

Konveks ve konveks olmayan cezalı regresyon yöntemlerinin karşılaştırılması üzerine bir çalışma

Murat GENÇ*

Yönetim Bilişim Sistemleri Bölümü, Tarsus Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Takbaş Mahallesi Kartaltepe Sokak 33400 Tarsus, Mersin, Türkiye

Geliş Tarihi (Received Date): 19.05.2023
Kabul Tarihi (Accepted Date): 20.10.2023

Öz

Doğrusal regresyonda cezalı regresyon yöntemleri veri kümesinin yapısına bağlı olarak ön tahminde daha doğru sonuçlar elde edilmesi için kullanılır. Ayrıca cezalı regresyon yöntemleri kullanılarak yanıt değişkeni ile ilişkili olan açıklayıcı değişkenlerin tespiti mümkündür. Bu çalışmada ridge, LASSO, elastik net, uyarlamalı LASSO konveks cezalı regresyon yöntemleri ile SCAD ve MCP konveks olmayan cezalı regresyon yöntemlerinin gerçek katsayı vektörünün özelliklerine bağlı olarak performansları simülasyon çalışmaları ile karşılaştırılmıştır. Yöntemlere dayalı olarak oluşturulan modellerin ön tahmin performansının karşılaştırılması için test kümesi hata kareler ortalaması kullanılırken yöntemlerin değişken seçimindeki performanslarının karşılaştırılması için yanlış sınıflama oranı, yanlış pozitif oranı ve aktif küme büyüklükleri elde edilmiştir. Simülasyon çalışmalarına göre gerçek katsayı vektörünün yapısının konveks ve konveks olmayan cezalı regresyon yöntemleri ile oluşturulan modellerin performansı üzerinde kayda değer bir etkisinin olduğu görülmüştür.

Anahtar kelimeler: Doğrusal regresyon, LASSO, SCAD, MCP, Çoklu İç İlişki.

A study on comparison of convex and non-convex penalized regression methods

Abstract

In linear regression, penalized regression methods are used to obtain more accurate predictions depending on the structure of the data set. In addition, it is possible to determine the explanatory variables associated with the response variable by using penalized regression methods. In this study, the performances of ridge, LASSO, elastic net, adaptive LASSO convex penalized regression methods and SCAD and MCP non-convex penalized regression methods are compared depending on the properties of the true coefficient vector based on simulation studies. While the mean squared error on the

*Murat GENÇ, muratgenc@tarsus.edu.tr, <https://orcid.org/0000-0002-6335-3044>

test set is used to compare the prediction performance of the models based on the methods, the false classification rate, false positive rate and active set sizes are obtained to compare the performance of the methods in variable selection. According to the simulation studies, it has been observed that the structure of the true coefficient vector has a remarkable effect on the performance of the models created by convex and non-convex penalized regression methods.

Keywords: Linear Regression, LASSO, SCAD, MCP, Multicollinearity.

1. Giriş

Doğrusal regresyon modelleri, açıklayıcı değişkenler ile yanıt değişkeni arasındaki ilişkiyi modellemede yaygın olarak kullanılan istatistik modellerdir. Doğrusal regresyon modelleri birçok bilim alanında değişkenler arasındaki ilişkiyi modellemede kullanılmaktadır. Doğrusal regresyon modeli oluşturmanın ilk aşaması model katsayılarının tahmin edilmesidir.

$n \times 1$ boyutlu yanıt değişken değerlerinin vektörü \mathbf{y} , $n \times p$ boyutlu açıklayıcı değişken değerlerinin matrisi \mathbf{X} , $p \times 1$ boyutlu bilinmeyen katsayıların vektörü $\boldsymbol{\beta}$, ve $n \times 1$ boyutlu 0 ortalamalı ve $\sigma^2 \mathbf{I}$ varyanslı hata terimleri vektörü $\boldsymbol{\varepsilon}$ olmak üzere bir doğrusal regresyon modeli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

olarak tanımlanır. Doğrusal regresyonda katsayı tahmini için kullanılan en temel yöntem en küçük kareler (EKK) yöntemidir. Denklem (1)'deki katsayıların EKK tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{ekk} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2)$$

olarak verilir. EKK ile regresyon katsayılarının tahmini yaygın olarak yapılmaktadır. Ancak veri kümesinde açıklayıcı değişkenler arasında bulunan yüksek düzeyde doğrusal bağımlılık olarak tanımlanan çoklu iç ilişkinin olması halinde EKK tatmin edici olmayan katsayı değerleri verir. EKK'nın çoklu iç ilişki durumunda ortaya çıkan dezavantajlarını ortadan kaldırmak için ridge [1], kısıtlı en küçük kareler [2] ve kısıtlı ridge tahmin edici [3] gibi regresyon katsayıları üzerinde çeşitli kısıtlar kullanan tahmin ediciler önerilmiştir. Ayrıca EKK yöntemi değişken seçimi yapamadığı için yanıt değişken üzerinde etkisi fazla olan değişkenlerin seçilmesinde kullanılamaz. Yanıt değişkeni etkileyen değişkenlerin tespiti için çeşitli yöntemler önerilmiştir. En iyi alt küme seçim yönteminde [4], açıklayıcı değişkenlerin tüm kombinasyonları kullanılarak oluşturulan modeller göz önüne alarak değişken seçimi yapılır. En iyi alt küme seçim yöntemi, hesaplama bakımından maliyetli bir yöntem olması nedeniyle, bu yöntemden farklı olarak modelin katsayı tahmini ile değişken seçiminin eşanlı olarak yapıldığı negatif olmayan garrote [5], bridge regresyon [6], LASSO (least absolute shrinkage and selection operator) [7], elastik net [8] ve uyarlanabilir (adaptive) LASSO [9] gibi konveks sürekli bir ceza teriminin kullanıldığı yöntemlerin yanı sıra SCAD (smoothly clipped absolute deviation) ve MCP (minimax concave penalty) [10] gibi konveks olmayan sürekli bir ceza teriminin kullanıldığı yöntemler de bulunmaktadır. Bu yöntemlerden bridge regresyon, $\|\boldsymbol{\beta}\|_1^q$, $\theta > 0$ ceza fonksiyonuna sahip bir cezalı regresyon yöntemidir. Bridge

