varken elde edilen tahminlere göre daha az etkindirler.

Bağımsız korelasyon yapısı aşağıdaki gibi tanımlanır:

$$R_{i,j} = \begin{cases} 1 & \text{eğer } i = j \\ 0 & \text{d.d.} \end{cases}$$

$$R(\alpha) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Korelasyon yapısı tanımlanırken bağımsızlık varsayımı dikkate alındığından burada herhangi bir parametrenin tahmin edilmesine gerek yoktur.

Değiştirilebilir (exchangeable) korelasyon yapısı: GEE’nde en çok kullanılan korelasyon yapılarından biridir. Korelasyon yapısının belirlenmesi aşağıdaki eşitlik ile ilgilidir:

$$\rho_{ij} = a^{|j-i|^k}$$

Bu eşitlikte $k=0$ olması durumunda kullanılan korelasyon yapısı değiştirilebilir korelasyon yapısı olarak adlandırılır (Zeger ve Liang, 1986). Bir diğer ifade ile $\rho_{12} = \rho_{13} = \dots = \rho_{1j}$ olmak üzere boylamasına çalışmanın tüm değişkenlerine ilişkin korelasyonun birbirine eşit olduğu söylenir. Bu durum çoğu zaman boylamasına verilerde gözlemlerin çok kısa aralıklarla elde edildiği durumlarda gerçekleşebilir.

Tekrarlı gözlemlerin olduğu durumlarda gözlemlerin elde edilmesinde mantıklı bir sıra söz konusu değilse yine değiştirilebilir korelasyon yapısı uygundur.

Tüm gözlemlerde korelasyonun aynı olduğu değiştirilebilir korelasyon yapısı aşağıda verildiği gibidir:

$$R_{i,j} = \begin{cases} 1 & \text{eğer } i = j \\ \rho & \text{d.d.} \end{cases}$$

$$R(a) = \begin{bmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 1 \end{bmatrix}$$

Bütün değişkenler için korelasyon katsayılarının eşit olduğu düşünüldüğünden tahmin edilmesi gereken parametre sayısı 1'dir.

Otoregresif korelasyon yapısı: Longitudinal veri analizinde çoğunlukla verilerin ard arda gözlemlenmesi nedeniyle birinci dereceden otoregresif korelasyon meydana gelir. Bu durum $\rho_{ij} = a^{|j-i|}$ denkleminde $k=1$ olması ile ifade edilir. GEE'nde çok sık kullanılan korelasyon yapılarından biridir.

Otoregresif korelasyon yapısı için çalışılan korelasyon matrisi aşağıda olduğu gibidir.

$$R_{i,j} = \begin{cases} 1 & \text{eğer } i = j \\ \rho^{|j-i|} & \text{d.d.} \end{cases}$$

$$R(a) = \begin{bmatrix} 1 & \rho & \dots & \rho^{t-1} \\ \rho & 1 & \dots & \rho^{t-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{t-1} & \rho^{t-2} & \dots & 1 \end{bmatrix}$$

Bir önceki korelasyon katsayısı da dikkate alındığından elde edilen parametre sayısı 1'dir.

Bu 4 tip korelasyon yapısından başka; M-bağımlı korelasyon yapısı, sabit korelasyon yapısı, üssel korelasyon yapısı adı altında başka korelasyon yapıları da vardır. Ancak bu korelasyon yapılarının kullanımı diğerlerine oranla daha azdır.

Hangi korelasyon yapısının kullanılacağına karar verilirken benimsenen yollardan biri, uygun görülen birkaç korelasyon yapısını denemek ve daha sonra model temelli varyansa en yakın deneysel varyansı veren korelasyon yapısını seçmektir (Yazıcı, 2001).

Bir GEE düşünelim. Lineer regresyonlardaki gibi, iç ilişkinin potansiyel problemi de oluşur. Yani, eğer genelleştirilmiş tahmin denklemlerindeki açıklayıcı değişkenler lineere yakınsa, tahmin ediciye dayalı öntahminler kötü performans gösterecektir. Böylece cezalandırma, önceki bölümde gösterildiği gibi istenir. Ancak cezalandırmanın klasik yaklaşımı, örneğin Bridge regresyon, Bölüm 3'teki gibi ortak likelihood fonksiyonlarının varlığı gerekir.

Cezalı skor denklemleri yaklaşımı, cezalandırmayı geneller ve cezalı skor denklemleri ortak likelihood fonksiyonlarına bağlı olmadığından, GEE'de içilişki problemini kullanmak için teknikleri sağlar ve IRLS yöntemi uygulanabilir.

Aşağıda genelleştirilmiş tahmin denklemlerine cezalı skor denklemleri uygulanacak ve daha iyi tahmin ve öntahmin elde etmek için cezalı GEE çözülecektir.

4.3. Cezalı Genelleştirilmiş Tahmin Denklemleri

Longitudinal çalışmalarda GEE geniş alanlarda kullanılır. GEE tahmin edicisi asimptotik tutarlı ve yeterli olmasına rağmen açıklayıcı değişkenlerin içilişkili olması durumunda karşılanabilir. Özellikle açıklayıcı değişkenlerin fazla sayılı olduğu durumlarda karışıktır. Bu, (4.3)'ün parametre tahmin edicisi $\hat{\beta}$ 'ya bağlı tahmin ve öntahmin doğruluğu sorununa neden olur.

Cezalandırmanın, lineer regresyonda içilişki problemini ele alma tekniklerini kanıtladığı bilinir. Cezalandırmaya klasik yaklaşım, ceza fonksiyonlu modelin sapmasını minimize eder. Örneğin $L(\beta)$ ortak likelihood fonksiyonu ise, Bridge ceza için cezalandırma problemi

$$\min_{\beta} \left(-2 \log L(\beta) + \lambda \sum |\beta_j|^\gamma \right)$$

dir. Ancak genelde GEE için ortak likelihood fonksiyonu $L(\beta)$ bulunmaz. GEE'ne cezalandırma uygulamak için ortak likelihood fonksiyonuna bağlı olmayan özel teknikler gereklidir.

Aşağıda GEE 'ne Bridge ceza uygulanmıştır.

$$\begin{cases} S_1(\beta, X, y) + d(\beta_1, \lambda, \gamma) = 0 \\ \dots \\ S_p(\beta, X, y) + d(\beta_p, \lambda, \gamma) = 0 \end{cases} \quad (4.5)$$

$d(\beta_j, \lambda, \gamma) = \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j)$, S_j 'ler GEE 'nin (-) tahmin fonksiyonları veya ortak likelihood fonksiyonunun (-) skor fonksiyonlarıdır.

Cezalı skor denklemleri yaklaşımı ile GEE cezalandırılabilir. Açıklayıcı değişkenler arasında içilişki olduğunda daha iyi öntahmin ve küçük varyans elde etmek için cezalı GEE, GEE tahmin edicilerini 0'a bürer. Cezalı GEE tahmin

edicisinin çözümü için aşağıdaki yöntem izlenir (Liang ve Zeger, 1986) ve IRLS yönteminde Ağırlıklandırılmış EKK'e cezalandırma uygulanır.

Cezalı GEE için Algoritma

1. $\hat{\beta}_0$ başlangıç değeri ile başla.
2. Şimdiki tahmin $\hat{\beta}$ 'ya bağlı sapan rezidüleri veya Pearson kullanan $R(a)$ çalışılan korelasyon matrisini ve a, Φ parametrelerini tahmin et.
3. Düzeltilmiş bağımlı değişken $z = D\hat{\beta} + S$ tanımla.
4. MNR (Atış) metodunu kullanarak $\hat{V}$ ağırlıkları ile X üzerinde z 'nin regresyonuna cezalandırma uygulayarak, sabit $\lambda \geq 0$, $\gamma \geq 1$ için $\hat{\beta}$ tahmin edicisini güncelle.
5. $\hat{\beta}$ 'da yakınsama elde edinceye kadar 2.- 4. adımlarını tekrarla.

Bridge (LASSO) tahmin edicileri için cezalı GEE çözümü ile, açıklayıcı değişkenler arasında içilişki olduğunda, daha iyi tahmin ve öntahmin elde edilir.

5. BÜZÜLME PARAMETRELERİNİN SEÇİMİ

5.1. Giriş

Regresyon problemlerinde,

- (1) Veriye iyi uyum sağlamak
- (2) Basit ya da açıklanabilir bir modeli korumak

kurallarına uygun olarak model seçmek gereklidir. İlkinde mümkün olduğu kadar çok açıklayıcı değişkenle model açıklanmak istenirken; ikincisinde, istatistiksel olarak anlamlı olmayan açıklayıcı değişkenlerin modelde olmaması istenmektedir. Ancak çok fazla sayıda açıklayıcı değişken varsa, aynı zamanda hem (1) hem de (2)'yi sağlayan iyi modeli seçmek genelde zordur. Çok sık olarak, çok açıklayıcı değişkenli geniş modele sahip olunmaktadır. Bu modellerde asıl problem over-fitting (aşırı-uyum)'dir.

Over-fitting (aşırı uyum), modeller gereğinden fazla açıklayıcı değişken içeriyorsa ortaya çıkar ve verilen tüm veri noktalarında, veri modele aşırı uyum sağlar. Öntahminde model çok kötü performans gösterir.

5.2. Çapraz Geçerlilik Testi ve Genelleştirilmiş CV

Stone (1974), over-fitting problemini kullanmak için çapraz geçerlilik testini (CV) önermiştir. Her defasında bir gözlem noktasını dışarıda bırakarak model seçer ve geriye kalan veri noktaları ile model oluşturarak dışarıda bırakılan gözlemlerde ortalama öngörü hatasını minimize eder. Yani;

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}^{-i})^2$$

ve

$$\hat{y}^{-i} = x_i^T \hat{\beta}^{-i}(\lambda)$$

olmak üzere

$$\min_{\lambda} CV(\lambda)$$

dir. $\hat{\beta}^{-i}(\lambda)$, (x_i, y_i) dışında diğer gözlemlere dayalı modelin tahminidir ve λ , model seçimi için ayar parametresidir. Model uydurma ve seçimlerinde CV metodunun birçok uygulaması vardır. Özellikle Stone (1974), Hastie ve Tibshirani (1990), Wahba (1990), Shao (1993) ve Zhang (1992)'de bulunabilir.

Craven ve Wahba (1979), λ düzeltme parametresini optimize etmek için lineer düzeltme splineleri için GCV'yi önermiştir. $Y = g + \varepsilon$ modelinin $\hat{g} = A(\lambda)y$ lineer operatörü için

$$GCV(\lambda) = \frac{\|(I - A(\lambda))y\|^2}{n[tr(I - A(\lambda))/n]^2}$$

formudur.

GCV'nin bir avantajı, n kere tahmin hesaplamaya gerek yoktur. Dışarıda kalan her bir nokta, çapraz geçerlilik testi için seçilir. Tam modelin toplam sapmasını (RSS), örneklem boyutunu ve modelin serbestlik derecesini hesaplamaya yeter. Böylelikle hesaplama olarak daha az masraflıdır ve S+ gibi programlama dilleri ile hesaplanabilir (Fu, 1998).

5.3. GCV ile λ ve γ Parametrelerinin Seçimi

λ ve γ ceza parametreleri seçmek için, Craven ve Wahba'nın GCV yöntemi kullanılır. İlk olarak; (2.3)'ten lineer regresyon modelinin Bridge tahmin edicisi

$$(X^T X + \lambda D)\beta = X^T y \quad (5.1)$$

eşitliğini sağlar. Modelin $p(\lambda, \gamma)$ parametrelerinin efektif (etkili) sayısını, Craven ve Wahba modelin serbestlik derecesinde ceza etkisini vermiştir.

$$p(\lambda, \gamma) = \text{tr} \left(X \left(X^T X + \lambda D \right)^{-1} X^T \right) - n_0$$

Buradaki D , $p \times p$ tipinde

$$D_j = \begin{cases} \gamma |\hat{\beta}_j|^{\gamma-2} / 2, & \hat{\beta}_j \neq 0 \\ 0, & \hat{\beta}_j = 0 \end{cases} \quad j = 1, \dots, p \text{ için}$$

elemanlarının diyagonal matrisi, n_0 ; $\gamma = 1$ için $\hat{\beta}_j = 0$ olan $\hat{\beta}_j$ 'nin sayısıdır.

GCV, n boyutlu için;

$$GCV = \frac{RSS(\hat{\beta})}{n(1 - p(\lambda, \gamma)/n)^2} \quad (5.2)$$

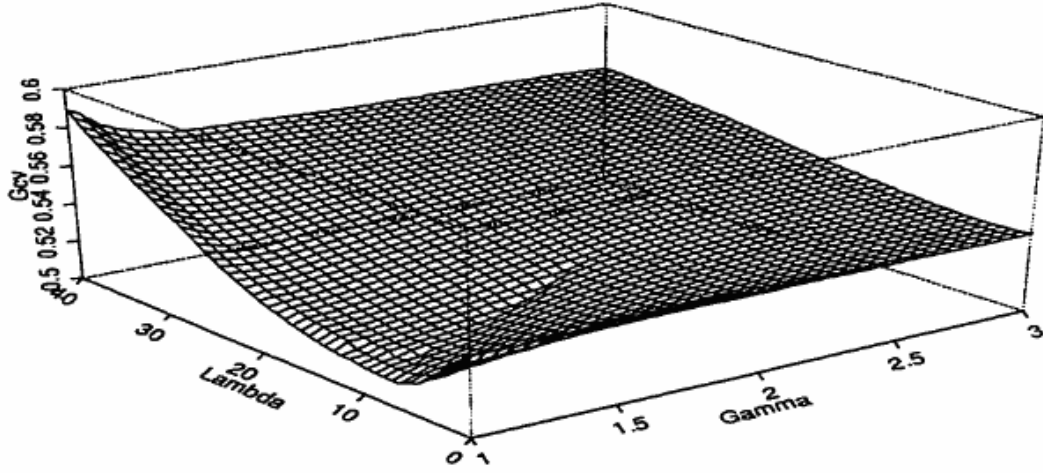
olarak tanımlanmıştır. (5.2) eşitliği

$$\frac{n \, RSS(\hat{\beta})}{(n - p(\lambda, \gamma))^2}$$

olarak da yazılabilir ve modelin geri kalan serbestlik derecesi üzerinde, ortalama hata kareler miktarı anlamına gelir.

λ ve γ parametreleri seçmek için, $\lambda \geq 0$ ve $\gamma \geq 1$ çizgisi üzerinde her bir (λ, γ) çifti için GCV hesaplanır. λ ve γ , Şekil 5.1'deki gibi GCV'nin minimum değerinin elde edilmesiyle seçilir.