regresyonun ceza fonksiyonunun $\theta = 1$ durumu konveks bir ceza fonksiyonu üretir ve LASSO yöntemine karşılık gelir. Elastik net, LASSO'da ceza terimine ℓ_2 norma dayalı bir terim eklenmesi ile elde edilir ve LASSO'ya göre daha esnek bir yöntemdir. Uyarlanabilir LASSO, uyarlanabilir ağırlıklar kullanılarak LASSO tahminlerinin güncellenmesine dayalı iki aşamalı bir yöntemdir. SCAD ve MCP ise parçalı karesel ceza terimleri kullanan konveks olmayan cezalı regresyon yöntemleridir.

Açıklayıcı değişkenlerin gözlem matrisinin boyutlarına göre veri kümesi, klasik veya yüksek boyutlu olarak adlandırılır. Gözlem sayısı değişken sayısından büyük olan veri kümeleri klasik veri kümesi olarak adlandırılırken, aksi halde yüksek boyutlu veri kümesi olarak adlandırılır. Literatürde her iki veri yapısına dayalı olarak yapılan çeşitli çalışmalar bulunmaktadır [11, 12]. Ayrıca konveks ve konveks olmayan farklı doğrusal regresyon yöntemlerinin karşılaştırılması amaçlı çeşitli çalışmalar bulunmaktadır. [13], Yüksek boyutlu veri analizi için temel bileşenler regresyonda bootstrap yöntemini kullanarak değişken seçimi için bir simülasyon çalışması gerçekleştirmiştir. [14], cezalı regresyon yöntemlerinin yüksek boyutlu veri kümelerinde performansı üzerine bir çalışma yapmıştır. [15], klasik veri kümelerinde gerçek katsayı vektörünün yapısına bağlı olarak konveks cezalı regresyon yöntemlerinin karşılaştırılması üzerine bir çalışma yapmıştır. [16], post LASSO yönteminin ikinci aşamasındaki tahmin edicinin model performansı üzerindeki etkisini bir simülasyon çalışması ile incelemiştir.

Bu çalışmada, klasik veri kümelerinde gerçek katsayı vektörünün yapısı, sıfırdan farklı regresyon katsayılarının miktarı ve konumuna bağlı olarak konveks ve konveks olmayan cezalı regresyon yöntemlerinin performans incelemesi ele alınmıştır. Bu amaçla konveks ceza fonksiyonlu ridge, LASSO, elastik net ve uyarlanabilir LASSO yöntemleri ile konveks olmayan ceza fonksiyonlu SCAD ve MCP yöntemleri kullanılmıştır.

Çalışmanın ikinci bölümünde karşılaştırmaya tabi tutulan cezalı regresyon yöntemleri özetlenmiş ve yöntemlerin karşılaştırılmasında kullanılan kriterler verilmiştir. Ayrıca karşılaştırmada faydalanılan simülasyon çalışmalarının özelliklerine değinilmiştir. Çalışmanın üçüncü bölümünde, gerçekleştirilen simülasyon çalışmalarının bulguları verilmiş ve sonuçlar detaylı bir şekilde irdelenmiştir. Çalışmanın dördüncü bölümünde sonuçlar verilmiş ve çalışma tamamlanmıştır.

2. Materyal ve metod

2.1. Cezalı regresyon yöntemleri

Veri kümesinin yapısına bağlı olarak EKK yönteminin yeni gözlemlerin tahmininde yetersiz olması ve değişken seçimi yapamaması nedeniyle, EKK'ya alternatif olarak model katsayılarının belirli kısıtlar altında hesaplanmasına dayalı çeşitli yöntemler önerilmiştir. Bu yöntemler, cezalı regresyon yöntemleri olarak adlandırılır ve yaygın bir kullanım alanına sahiptir [17]. Bu yöntemler, kararlı regresyon katsayı tahminleri elde etmek için aşırı büyük katsayıların büzülmesi ilkesine dayanır. Ayrıca söz konusu yöntemlerden bazıları bir kısım regresyon katsayı tahminlerini sıfır olarak belirleyebilir ve böylece otomatik değişken seçimi yapabilir.