Genelleştirilmiş lineer modellerde, Gauss-olmayan yanıt değişkenler için hata kareler toplamı anlamlı olmadığından GCV düzeltilir. $-2\log(Lik)$, sapma yerine; GCV'deki hata kareler toplamı (RSS) kullanılır. Lik yanıt değişkeninin ortak likelihood fonksiyonudur.



Şekil 5.1. GCV ile λ ve γ parametrelerinin seçimi (Fu, 1998)

$p(\lambda, \gamma)$ parametresinin efektif sayısı için 2 özel durum dikkate alınır.

1. $\lambda = 0$. Modele hiçbir ceza uygulanmaz. $p(\lambda, \gamma)$ projeksiyon matrisinin izidir ve lineer modelde parametrelerin sayısı p 'ye eşittir.
2. $\lambda \gg 1$ ve $\gamma = 1$. LASSO parametreleri büzdüğünden ve $\hat{\beta}_j = 0$ olduğundan, yeteri kadar büyük λ için, $D = \text{diag}(0)$ ve $n_0 = p$ 'dir. Tüm $\hat{\beta}_j = 0$ olduğunda; model boş (null) modeldir. Böylece model parametrelerinin efektif sayısı 0'a eşittir. Bu da $p(\lambda, \gamma) = p - p = 0$ demektir. Diğer durumlarda $p(\lambda, \gamma)$ 0'dan büyüktür ve modeldeki parametre sayısı p 'den küçüktür.

5.4. Cezalı GEE için Yarı-GCV

GCV metodu, genelleştirilmiş lineer modeller için λ ve γ parametrelerini seçmekte kullanılır. Ancak 4. bölümde üzerinde durulduğu gibi genelde GEE için ortak likelihood fonksiyonları yoktur. Genelde GCV metodu cezalı GEE'ne uygulanamaz. Bu nedenle düzeltme uygulanmalıdır.

Cezalı GEE’nde GCV metodunu genelleştirmek için korelasyon yapısı dahil edilmelidir. Korelasyonun yapısı dahil edilmesi ile GLM’deki gibi GCV’nin aynı etkisi elde edilebilir. GLM için GCV’de kullanılan sapma, sapma hatalarının kareleri toplamıdır. Korelasyon nedeniyle sapma, GEE’nde doğru anlama sahip olmamasına rağmen, sapma hataları

$$\text{sign}(y_{kt} - \hat{\mu}_{kt}) \sqrt{-2 \log L(y_{kt} - \hat{\mu}_{kt})}$$

olarak her bir gözlem noktasında hesaplanabilir. Buradaki $L(y_{kt} - \hat{\mu}_{kt})$, marjinal dağılımına dayalı Y_{kt} gözleminin likelihoodudur. İlişkili gözlemler için ağırlıklı sapma $D_w(\lambda, \gamma)$, bağımsız gözlemler için sapmaya benzer etki elde etmek için sapma rezidülerine korelasyon dahil edilmesiyle aşağıdaki gibi verilir.

$$D_w(\lambda, \gamma) = \sum_{k=1}^K r_k^T [R_k(a)]^{-1} r_k$$

Buradaki r_k , k. bileşenin sapma hata vektörüdür. $n_k \times n_k$ boyutlu $R_k(a)$, işleyen korelasyon matrisidir.

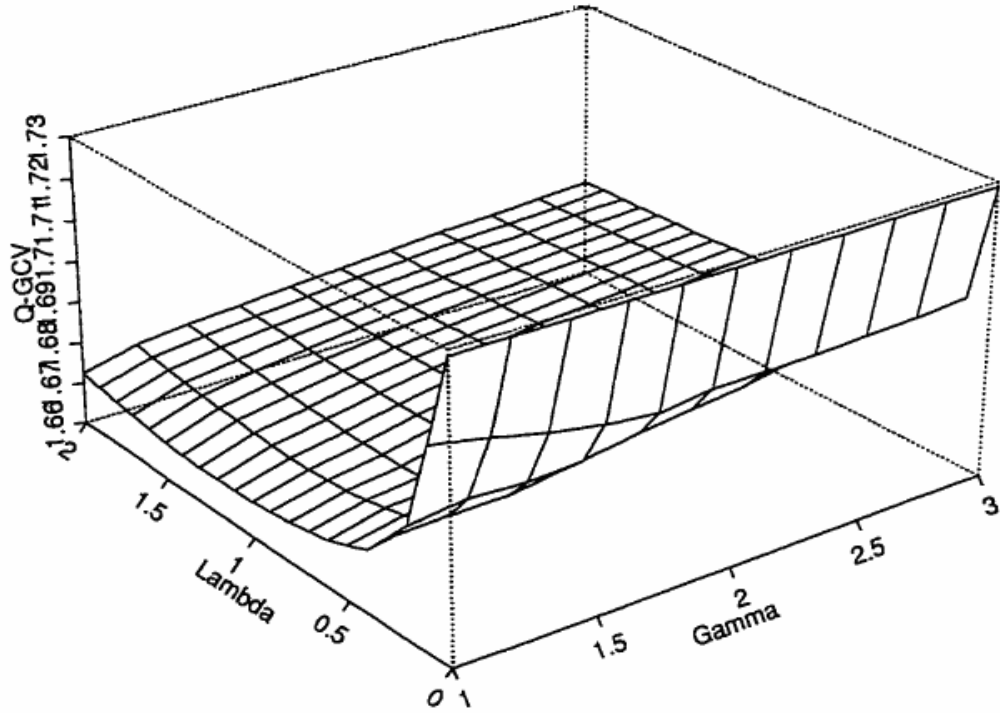
Yarı-GCV;

$$GCV_q = \frac{D_w(\lambda, \gamma)}{n(1 - p(\lambda, \gamma)/n)^2} \quad (5.3)$$

olarak tanımlanır. Buradaki n, Y_{kt} ilişkili gözlemlerin serbestlik derecesinin efektif sayısıdır ve

$$n = \sum_{k=1}^K \frac{n_k^2}{|R_k(a)|}$$

olarak tanımlanır. $|R_k(a)|$, $R_k(a) = (\rho_{ij})$ ’nin tüm $\sum \rho_{ij}$ elemanlarının toplamıdır. GEE’nin korelasyon yapısı ne Pearson rezidüleri ne de sapma rezidüleri ile tahmin edilebildiğinden, sapma rezidüleri sırayla korelasyon yapısı dahil etmek tavsiye edilir.



Şekil 5.2. Yarı-GCV ile λ ve γ parametrelerinin seçimi (Fu, 1998)

Parametre seçim yöntemi, GLM'deki ile aynıdır. Yani her bir sabit (λ, γ) çifti için Bridge (LASSO) tahmin edici $\hat{\beta}(\lambda, \gamma)$ hesaplanır, sonra $p(\lambda, \gamma)$ parametresinin efektif sayısı hesaplanır. Böylece yarı-GCV, sapma rezidüleri ile (5.3)'ün kullanılmasıyla hesaplanır ve $R(a)$ korelasyon matrisi, cezalı GEE için IRLS yönteminin son adımından elde edilir. λ ve γ parametreleri Şekil 5.2'de gösterildiği gibi yarı-GCV'nin minimize olduğu noktanın üstünde seçilir.