\mathbf{x}_i $p \times 1$ boyutlu i . gözlem vektörü, y_i , i . gözlem yanıt değeri olmak üzere; $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ bağımsız ve aynı dağılımlı veri kümesi verilsin. Denklem (1)'de verilen doğrusal regresyon modeli için amaç fonksiyonu

$$Q(\boldsymbol{\beta}) = \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + p_\lambda(\boldsymbol{\beta}) \quad (3)$$

olarak verilir. Burada $\boldsymbol{\beta}$ tahmin edilecek katsayılar vektörü, $p_\lambda(\cdot)$ ceza fonksiyonu ve λ ayar parametresidir. Cezalı regresyon yöntemlerinde katsayı tahminleri Denklem (3)'te verilen amaç fonksiyonunun minimize edilmesi ile elde edilir.

Denklem (2)'deki EKK tahmin edicisi tüm yansız tahmin ediciler sınıfı içinde en küçük varyansa sahip olan tahmin edicidir. Ancak, veri kümesinde çoklu iç ilişki olması halinde EKK tahminleri kararlı olmaktan uzaktır. Bu problemin üstesinden gelenebilmesi için literatürde katsayı tahminlerine çeşitli şekillerde etki eden ceza fonksiyonları önerilmiştir.

Cezalı regresyon yöntemleri, ceza fonksiyonunun konveks veya konveks olmamasına göre sırasıyla konveks cezalı regresyon yöntemleri ve konveks olmayan cezalı regresyon yöntemleri olarak adlandırılır. Literatürde veri kümelerinin özelliklerine bağlı olarak performansı değişen konveks ve konveks olmayan cezalı regresyon yöntemleri önerilmiştir. Alt Bölüm 2.1.1 ve Alt Bölüm 2.1.2'de söz konusu yöntemler özetlenmiştir.

2.1.1. Konveks cezalı regresyon yöntemleri

Literatürde regresyon katsayılarını büzerek EKK'ya göre daha yüksek ön tahmin doğruluğuna ulaşılmasını sağlayan çeşitli konveks cezalı regresyon yöntemleri bulunmaktadır. Bu yöntemlerden bazıları ceza fonksiyonunun yapısına bağlı olarak otomatik değişken seçimi de yapabilmektedir.

[1] yanlılık ve varyans arasında bir ödünleşmeye dayalı bir yöntem olarak ridge regresyonu önermiştir. Ridge regresyonda, model katsayılarının yansızlığından ödün verilerek, varyansın aşırı şişmesi probleminin aşılması hedeflenir. Ridge regresyonda katsayı tahminlerine

$$\hat{\boldsymbol{\beta}}_R = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \right\} \quad (4)$$

probleminin çözülmesi ile ulaşılır. Denklem (4)'te $\lambda > 0$, ayar parametresinin değeri büyüdükçe katsayıların büzülme miktarı artar. \mathbf{I}_p , $p \times p$ boyutlu birim matris olmak üzere Denklem (4)'te verilen problemin çözülmesi ile ridge regresyon katsayı tahminleri

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}^T \mathbf{X} + \lambda n \mathbf{I}_p)^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

olarak bulunur. Ridge regresyon yönteminde yanlılık ile varyans arasındaki ödünleşmeden faydalanılarak, EKK'ya göre daha doğru ön tahmin değerleri bulunur. Ridge regresyonun EKK'ya olan bu üstünlüğüne rağmen, otomatik değişken seçimi yapamaması nedeniyle ridge regresyona alternatif tahmin ediciler önerilmiştir.

Ridge regresyonun değişken seçimi konusundaki eksikliğini göz önüne alarak [7], LASSO yöntemini önermiştir. LASSO yönteminde katsayı tahminleri

$$\hat{\beta}_L = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (6)$$

probleminin çözülmesi ile elde edilir. LASSO’da Denklem (6)’da bulunan ℓ_1 norma dayalı $\|\beta\|_1$ terimi sayesinde ayar parametresinin değeri yeterince büyük olduğunda LASSO katsayı tahminlerinin bazıları sıfır olur. Bu nedenle LASSO bir katsayı büzme yöntemi olmasının yanı sıra bir otomatik değişken seçim yöntemidir. Denklem (6)’daki problemin kapalı formda bir çözümü yoktur. Bu nedenle LASSO tahminlerinin elde edilmesi için çeşitli algoritmalar önerilmiştir. En küçük açı regresyonu (least angle regression – LARS) [18], değişen yönlü çarpanlar yöntemi (alternating direction method of multipliers – ADMM) [19], koordinat iniş (coordinate descent) [20] algoritmaları LASSO tahminlerinin elde edilmesinde kullanılabilen bazı algoritmalarıdır.

Açıklayıcı değişkenler arasında yüksek korelasyon olması durumunda, LASSO regresyon tahmin edici tahminler veremez [7]. Bu durumda LASSO’ya alternatif olarak kullanılacak bir yöntem elastik net (ENET) [8] yöntemidir. ENET, LASSO’nun ceza fonksiyonuna bir ℓ_2 norm teriminin eklenmesi ile elde edilir. ENET tahmin edicisi

$$\hat{\beta}_E = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2) \right\} \quad (7)$$

probleminin çözülmesi ile bulunur. Burada $0 \leq \alpha \leq 1$ ikinci ayar parametresidir. Ridge ve LASSO, ENET’te yeni ayar parametresinin sırasıyla 0 ve 1 değerlerine karşılık gelir. Ayrıca ENET’in yüksek korelasyonlu değişkenleri gruplama özelliği bulunmaktadır. Dolayısıyla veri kümesinde yüksek korelasyonlu değişkenlerin varlığında ENET’in LASSO’ya göre daha iyi bir performans göstermesi beklenir.

LASSO’nun değişken seçimi konusunda tutarlı olmadığı durumlarda, [9] tarafından önerilen uyarlanabilir LASSO (U-LASSO) model katsayılarının tahmininde kullanılabilir. U-LASSO değişken seçiminde tutarlı bir yöntemdir ve yaklaşık minimax optimaldir [9]. U-LASSO, LASSO’nun ceza fonksiyonunda uyarlanabilir ağırlıklar kullanılması ilkesine dayanır. U-LASSO’nun katsayı tahminleri

$$\hat{\beta}_{U-LASSO} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \sum_{j=1}^p w_j |\beta_j| \right\} \quad (8)$$

problemin çözülmesi ile elde edilir. Denklem (8)’de w_j , $i = 1, 2, \dots, p$ uyarlanabilir ağırlıklardır. [9], uyarlanabilir ağırlıklar vektörünü $\hat{\mathbf{w}} = 1/|\hat{\beta}_{ekk}|$ olarak seçmiştir. Buna göre U-LASSO’nun hesaplanabilmesi için iki aşama söz konusudur:

1. EKK katsayı tahminlerinin ve bu tahminlerden faydalanılarak uyarlanabilir ağırlıklar vektörünün hesaplanması.
2. Birinci aşamadaki ağırlıklar ile LASSO cezalandırma fonksiyonu yeniden ağırlıklandırılarak Denklem (8)’deki problemin çözülmesi.

2.1.2 Konveks olmayan cezalı regresyon yöntemleri

LASSO yöntemi hesaplama kolaylığı sağlayan algoritmalar ve ön tahmindeki yüksek doğruluk değeri nedeniyle ön plana çıkmaktadır. Ancak LASSO yönteminin değişken seçiminde tutarlı sonuçlar vermesi için tasarım matrisinin oldukça sınırlayıcı bazı şartları sağlaması gerekir [21].

LASSO yöntemi ile değişken seçiminde ortaya çıkan problemin aşılabilmesi için [10] tarafından konveks olmayan bir cezalı regresyon yöntemi olarak SCAD önerilmiştir. SCAD, yaklaşık olarak bir olasılıkla gerçekte sıfır olmayan katsayıları, modelin gerçek seyreklik (sparsite) düzeyi önceden biliniyormuş gibi tahmin edebilir (oracle özelliği). SCAD yönteminin ceza fonksiyonu

$$p_{\lambda,\gamma}^{SCAD}(\beta_j) = \begin{cases} \lambda\beta_j, & \beta_j \leq \lambda \\ \frac{\gamma\lambda\beta_j - 0.5(\beta_j^2 + \lambda^2)}{\gamma - 1}, & \lambda < \beta_j \leq \gamma\lambda \\ \frac{\lambda^2(\gamma^2 - 1)}{2(\gamma - 1)}, & \beta_j > \gamma\lambda \end{cases} \quad (9)$$

olarak tanımlanır. Denklem (9)'da verilen ceza fonksiyonu Denklem (3)'te verilen amaç fonksiyonuna uygulanıp amaç fonksiyonu minimize edilerek SCAD yönteminin katsayı tahminlerine

$$\hat{\beta}_{SCAD} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \sum_{j=1}^p p_{\lambda,\gamma}^{SCAD}(\beta_j) \right\} \quad (10)$$

problemi çözümlenerek ulaşılır.

[21], değişken seçiminde farklı bir cezalı yöntem olarak MCP yöntemini önermiş ve MCP yönteminin uygun ayar parametresi değerleri için değişken seçiminde tutarlı olduğunu göstermiştir. MCP yönteminin ceza fonksiyonu

$$p_{\lambda,\gamma}^{MCP}(\beta_j) = \begin{cases} \lambda\beta_j - \frac{\beta_j^2}{2\gamma}, & \beta_j \leq \gamma\lambda \\ 0.5\gamma\lambda^2, & \beta_j > \gamma\lambda \end{cases} \quad (11)$$

olarak tanımlanır. Denklem (11)'de verilen ceza fonksiyonu, Denklem (3)'te verilen amaç fonksiyonuna yazılıp, amaç fonksiyonu minimize edilerek MCP yönteminin katsayı tahminlerine

$$\hat{\beta}_{MCP} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \sum_{j=1}^p p_{\lambda,\gamma}^{MCP}(\beta_j) \right\} \quad (12)$$

probleminin çözülmesi ile ulaşılır. SCAD ve MCP yöntemlerinin Denklem (9) ve Denklem (11)'de verilen ceza fonksiyonlarına göre bu yöntemlerin model katsayılarını büzme şekillerinde farklılıklar bulunmakta olup her iki yöntemin de LASSO yöntemine göre daha esnek tasarım matrisi koşullarında ve yönteme bağlı olarak değişen bir

cezalandırma düzeyinde tutarlı değişken seçimi yapabildiği gösterilmiştir [10, 21]. Konveks olmayan ceza fonksiyonu nedeniyle bu yöntemlerde model oluşturmak, konveks ceza fonksiyonlu yöntemlerdekinden daha zordur. [22], ceza fonksiyonuna yerel bir doğrusal yaklaşım (local linear approximation – LLA) uygulayarak konveks olmayan ceza fonksiyonlu yöntemler için LARS algoritmasına dayalı bir optimizasyon yöntemi önermiştir. [23] SCAD ve MCP yöntemlerinin katsayı tahminlerinin elde edilmesi için bir koordinat iniş algoritması önermiştir.