Uyarılar

1. $D_w(\lambda, \gamma)$ 'yi ağırlıklı sapma olarak düşünelim. Bağımsız gözlemler için korelasyon matrisi $R(a)$ birim matris olduğunda sapmayı azaltır. Bu nedenle yarı-GCV, GCV'ye indirgenir.
2. İlişkili gözlemlerin serbestlik derecelerinin efektif sayısı, $R(a)$ korelasyon matrisine bağlıdır. λ ve γ 'nın farklı değerleri, farklı tahminlere ve $R(a)$ 'nın farklı değerlerine neden olduğundan, n de λ ve γ gibi farklıdır. Ancak serbestlik derecesinin efektif sayısı, gözlemler ve bileşen için esas olduğundan, n ; λ ve γ 'dan bağımsız olmalıdır. Bu nedenle farklı λ ve γ 'ya göre yarı-GCV hesaplamak için sabit n kullanılmalıdır. $\lambda = 0$ 'dan n 'in tahminini kullanmak tavsiye edilir.

Ağırlıklı sapma, aşağıdaki gibi ilişkili Gauss yanıtlarına neden olur.

Σ , ε 'un diyagonal olmayan varyans-kovaryans matrisi ve $\varepsilon \sim N(0, \Sigma)$ olmak üzere, $Y = X\beta + \varepsilon$ modelinden $Y = (Y_1, Y_2, \dots, Y_n)^T$ ilişkili yanıtlar olsun. Sırasıyla bağımsız değişkenler için GCV metodu uygulanır, $\Sigma = Q^T \Lambda Q$ 'yu sağlayan $P = \Lambda^{-1/2} Q$ olmak üzere, dönüşümü $Z = PY$ alalım. Böylece Z , $N(PX\beta, I)$ normal dağılımına sahiptir.

$$\begin{aligned}
 GCV(\lambda, \gamma) &= \frac{(Z - PX\hat{\beta})^T (Z - PX\hat{\beta})}{n(1 - p(\lambda, \gamma)/n)^2} \\
 &= \frac{(Y - X\hat{\beta})^T P^T P(Y - X\hat{\beta})}{n(1 - p(\lambda, \gamma)/n)^2} \\
 &= \frac{(Y - X\hat{\beta})^T \Sigma^{-1} (Y - X\hat{\beta})}{n(1 - p(\lambda, \gamma)/n)^2}
 \end{aligned}$$

Yani GCV, rezidülerde korelasyon yapısının dahil edilmesiyle elde edilir. Benzer şekilde cezalı GEE’nde aynı etkiyi elde etmek için, (5.3) denklemindeki gibi sapma rezidüleri korelasyon yapısı dahil edilir.

İlişkili gözlemlerin serbestlik derecesinin efektif sayısı, ilişkili Gauss gözlemlerine de neden olur. $\rho_{ii}=1$ köşegen elemanlı $R=(\rho_{ij})$ matrisi olsun. $Y=(Y_1, Y_2, \dots, Y_n)^T$, $N(0, \sigma^2 R)$ dağılımına sahip olsun. $\bar{Y}$ örneklem ortalamasının varyansı;

$$\begin{aligned}
 Var(\bar{Y}) &= \frac{1}{n^2} Var\left(\sum Y_i\right) = \frac{1}{n^2} Cov\left(\sum Y_i, \sum Y_i\right) \\
 &= \frac{1}{n^2} \sum_i \sum_j Cov(Y_i, Y_j) \\
 &= \frac{\sigma^2}{n^2} \sum_i \sum_j \rho_{ij} \\
 &= \frac{\sigma^2}{n^2} |R| \\
 &= \frac{\sigma^2}{n^2 / |R|}
 \end{aligned} \tag{5.4}$$

dir.

Y_i gözlemlerinin bağımsız olduğu özel durumlarda R birim matristir, böylece örneklem ortalamasının varyansı $Var(\bar{Y}) = \sigma^2/n$ olur. Paydadaki n , $Y_1, Y_2, \dots, Y_n$ bağımsız gözlemlerinin serbestlik derecesinin sayısıdır. $Y_1, Y_2, \dots, Y_n$ ilişkili gözlemlerinin serbestlik derecesinin efektif sayısı, (5.4) denkleminin paydasını $n^2/|R|$ olarak tanımlayalım. Negatif olmayan ilişki sabiti $\rho_{ij} \geq 0$ için, bu serbestlik derecesinin efektif sayısı Y_1 ’in n kez tekrarı için ve $(Y_1, Y_2, \dots, Y_n)$ n bağımsız gözlemi için 1 ile n arasındadır.

Negatif korelasyon ile bazı problemler oluşabilir. Ancak uygulamada, negatif korelasyonlu gözlemlerin serisine rastlamak çok nadir bir durumdur. Özellikle longitudinal (boylamasına) araştırmalarda aynı bileşenden pozitif ilişkili yanıtlar beklenir. Böylece serbestlik derecesinin efektif sayısı, genelde boylamasına araştırmalarda iyi çalışır.

6. SİMÜLASYON ÇALIŞMALARI

Bu bölümde doğru modeldeki Bridge regresyonun büzülme etkisini sınamaya dayalı istatistiksel simülasyonlar verilecektir. Bridge ceza modeli; cezalı olmayan, LASSO ceza ve Ridge ceza modelleri ile karşılaştırıldı. Regresyon parametrelerinin standartlaştırılmış hata kareler ortalaması (MSE);

$$MSE = \text{ort} \left(\hat{\beta} - \beta \right)^T X^T X \left(\hat{\beta} - \beta \right)$$

ve öntahmin hata kareleri $PSE = \text{ort} \text{Dev}(y, \hat{\mu})$ model rasgele hatasının ortalaması alınarak hesaplanır

PSE, X gibi aynı korelasyon yapısına sahip ilişki uzayında, bazı rasgele seçilmiş noktadaki ortalama olarak hesaplanır. Her bir niceliğin de standart hatası hesaplanır. Bu bölümdeki simülasyonlar, Fu (1998) makalesinden yararlanılarak hazırlanmıştır.

6.1. Lineer Regresyon Modeli

40 gözlemlili ve 5 açıklayıcı değişkenli

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \quad (6.1)$$

basit modelinin simülasyonunda Bridge model ile OLS, LASSO ve Ridge'i karşılaştıralım. $\varepsilon \sim N(0, \sigma^2)$ 'dir. Buradaki β ; gerçek parametre ve x^i ; i . gözleme ait vektördür.

İçilişkide büzülme etkisini sınamak için X'in korelasyon matrisinde gösterildiği gibi güçlü lineer korelasyonlu X regresyon matrisi seçilir. x_4 ve x_5 arasında $\rho = 0.995$ alınarak çok güçlü korelasyon katsayısı seçilir. X matrisi aşağıdaki gibi üretilir. İlk olarak $N(0,1)$ standart normal dağılımlı rasgele sayıları ile

40×5 tipinde bir matris üretilir. Sonra X 'in ardışık kolon vektörlerinin çift yönlü korelasyon katsayıları, $U(-1,1)$ düzgün dağılımından üretilir.

Sabit terimsiz regresyon parametrelerini büzmek için,

$$x_j = \sqrt{n} (x_j - \text{ort}(x_j)) / \|x_j - \text{ort}(x_j)\|$$

ile veri ölçülür ve merkezileştirilir. Buradaki x_j ; X 'in j . kolon vektörüdür.

Doğru model 0 katsayısını içerdiğinde LASSO Ridge'den daha iyi ve doğru model küçük fakat sıfır olmayan parametreler içerdiğinde de LASSO Ridge'den daha kötü performans gösterdiğinden; Gerçek β değerinin iki kümesi, 0 katsayılı modeldeki büzülme etkisini sınaması için seçilir. Sıfır içermeyen fakat küçük katsayılı modeller:

$$(a) \text{ modeli için } \beta_0 = 0 \text{ sabit terimi ile } \beta_{true} = (0, 0, 0.5, 0, -1)^T$$

ve

$$(b) \text{ modeli için } \beta_0 = 0 \text{ sabit terimi ile } \beta_{true} = (0.5, 3, -1.0, 2.5, 9)^T \text{ 'dir.}$$

Y yanıtı (6.1) modelinden üretilir.

(6.1) Lineer modelinin korelasyon matrisi

$$\begin{pmatrix} x_1 & 1.000 & 0.110 & -0.144 & 0.036 & 0.066 \\ x_2 & 0.110 & 1.000 & -0.315 & 0.021 & 0.034 \\ x_3 & -0.144 & -0.315 & 1.000 & -0.118 & -0.109 \\ x_4 & 0.036 & 0.021 & -0.118 & 1.000 & 0.995 \\ x_5 & 0.066 & 0.034 & -0.109 & 0.995 & 1.000 \end{pmatrix}$$

Tablo 6.1. 200 kez tekrarlayan simülasyonla model karşılaştırılması

Model (a)

β	EKK	Bridge	LASSO	Ridge
$\beta_0 = 0.0$	0.006 (0.071)	0.006 (0.071)	0.006 (0.071)	0.006 (0.071)
$\beta_1 = 0.0$	0.004 (0.074)	-0.004 (0.046)	-0.004 (0.046)	-0.011 (0.069)
$\beta_2 = 0.0$	-0.007 (0.087)	-0.013 (0.060)	-0.013 (0.060)	-0.018 (0.083)
$\beta_3 = 0.5$	0.495 (0.081)	0.457 (0.080)	0.457 (0.080)	0.469 (0.080)
$\beta_4 = 0.0$	0.009 (0.837)	-0.267 (0.366)	-0.267 (0.366)	-0.394 (0.196)
$\beta_5 = -1.0$	-1.007 (0.843)	-0.700 (0.375)	-0.700 (0.375)	-0.585 (0.204)
MSE	1.385 (0.061)	1.104 (0.059)	1.104 (0.059)	1.212 (0.057)
PSE	3.178 (0.141)	2.482 (0.132)	2.482 (0.132)	2.745 (0.127)

Model (b)

β	EKK	Bridge	LASSO	Ridge
$\beta_0 = 0.0$	0.003(0.757)	0.003(0.757)	0.003 (0.757)	0.003 (0.757)
$\beta_1 = 0.5$	0.353 (0.860)	0.386 (0.783)	0.380 (0.791)	0.440 (0.783)
$\beta_2 = 3.0$	3.018 (0.789)	2.956 (0.768)	2.987 (0.773)	2.913 (0.730)
$\beta_3 = -1.0$	-1.007 (0.811)	-0.959 (0.769)	-0.960 (0.779)	-0.983 (0.757)
$\beta_4 = 2.5$	2.248 (0.837)	3.679 (4.715)	3.496 (4.877)	4.909 (1.610)
$\beta_5 = 9.0$	9.205 (7.826)	7.666 (4.883)	7.910 (5.024)	6.257 (1.819)
MSE	145.17 (5.97)	129.60 (5.50)	130.16 (5.52)	127.90 (5.70)
PSE	329.30 (14.22)	290.29 (12.87)	292.10 (12.90)	286.42 (13.26)

Tablo 6.1; parametre tahminleri, parantezlerde standart hataları, OLS'nin MSE ve PSE'si, Bridge, LASSO ve Ridge modellerini gösterir. $\hat{\beta}_4$ ve $\hat{\beta}_5$ 'nin standart hataları içilişki nedeniyle hem (a), hem de (b) modellerinde diğerlerine nispeten daha büyüktür.

(a) modelinde; Bridge ve LASSO sonra da Ridge en küçük MSE ve PSE değerini elde eder. İçilişki nedeniyle OLS en büyük MSE'ye ve en büyük PSE'ye sahiptir.

(b) modelinde; En küçük MSE'ye sırasıyla Ridge, Bridge, LASSO ve OLS sahiptir. Aynı zamanda en küçük öntahmin hatası PSE'ye de Ridge sahiptir. En büyük öntahmin hatası ise OLS'nindir.

Yukarıdaki örnekte görülüyor ki, Bridge regresyon OLS tahminlerini bürer ve küçük varyans, küçük MSE ve küçük öntahmin hatası elde edilir.

Bridge tahmin edici, LASSO ve Ridge tahmin ediciye kıyasla ve OLS tahmin ediciye göre daha iyi performans gösterir.

6.2. Karmaşık Lineer Regresyon Modeli

Bölüm 6.1'de basit lineer regresyon modelleri çalışıldı, OLS, Bridge, LASSO ve Ridge gibi farklı cezalarla bürölme etkisi ve MSE, PSE değerleri karşılaştırıldı. Bu bölümde açıklayıcı değişkenlerin farklı korelasyon yapısı ile daha karmaşık lineer regresyon modellerinde farklı bürölme etkisi çalışılacaktır. Gerçek parametreler, Bölüm 2.7'de bahsedildiği gibi γ 'nın farklı değerleri için Bridge cezanın prior dağılımından üretilir.

Model

$n = 30$ örneklem boyutlu, 10 açıklayıcı değişkenli

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_{10} x_{10} + \varepsilon$$

lineer regresyon modeli üzerinde çalışalım. 10 regresyon matrisi X_m , $m=1,2,...,10$, $U(-1,1)$ düzgün dağılımından üretilen farklı çift yönlü korelasyon katsayılı $\{\rho\}_m$ ile 30×10 boyutlu ortonormal X matrisinden üretilmiştir.

Veri

Her bir X_m için; gerçek β_k , $k=1,2,...,30$, üretilmiştir. Buradaki β_k 'nin her bir bileşeni Bridge prior $\Pi_\gamma(\beta)$ 'dan, yani $\lambda=1$ ve $\gamma \geq 1$ ile $\Pi_{\lambda,\gamma}(\beta)$ 'dan üretilmiştir. Her bir X_m ve β_k ile 30 gözlem $N(0, \sigma^2)$ normal dağılımından rasgele hata ε_i ile $Y = X_m \beta_k + \varepsilon$ modelinden üretilmiştir. Farklı ceza modelleri: OLS, Bridge, LASSO ve Ridge için MSE ve PSE;

$$MSE = (\hat{\beta} - \beta)^T X_m^T X_m (\hat{\beta} - \beta)$$

ve

$$PSE = \text{ort} \left(y_i - x_i^T \hat{\beta} \right)^2$$

olarak hesaplanmıştır.

20 rasgele seçilmiş nokta (x_i, y_i) , aynı modelden üretilmiştir. x_i ; her bir öntahmin noktasına ait vektördür. MSE ve PSE, rasgele model hatası ε 'un 50 tekrarlı ortalamasıdır. β_k ; $\Pi_\gamma(\beta)$ prior dağılımından üretilir. MSE ve PSE, OLS, Bridge, Ridge ve LASSO modelleri için hesaplanır. Böylece MSE ve PSE'nin $10 \times 30 = 300$ kümesi hesaplanır. Yukarıdaki yöntem $\gamma = 1, 1.5, 2, 3, 4$ değerleri için tekrarlanır.