2.3. Simülasyon çalışmaları ve karşılaştırma kriterleri

Bu çalışmada konveks cezalı regresyon yöntemleri ridge, LASSO, elastik net ve uyarlanabilir LASSO ile konveks olmayan cezalı regresyon yöntemleri SCAD ve MCP simülasyon çalışmaları ile karşılaştırmaya tabi tutulmuştur. Simülasyon çalışmalarında kullanılan veri kümeleri Denklem (1)'de verilen doğrusal regresyon modeline göre [7]'de açıklanan yöntemeye dayalı olarak oluşturulmuştur. Her bir simülasyon çalışmasında açıklayıcı değişken sayısı 40 olup tekrar sayısı 100 olarak belirlenmiştir. Simülasyon çalışmalarında, i . ve j . açıklayıcı değişkenler arasındaki korelasyon $\rho^{|i-j|}$ olup sırasıyla orta ve yüksek düzeyde korelasyon düzeyinin etkisini ölçmek için ρ 'nun iki değeri 0.7 ve 0.9 olarak alınmıştır. Gerçek katsayılar vektörünün, yöntemlerin performansı üzerindeki etkisinin incelenmesi için simülasyon çalışmaları iki grup altında toplanmıştır.

- S1. Birinci gruptaki simülasyon çalışmalarında gerçek katsayılar vektörünün aktif (sıfırdan farklı katsayıya sahip) değişkenlere karşılık gelen katsayıları aktif olmayan (sıfır katsayılı) değişkenlere karşılık gelen katsayılarından önce gelmektedir. Bu grupta aktif katsayı değerleri, ilk terimi 0.5 ve ortak farkı 0.5 olan bir aritmetik dizi oluşturmaktadır. Daha açık bir ifadeyle gerçek katsayı vektörü

$$\beta = \left[\underbrace{0.5, 1, \dots, s/2}_s, \underbrace{0, 0, \dots, 0}_{40-s} \right] \text{ formuna sahiptir. Farklı seyreklik düzeylerini}$$

temsil etmek üzere; aktif değişken sayısı olarak, $s = 10, 20, 30, 40$ durumları alınmıştır. s değeri büyüdükçe seyreklik düzeyi azalmaktadır.

- S2. İkinci gruptaki simülasyon çalışmalarında aktif değişkenlerin katsayılarının büyüklüğü eşit alınmış, aktif olmayan değişkenlerin sayısı ve konumu değiştirilerek seyreklik ve konumun etkisinin incelenmesi hedeflenmiştir. Bu gruptaki simülasyon çalışmalarında, gerçek katsayı vektörü $\beta = [a, \mathbf{0}_k, a, \mathbf{0}_k, \dots, a, \mathbf{0}_k]$ formuna sahiptir. Burada a sinyal değerini (aktif değişken katsayısının büyüklüğü), k ise sıfır vektörünün boyutunu göstermektedir. k 'nın değeri gerçek katsayı vektörünün seyreklik düzeyini temsil etmektedir. Bu çalışmada $a = 1, 3, 9$ ve $k = 1, 4, 9$ değerleri göz önüne alınmıştır.

Ayar parametresi tahmini ve performans kriterlerinin hesaplanması için veri kümesinin eğitim değerlendirme ve test kümelerine ayrıştırılması yaygındır [8, 24]. Öncelikle karşılaştırılacak cezalı regresyon yöntemleri seçilen ayar parametresi değerlerinde eğitime tabi tutulur. Bu modeller için hata kareler ortalaması değerlendirme kümesi üzerinde hesaplanır. Değerlendirme kümesinde en küçük hata kareler ortalamasına sahip model en iyi model olarak belirlenir. Oluşturulan modellerin performansı ise test kümesi üzerinde hata kareler ortalaması hesaplanarak tespit edilir.

Bu çalışmada, her bir simülasyon çalışması için bağımsız olarak 100 birimlik eğitim, 100 birimlik değerlendirme ve 400 birimlik test kümesi oluşturulmuştur. Denklem (1)'deki doğrusal regresyon modelinde hataların standart sapması [7]'de olduğu gibi $\sigma = 3$

alınmıştır. Elastik netin ceza fonksiyonu ridge ve LASSO yöntemlerinin ceza fonksiyonlarının konveks bir bileşimi olup burada elastik netin ridge ve LASSO yöntemlerinden farklılık düzeyini gözlemlemek için $\alpha = 0.5$ seçilmiştir. Ayrıca SCAD ve MCP'nin ikinci ayar parametresinin değeri R'nin `ncvreg` [23] paketinde varsayılan değerler olan $\gamma_{SCAD} = 3.7$ ve $\gamma_{MCP} = 3$ olarak seçilmiştir.

Cezalı regresyon yöntemleri ile oluşturulan modellerin performanslarının karşılaştırılmasında çeşitli kriterler kullanılmıştır. Σ , açıklayıcı değişkenlerin kovaryans matrisi ve $\tilde{\beta}$, ilgili cezalı regresyon yönteminin katsayılar vektörünü temsil etmek üzere, test kümesi üzerinde test kümesi hata kareler ortalaması

$$TMSE = (\beta - \tilde{\beta})^T \Sigma (\beta - \tilde{\beta})$$

modellerin ön tahmindeki performanslarını karşılaştırmak için kullanılmıştır. Her bir yönteme dayalı olarak tahmin edilen aktif kümeler elde edilmiştir. Buna dayalı olarak her bir cezalı regresyon yöntemi tarafından hatalı bir şekilde aktif kümeye alınan veya aktif kümeye alınmayan açıklayıcı değişkenlerin oranı olarak tanımlanan yanlış sınıflama oranı ve gerçekte aktif kümenin dışında olmasına rağmen aktif kümede tahmin edilen açıklayıcı değişkenlerin oranı olarak tanımlanan yanlış pozitif oranı yöntemlerin değişken seçimindeki performansının bir göstergesi olarak verilmiştir. Son olarak her bir yönteme dayalı olarak elde edilen aktif küme büyüklükleri rapor edilmiştir.

Simülasyon çalışmalarındaki hesaplamalar için R yazılımında bulunan `glmnet` [20] ve `ncvreg` [23] paketlerinden faydalanılmıştır.

3. Bulgular ve tartışma

S1 grubu simülasyon çalışmalarından elde edilen sonuçlar Tablo 1'de verilmiştir. Tablo 1'e göre korelasyon miktarı orta düzeyde iken değişken seçimi yapabilen yöntemler tüm seyreklik değerlerinde, ridge regresyona göre daima daha iyi bir TMSE değeri vermiştir. Korelasyon miktarı yüksek olduğunda U-LASSO, MCP ve SCAD yöntemlerinde bu sonucun aksi bir durum söz konusu iken LASSO ve ENET ($t = 10, 20, 30$), ridge yöntemine göre daha iyi bir TMSE değeri vermiştir. Seyrek olmayan değerde ($t = 40$) yüksek korelasyonlu durumda ENET ve ridge yöntemlerinin TMSE performansı diğer yöntemlerden daha yüksektir. ENET içerdiği ℓ_2 norma dayalı terim nedeniyle diğer değişken seçimi yapabilen yöntemlere göre yüksek korelasyondan daha az etkilenmiştir. Ayrıca bu gruptaki simülasyon çalışmalarında genel olarak konveks cezalı regresyon yöntemleri, konveks olmayan cezalı regresyon yöntemlerine göre daha iyi bir TMSE değeri üretmiştir.

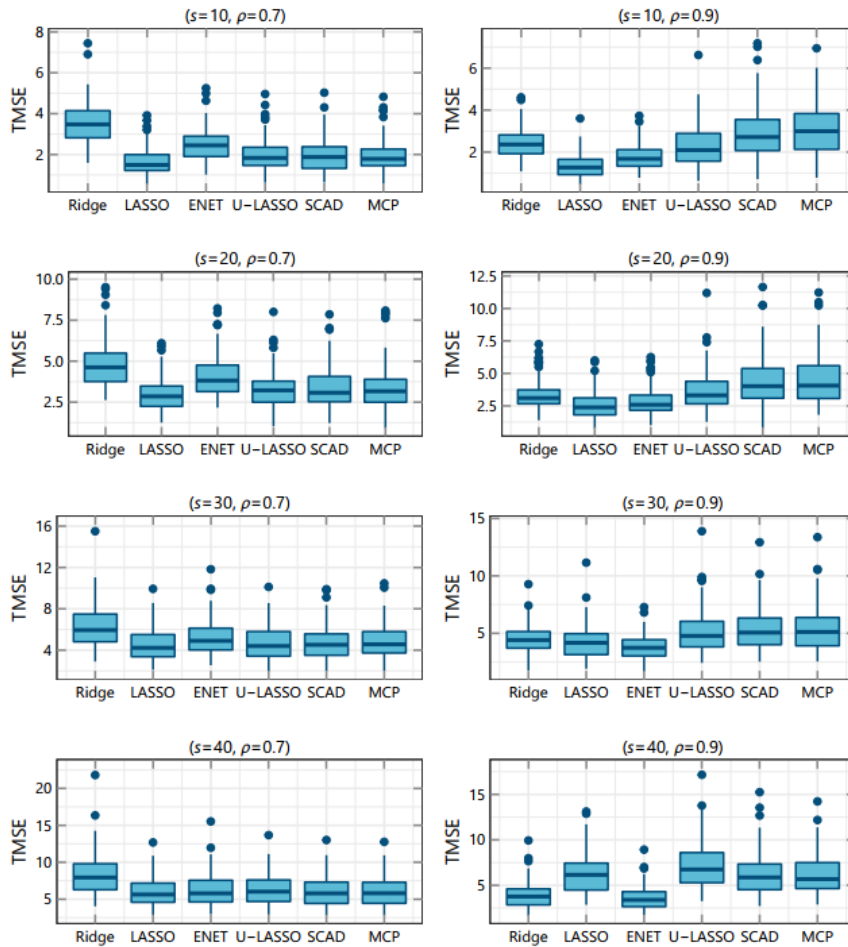
S1 grubundaki simülasyon çalışmalarında $t = 40$ durumunda hiçbir gerçek katsayı değeri sıfır olmadığından dolayı bu değere karşılık gelen yanlış sınıflama oranı ve yanlış pozitif oranı rapor edilmemiştir. Bu ölçüler bakımından korelasyon orta düzeyde iken konveks olmayan ceza terimli yöntemlerin daha iyi olduğu söylenebilir. Yüksek korelasyon düzeyinde ise genel olarak LASSO ve U-LASSO konveks cezalı regresyon yöntemleri daha iyi bir sonuç vermiştir. Ayrıca aktif küme büyüklüğü açısından t 'nin küçük değerlerinde konveks olmayan ceza terimli yöntemlerin gerçek seyreklik düzeyine daha yakın bir model verdiği ifade edilebilir.