Metot

Farklı cezaların her bir MSE ve PSE kümesi, $\Pi_\gamma(\beta)$ 'dan üretilen aynı β_k 'lerden hesaplanır ve değerleri farklı β_k ile geniş aralıkta değişmektedir. Fakat

modeller arası farklar küçüktür. OLS yardımıyla, bağıntılı (related) MSE_r ve bağıntılı (related) PSE_r 'nin karşılaştırılmasıyla seçilir.

$$MSE_r = \frac{MSE - MSE_{OLS}}{MSE_{OLS}}$$

ve

$$PSE_r = \frac{PSE - PSE_{OLS}}{PSE_{OLS}}$$

Karşılaştırıldığında, bağıntılı MSE_r ve bağıntılı PSE_r orijinal MSE ve PSE'ye tercih edilir (Fu, 1998).

Sonuç

Her bir sabit γ değeri için, 300 kümeli MSE_r ve PSE_r 'nin ortalama ve standart hataları hesaplanmış ve Tablo 6.2'teki gibi verilmiştir. $\gamma=1$ ve 1.5 için gösterildiği gibi Bridge, LASSO ve Ridge'nin MSE ve PSE'si OLS'ninkinden daha küçüktür. Tablo 6.2'den görüldüğü gibi Bridge ve LASSO küçük γ değerleri için iyi performans gösterirken, büyük γ değerinde o kadar iyi performans göstermemektedir.

γ 'nın büyük değerleri ($\gamma=1.5, 2, 3, 4$) için Ridge; Bridge ve LASSO'dan daha iyidir. Bölüm 2.6 ve 2.7 'de de bahsedildiği gibi γ 'nın büyük değerleri, model için küçük fakat sıfır olmayan β parametreleri türetir ve γ 'nın küçük değerleri sıfır da içeren β regresyon parametreleri üretir. Bu da, doğru model sıfır parametreleri içeriyorsa LASSO iyi performans gösterirken, doğru model küçük fakat sıfır olmayan parametreleri içeriyorsa LASSO'nun kötü performans gösterdiği anlamına gelir. Bridge de LASSO'ya benzer etki göstermektedir. $\gamma=1, 1.5$ gibi küçük değerlerde iyi, fakat γ 'nın büyük değerlerinde kötü performans gösterir.

Tablo 6.2. Farklı γ değerleri için MSE_r ve PSE_r 'nin ortalama ve standart hataları

γ	Bridge		Lasso		Ridge	
	MSE_r	PSE_r	MSE_r	PSE_r	MSE_r	PSE_r
1	-.0860(.0044)	-.0021(.0002)	-.0841(.0043)	-.0020(.0004)	-.0595(.0030)	-.0013(.0002)
1.5	-.0225(.0054)	-.0009(.0003)	-.0224(.0054)	-.0009(.0003)	-.0566(.0032)	-.0017(.0002)
2	.0176(.0053)	-.0002(.0005)	.0176(.0053)	-.0002(.0005)	-.0519(.0028)	-.0021(.0003)
3	.0349(.0048)	.0005(.0003)	.0350(.0048)	.0005(.0003)	-.0566(.0029)	-.0016(.0002)
4	.0377(.0048)	.0001(.0003)	.0377(.0048)	.0001(.0003)	-.0577(.0027)	-.0018(.0002)

Geniş aralıkta MSE ve PSE 'lerinin değerleri farklıdır. Fu (1998), orijinal MSE ve PSE değerlerindense, farklı ceza modelleri arasında MSE_r ve PSE_r 'yi tercih etmenin daha uygun olduğu sonucuna varmıştır.

Yukarıdaki sonuca göre; Bridge regresyon küçük MSE ve PSE elde eder, büyük regresyon parametrelili lineer regresyon modeli için LASSO ve Ridge'e göre daha iyi performans verir. Fakat sıfır olmayan küçük parametrelili doğru modele sahipse kötü performans gösterebilir.

7. SONUÇ VE ÖNERİLER

7.1. Sonuçlar

İçilişki, regresyonda bir problemdir. Tahmin ve öntahmini etkiler, böylece araştırmalara çok büyük etkisi vardır. İçilişkiyi bulabilmek için Temel Bileşenler Analizi, büzülme modeli gibi birçok önemli metot vardır.

Ridge regresyon ve LASSO gibi iki önemli üyesi olan Bridge regresyon, cezalı regresyonların özel sınıflarındandır, içilişki problemini ele almada önemli rol oynarlar. Tahmin ediciyi 0'a doğru büzerek, iyi tahmin, öntahmin ve küçük varyans elde eder.

Bridge tahmin edici basit yapıya sahiptir ve basit algoritmalarla hesaplanabilir. Lineer regresyon modelleri için tahmin ve öntahmine göre Bridge tahmin edicinin daha iyi performans gösterdiği simülasyon çalışmalarından da görülmektedir. $\gamma > 1$ Bridge tahmin edicinin çözümü için MNR metodu ve $\gamma = 1$ Bridge tahmin edici için Atış metodu kullanılır. Bölüm 3'te cezalı skor denklemleri, Bölüm 4'te GEE'nin ortak likelihood fonksiyonundan bağımsız olduğundan ve genelde GEE için ortak likelihood fonksiyonu olmadığından Bölüm 5'te bahsedilen GCV metodu, cezalı GEE için uygulanamaz.

7.2. Öneriler

Daha ileri araştırmalarda, istatistiksel modellemede cezalandırma konusuna daha çok ihtiyaç duyulacaktır. X 'de içilişki olduğu durumlarda geliştirilen yöntemler EKK regresyonu ile karşılaştırıldığında iyi sonuçlar vermesine rağmen, parametre tahmininde daha iyi sonuçlar verebilecek yeni tahmin yöntemleri geliştirilmelidir.

KAYNAKLAR

- BÜHLMANN, P. ve YU B., 2003. Boosting with the L2 loss: Regression and classification, *Journal of the American Statistical Association* 98:324–339.
- CRAVEN, P., ve WAHBA, G., 1979. Smoothing Noisy Data With Spline Functions, *Numerische Mathematik*, 31:377–403.
- DAVIS, C.S., 2002. *Statistical Methods for the Analysis of Repeated Measurements*, Heidelberg: Springer Verlag, 2002, p 415.
- DIGGLE, P. J., LIANG, K. Y. ve ZEGER, Ç. L. 1994. *Analysis of Longitudinal Data*, Clarendon, Oxford.
- DOBSON, J.A., 1990. *An Introduction to Generalized Linear Models*, Chapman and Hall, New York. 174.
- EFRON, B., ve TIBSHIRANI, R.J., 1993. *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- FAHRMEIR, L. ve G. TUTZ, 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd edition.
- FAN, J. ve LI, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Statist. Ass.*, 96:1348–1360.
- FAN, J. ve LI, R., 2006. Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery, *Proceedings of the International Congress of Mathematicians* (M. Sanz-Sole, J. Soria, J.L. Varona, J. Verdera, eds.) Vol. III:595–622.
- FRANK, I.E., ve FRIEDMAN, J.H., 1993. A Statistical View of Some Chemometrics Regression Tools,” *Technometrics*, 35:109–148.
- FRIEDMAN, J., HASTIE, T. ve TIBSHIRANI, R., 2000. Additive Logistic Regression: A Statistical view of Boosting (with discussion), *Annals of Statistics* 28:337-307
- FRIEDMAN, J., HASTIE, T., ROSSET, S., TIBSHIRANI, R. ve ZHU, J., 2004. Discussion of boosting papers, *Ann. Statist.*, 32:102–107.
- FU, W., 1998. Penalized regression: the bridge versus the LASSO, *J. Computational and Graphical Statistics*, 7:397–416.

- FURNIVAL, G.M., ve WILSON, R.W., JR., 1974. Regressions by Leaps and bounds, *Technometrics*, 16:499–511.
- GREEN, P.J., 1984. Iteratively Reweighted Least Squares For Maximum Likelihood Estimation, And Some Robust And Resistant Alternatives (with discussion). *Journal of Royal Statistical Society, B* 46:149-192.
- GILL, P.E., MURRAY, W., ve WRIGHT, M.H., 1981. *Practical Optimization*, London: Academic Press.
- GOLUB, G. ve VAN LOAN, C., 1983. *Matrix Computations*, Baltimore: Johns Hopkins University Press.
- HASTIE, T.J. ve TIBSHIRANI, R.J., 1990. *Generalized Additive Models*. Chapman And Hall, New York.
- HOCKING, R. R., 1996. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. Wiley, New York.
- HOERL, A.E., ve KENNARD, R.W., 1970a. Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12:55–67.
- HOERL, A.E., ve KENNARD, R.W., 1970b. Ridge Regression: Applications to Nonorthogonal Problems, *Technometrics*, 12:69–82.
- HUANG, F., 2003. A Prediction Error Property of the Lasso and Its Generalization, *Aust. N. Z. J. Stat.* 45:217–228.
- HUANG, J., HOROWITZ J. L. ve SHUANGGE M., 2008. Asymptotic Properties Of Bridge Estimators in Sparse High-Dimensional Regression Models, *Ann. Statist.* Volume 36, Number 2 (2008), 587-613.
- KNIGHT, K., ve FU, W., 2000. Asymptotics For Lasso-Type Estimators, *Annals of Statistics*, 28:1356–1378.
- LAIRD, N. M. ve WARE, J. H. 1982. Random-effects models for longitudinal data. *Biometrics*, 38:963-974.
- LAWSON, C., ve HANSEN, R., 1974. *Solving Least Squares Problems*, Englewood Cliffs, NJ: Prentice-Hall.
- LI, B. ve MCCULLAGH, P. 1994. Potential Functions and Conservative Estimating Functions. *The Annals of Statistics* Vol. 22, No. 1, 340-356.

- LIANG, K.Y. ve ZEGER, S. L. 1986. Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* 73:13-22.
- LIANG, K. Y., ZEGER, S. L. ve QAQISH, B. 1992. Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society B* 34:4-40.
- LIPSITZ, S. R., FITZMAURICE, G. M., ORAV, E. J. and LAIRD, N. M., 1974. Performance of Generalized Estimating Equations in Practical Situations, *Biometrics*, p. 270-278.
- LIPSITZ, S. R. and FITZMAURICE, G. M., 1996. Estimating Equations for Measures Of Association Between Repeated Binary Responses, *Biometrics* 52, p. 903-912.
- LITTELL, C.R, MILLIKEN, A.G, STROUP, W.W., WOLFINGER, D.R., 1996. SAS System for Mixed Models, SAS Institute Inc., Cary, NC.
- MCCULLAGH, P., 1980. Regression model for ordinal data (with discussion), *Journal of the Royal Statistical Society B* 42:109–127.
- MCCULLAGH, P. ve J. A. NELDER, 1989. *Generalized Linear Models*, (2nd ed.). New York: Chapman & Hall.
- MCCULLAGH, P. 1991. Quasi-likelihood and estimating functions In *Statistical Theory and Modelling: In Honour of Sir David Cox* (D. V.Hinkley, N. Reid and E.J.Snell, eds.) 265-268 Chapman and Hall, London.
- NELDER, J. A. ve WEDDERBURN, R. W. M., 1979. Generalized Linear models, *Journal Of Royal Statistical Society A* 135:370-384.
- SEBER, G.A.F., 1977. *Linear Regression Analysis*, New York: Wiley.
- SEN, A., ve SRIVASTAVA, M., 1990. *Regression Analysis Theory, Methods, and Applications*, New York: Springer.
- SHAO, J., 1993. Linear model selection by cross-validation. *Journal of the American Statistical Association*. 88:486-494.
- STONE, M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of Royal Statistical Society B*, 36:111-147.
- TIBSHIRANI, R., 1996. Regression Shrinkage and Selection via the LASSO, *Journal of Royal Statistical Society B*, 58:267–288.

- TIBSHIRANI, R., HASTIE, T., NARASIMHAN, B. ve CHU, C., 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression, *Proc. Natn. Acad. Sci. USA*, 99:6567–6572.
- TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. ve KNIGHT, K., 2005. Sparsity And Smoothness Via The Fused Lasso, *Journal of the Royal Statistical Society, Series B*, 67:91–108.
- TUSHER, V., TIBSHIRANI, R. ve CHU, C., 2001. Significance Analysis of Microarrays Applied To Transcriptional Responses To Ionizing Radiation, *Proc. Natn. Acad. Sci. USA*, 98:5116–5121.
- VARMUZA, K. ve FILZMOSER, P., 2009. Introduction to Multivariate Statistical Analysis in Chemometrics, Taylor&Francis Group, CRC Press,2009.
- WAHBA, G., 1990. Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia.
- WANG, H., LI, G. ve TSAI, C., 2007. Regression Coefficient And Autoregressive Order Shrinkage And Selection Via The Lasso, *Journal of the Royal Statistical Society, Series B*, 69:63–78.
- WEDDERBURN, R. W. M., 1974. Quasi-likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. *Biometrika* 61, 4:39-47.
- YUAN, M. ve LIN, Y., 2004. Model Selection and Estimation in Regression with Grouped Variables, Technical Report No. 1095, University of Wisconsin.
- YAZICI, B., 2001. Kategorik Veri Analizinde Eş Değişken Bulunması Durumunda Genelleştirilmiş Tahmin Denklemleri Yaklaşımı ve Bir Uygulama, Doktora Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü.
- ZEGER, S. L. and LIANG, K. Y., 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, 42:121-130.
- ZHANG, P., 1992. On the distributional properties of model selection criteria. *Journal Of the American Statistical Association*, 87:733-737.
- ZHANG,T., 2004. Statistical behavior and consistency of classification methods based on convex risk minimization,” *Ann. Statist.*, 32:469–475.
- ZHU, J. VE HASTIE, T., 2004. Classification of gene microarrays by Penalized logistic regression, *Biostatistics*, 5:427–444.

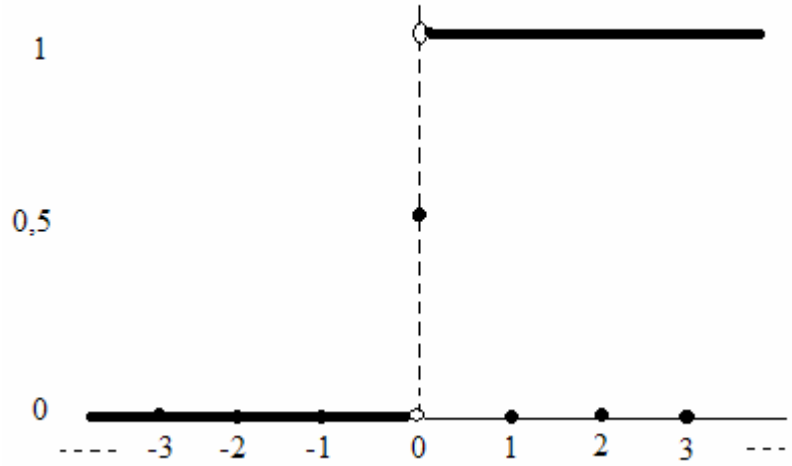
ÖZGEÇMİŞ

1986 yılında Adana’da doğdu. 2003 yılında Seyhan Hacı Ahmet Atıl (Y.D.A) Lisesini bitirdi ve aynı yıl Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik bölümünde lisans eğitime başlayarak 2007 yılında mezun oldu. 2007 yılında Çukurova Üniversitesi Fen Bilimleri Enstitüsüne bağlı İstatistik Ana Bilim dalında Yüksek Lisans eğitime başladı.