Tablo 1. S1 grubu simülasyon çalışmalarında tahmin edicilerin kalite ölçüleri.

s	ρ	Yöntem	Medyan TMSE	Standart Sapma	Yanlış Sınıflama Oranı	Yanlış Pozitif Oranı	Aktif Küme Büyüklüğü
10	0.7	Ridge	3.4754	0.12	0.75	1.00	40
		LASSO	1.4887	0.07	0.18	0.23	16
		ENET	2.4481	0.09	0.50	0.67	30
		U-LASSO	1.8351	0.07	0.11	0.10	12
		SCAD	1.8739	0.16	0.15	0.17	14
		MCP	1.7874	0.06	0.10	0.07	11
10	0.9	Ridge	2.3649	0.07	0.75	1.00	40
		LASSO	1.2689	0.08	0.13	0.15	14
		ENET	1.6822	0.06	0.43	0.57	26.5
		U-LASSO	2.0993	0.11	0.19	0.20	14
		SCAD	2.7183	0.14	0.18	0.13	11
		MCP	2.9990	0.12	0.15	0.10	11
20	0.7	Ridge	4.6240	0.21	0.50	1.00	40
		LASSO	2.8527	0.08	0.15	0.30	26
		ENET	3.8164	0.2	0.43	0.85	37
		U-LASSO	3.2217	0.12	0.13	0.20	23
		SCAD	3.0752	0.16	0.13	0.20	23
		MCP	3.1644	0.14	0.10	0.10	21
20	0.9	Ridge	3.1012	0.1	0.50	1.00	40
		LASSO	2.3856	0.13	0.11	0.20	23
		ENET	2.5858	0.1	0.38	0.75	35
		U-LASSO	3.3111	0.13	0.18	0.30	24
		SCAD	4.0199	0.27	0.18	0.25	23
		MCP	4.0599	0.15	0.15	0.20	22
30	0.7	Ridge	5.9456	0.28	0.25	1.00	40
		LASSO	4.2287	0.16	0.13	0.45	34
		ENET	4.9067	0.24	0.23	0.90	39
		U-LASSO	4.4028	0.16	0.10	0.30	32
		SCAD	4.5254	0.21	0.08	0.20	31
		MCP	4.5683	0.2	0.08	0.10	30.5
30	0.9	Ridge	4.4145	0.11	0.25	1.00	40
		LASSO	4.1680	0.21	0.13	0.40	33
		ENET	3.7363	0.14	0.23	0.90	39
		U-LASSO	4.7749	0.25	0.13	0.40	33
		SCAD	5.0621	0.27	0.15	0.50	34
		MCP	5.1185	0.29	0.15	0.50	34
40	0.7	Ridge	7.9400	0.32	-	-	40
		LASSO	5.6404	0.23	-	-	40
		ENET	5.7942	0.33	-	-	40
		U-LASSO	6.0300	0.33	-	-	40
		SCAD	5.7938	0.31	-	-	40
		MCP	5.8115	0.3	-	-	40

40	0.9	Ridge	3.7707	0.17	-	-	40
		LASSO	6.1566	0.36	-	-	40
		ENET	3.4036	0.26	-	-	40
		U-LASSO	6.7458	0.31	-	-	39
		SCAD	5.8725	0.41	-	-	39
		MCP	5.6913	0.36	-	-	39

Şekil 1, S1 grubundaki simülasyon çalışmalarında Ridge, LASSO, ENET, U-LASSO, MCP ve SCAD cezalı regresyon yöntemleri ile elde edilen TMSE değerlerinin kutu grafiklerini göstermektedir. Kutu grafikleri, Tablo 1'e dayalı olarak verilen S1 grubu simülasyon çalışmalarının çıkarımlarını desteklemektedir.



Şekil 1. S1 grubu simülasyon çalışmalarında TMSE değerlerinin kutu grafikleri

S2 grubu simülasyon çalışmalarından elde edilen sonuçlar Tablo 2,3,4'te verilmiştir. Tablo 2,3,4'e göre tüm seyreklik düzeylerinde sinyallerin küçük değerleri için konveks cezalı regresyon yöntemlerinin daha iyi bir TMSE değerine sahip olduğu görülmektedir. Sinyal değerleri büyüdükçe, konveks olmayan cezalı regresyon yöntemlerinin performansı kayda değer ölçüde artmaktadır. Ayrıca büyük sinyal durumunda ridge regresyon dışında hem konveks hem de konveks olmayan cezalı regresyon yöntemlerinin yüksek korelasyon miktarından olumsuz etkilendiği söylenebilir.

Yanlış sınıflama oranı, yanlış pozitif oranı ve aktif küme büyüklüğü bakımından S2 grubundaki simülasyon çalışmalarında konveks olmayan cezalı regresyon yöntemlerinin

daha iyi bir performans gösterdiği görülmektedir. Sinyal değerleri büyük olduğunda, konveks olmayan cezalı regresyon yöntemleri, gerçek katsayı vektöründe bulunan orijinal değişkenleri doğru bir şekilde tahmin edebilmektedir. Değişken seçimi performansı bakımından bu yöntemleri konveks cezalı regresyon yöntemlerinden U-LASSO yöntemi takip etmektedir. LASSO ve ENET'in değişken seçiminde bu yöntemlere göre daha kötü bir sonuç verdiği söylenebilir.