EKLER
(Tanım ve Matlab Kodları)

Heavi-side fonksiyon (Oliver Heaviside)

Heaviside adım fonksiyonu, yarı-maksimum kuralını kullanır. Yaklaşık 0 olan rasgele değişkenin kümülatif dağılım fonksiyonudur. Örneğin aşağıdaki grafik, bir adım sıçramalı heavi-side fonksiyonunu göstermektedir.



Ek Şekil 1: Heavi-Side fonksiyonu

t - lambda ilişkisi

Sabit $\lambda \geq 0$ için $t(\lambda) = \sum |\beta_j(\lambda, \gamma)|^\gamma$. Ortonormal X matrisi ile, P3 'ün p bağımsız denklemleri

$$2\beta_j - 2 \sum_i x_i y_i + \lambda \gamma |\beta_j|^{\gamma-1} \text{sign}(\beta_j) = 0$$

$\sum_i x_i y_i = \hat{\beta}_{OLS,j}$ olduğundan OLS tahmin edicinin j. koordinatıdır.

$$\hat{\beta}_j - \hat{\beta}_{ols,j} + \frac{\lambda \gamma}{2} |\hat{\beta}_j|^{\gamma-1} \text{sign}(\hat{\beta}_j) = 0$$

eşitliğini sağlar. $c_j = \hat{\beta}_{OLS,j}$ ve $s_j = \hat{\beta}_j / c_j$ olsun.

$$\frac{\lambda\gamma}{2}\sum|\hat{\beta}_j|^\gamma = \sum c_j \hat{\beta}_j - \sum \hat{\beta}_j^2$$

Böylece

$$t(\lambda) = \frac{2}{\lambda\gamma} \sum c_j^2 s_j (1 - s_j).$$

$$\frac{\frac{\lambda\gamma}{2}\sum|\hat{\beta}_j|^\gamma}{\frac{\lambda\gamma}{2}} = \frac{\sum c_j \hat{\beta}_j - \sum \hat{\beta}_j^2}{\frac{\lambda\gamma}{2}}$$

$$\sum|\hat{\beta}_j|^\gamma = \frac{2}{\lambda\gamma} \left[\frac{\sum c_j \hat{\beta}_j}{\sum c_j^2 \frac{\hat{\beta}_j}{c_j}} - \frac{\sum \hat{\beta}_j^2}{\sum c_j^2 \frac{\hat{\beta}_j^2}{c_j}} \right]$$

$$\frac{\lambda\gamma}{2}|\hat{\beta}_j|^\gamma = c_j^2 s_j - c_j^2 s_j^2$$

$$\frac{\lambda\gamma}{2}|\hat{\beta}_j|^\gamma c_j^{-2} = c_j^{-2} c_j^2 s_j - c_j^{-2} c_j^2 s_j^2$$

$$\frac{\lambda\gamma}{2}|\hat{\beta}_j|^\gamma c_j^{-2} = s_j - s_j^2 \Rightarrow s_j^2 - s_j + \frac{\lambda\gamma}{2} \underbrace{c_j^{-2} |\hat{\beta}_j|^\gamma}_{\left(\frac{\hat{\beta}_j^\gamma}{c_j^\gamma}\right) c_j^{\gamma-2} = s_j^\gamma c_j^{\gamma-2}} = s_j^2 - s_j + \frac{\lambda\gamma}{2} s_j^\gamma c_j^{\gamma-2} = 0$$

$$\frac{\frac{\lambda\gamma}{2} c_j^{-2} |\hat{\beta}_j|^\gamma}{\frac{\lambda\gamma}{2} c_j^{-2}} = \frac{s_j (1 - s_j)}{\frac{\lambda\gamma}{2} c_j^{-2}}$$

$$|\hat{\beta}_j|^\gamma = \frac{2}{\lambda\gamma} c_j^2 s_j (1 - s_j)$$

Newton – Raphson Metodu

Doğrusal olmayan denklemleri çözmek için kullanılan yöntemlerden biri de Newton-Raphson Yöntemidir. Bu yöntem ile eğer denklem bir köke yakınsıyorsa, hız bir şekilde kökün bulunması sağlanabilmektedir. Newton-Raphson metodu için genel $f(x)$ fonksiyonunu Taylor serisinde açarsak;

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 + \dots$$

Sağ taraftaki sadece iki terimi ele alırsak, yakınsama ile;

$$f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i)$$

olur. Biz $f(x) = 0$ olan bir nokta arıyoruz. Eğer $f(x_{i+1}) = 0$ ve $0 = f(x_i) + f'(x_i)(x_{i+1} - x_i)$ olduğunu kabul edersek

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Olur ve bu da Newton-Raphson metodunun serisidir.

Newton-Raphson Algoritması

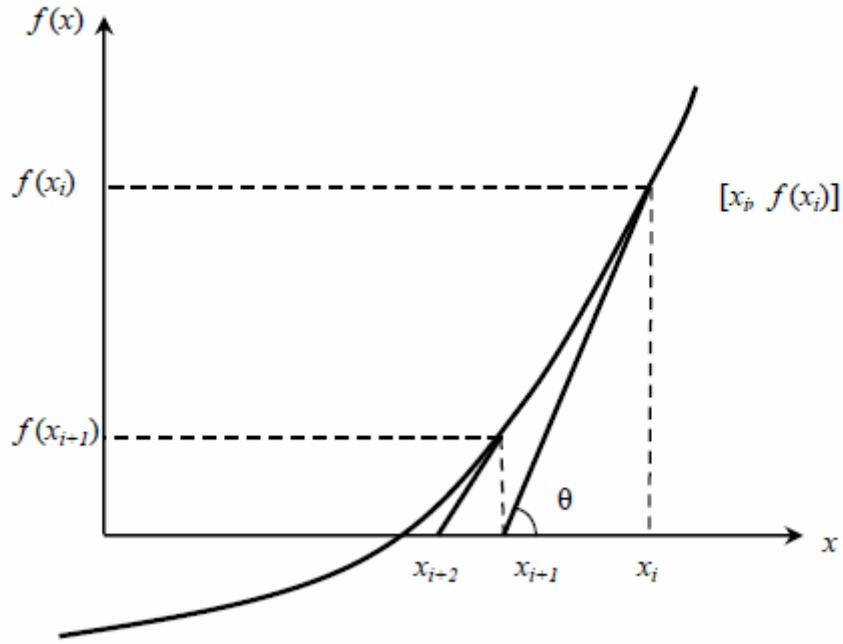
1. Sembolik olarak $f'(x)$ 'e bir değer ver.
2. x_i kökü için başlangıç değeri kullan ve

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

yeni x_{i+1} köklerini tahmin et.

3. $|\epsilon_a| = \left| \frac{x_{i+1} - x_i}{x_{i+1}} \right| \times 100$ ile mutlak göreceli yaklaşık hata $|\epsilon_a|$ 'yi bul.

4. Mutlak göreceli yaklaşık hata $|\epsilon_a|$ ile önceden belirlenen göreceli hata toleransı $|\epsilon_s|$ 'yi karşılaştır. $|\epsilon_a| > |\epsilon_s|$ ise 2. adıma dön, değilse algoritmayı durdur.



Ek Şekil 2. Newton-Raphson metodunun geometrik gösterimi