Tablo 2. S2 grubu simülasyon çalışmalarında tahmin edicilerin kalite ölçüleri ($k = 1$).

α	ρ	Yöntem	Medyan TMSE	Standart Sapma	Yanlış Sınıflama Oranı	Yanlış Pozitif Oranı	Aktif Küme Büyüklüğü
1	0.7	Ridge	2.6575	0.09	0.50	1.00	40
		LASSO	3.4971	0.16	0.33	0.55	30
		ENET	2.9179	0.13	0.40	0.75	35
		U-LASSO	4.6678	0.22	0.30	0.40	25
		SCAD	5.2912	0.24	0.35	0.55	28
		MCP	5.2358	0.31	0.34	0.45	26
1	0.9	Ridge	1.3135	0.03	0.50	1.00	40
		LASSO	2.3949	0.06	0.35	0.50	25
		ENET	1.6386	0.06	0.45	0.85	36
		U-LASSO	3.3703	0.12	0.35	0.40	20
		SCAD	4.1812	0.18	0.40	0.30	16
		MCP	3.9923	0.11	0.40	0.25	15
3	0.7	Ridge	5.2265	0.15	0.50	1.00	40
		LASSO	3.8831	0.16	0.30	0.60	32
		ENET	4.7422	0.15	0.45	0.90	38
		U-LASSO	3.0267	0.25	0.10	0.20	24
		SCAD	2.3700	0.09	0.03	0.05	21
		MCP	2.4014	0.12	0.00	0.00	20
3	0.9	Ridge	3.9098	0.12	0.50	1.00	40
		LASSO	3.5183	0.08	0.29	0.55	31
		ENET	3.7220	0.11	0.45	0.90	38
		U-LASSO	4.1123	0.22	0.20	0.40	27
		SCAD	4.8262	0.20	0.23	0.40	27
		MCP	4.8214	0.28	0.23	0.40	26
9	0.7	Ridge	15.0754	0.46	0.50	1.00	40
		LASSO	3.8879	0.17	0.30	0.60	32
		ENET	8.3847	0.42	0.48	0.95	39
		U-LASSO	2.2572	0.15	0.03	0.05	21
		SCAD	2.1574	0.10	0.00	0.00	20
		MCP	2.1203	0.11	0.00	0.00	20
9	0.9	Ridge	14.7462	0.34	0.50	1.00	40
		LASSO	3.5751	0.08	0.28	0.55	31
		ENET	9.2747	0.31	0.49	0.98	39.5
		U-LASSO	2.4309	0.11	0.03	0.05	21
		SCAD	2.4714	0.09	0.03	0.05	21
		MCP	2.7945	0.21	0.03	0.05	21

Tablo 3. S2 grubu simülasyon çalışmalarında tahmin edicilerin kalite ölçüleri ($k = 4$).

α	ρ	Yöntem	Medyan TMSE	Standart Sapma	Yanlış Sınıflama Oranı	Yanlış Pozitif Oranı	Aktif Küme Büyüklüğü
1	0.7	Ridge	2.1533	0.07	0.80	1.00	40
		LASSO	1.8979	0.14	0.33	0.38	19
		ENET	1.9331	0.1	0.44	0.53	25
		U-LASSO	2.3974	0.13	0.25	0.25	15
		SCAD	2.7359	0.13	0.20	0.19	12
		MCP	2.7313	0.19	0.15	0.13	11
		1	0.9	Ridge	1.1389	0.05	0.80
LASSO	1.4123			0.04	0.30	0.31	15
ENET	1.1998			0.04	0.48	0.56	25
U-LASSO	1.8746			0.05	0.28	0.22	11
SCAD	2.4490			0.09	0.23	0.16	8
MCP	2.4863			0.11	0.23	0.13	7
3	0.7			Ridge	4.2020	0.13	0.80
		LASSO	1.9720	0.12	0.33	0.41	21
		ENET	3.3934	0.12	0.65	0.81	34
		U-LASSO	1.1044	0.07	0.08	0.09	11
		SCAD	0.7945	0.05	0.03	0.03	9
		MCP	0.8122	0.05	0.00	0.00	8
		3	0.9	Ridge	3.1102	0.07	0.80
LASSO	2.0007			0.1	0.30	0.36	19.5
ENET	2.6300			0.09	0.60	0.75	32
U-LASSO	1.8662			0.16	0.13	0.16	12
SCAD	1.9288			0.2	0.08	0.09	10
MCP	2.4477			0.18	0.10	0.09	10
9	0.7			Ridge	8.4471	0.29	0.80
		LASSO	1.9735	0.12	0.33	0.41	21
		ENET	5.2776	0.24	0.73	0.91	37
		U-LASSO	0.8452	0.06	0.03	0.03	9
		SCAD	0.7744	0.04	0.00	0.00	8
		MCP	0.7550	0.04	0.00	0.00	8
		9	0.9	Ridge	8.4469	0.21	0.80
LASSO	1.9876			0.1	0.28	0.34	19
ENET	5.2471			0.2	0.73	0.91	37
U-LASSO	0.9076			0.06	0.03	0.03	9
SCAD	0.7953			0.05	0.00	0.00	8
MCP	0.8047			0.06	0.00	0.00	8

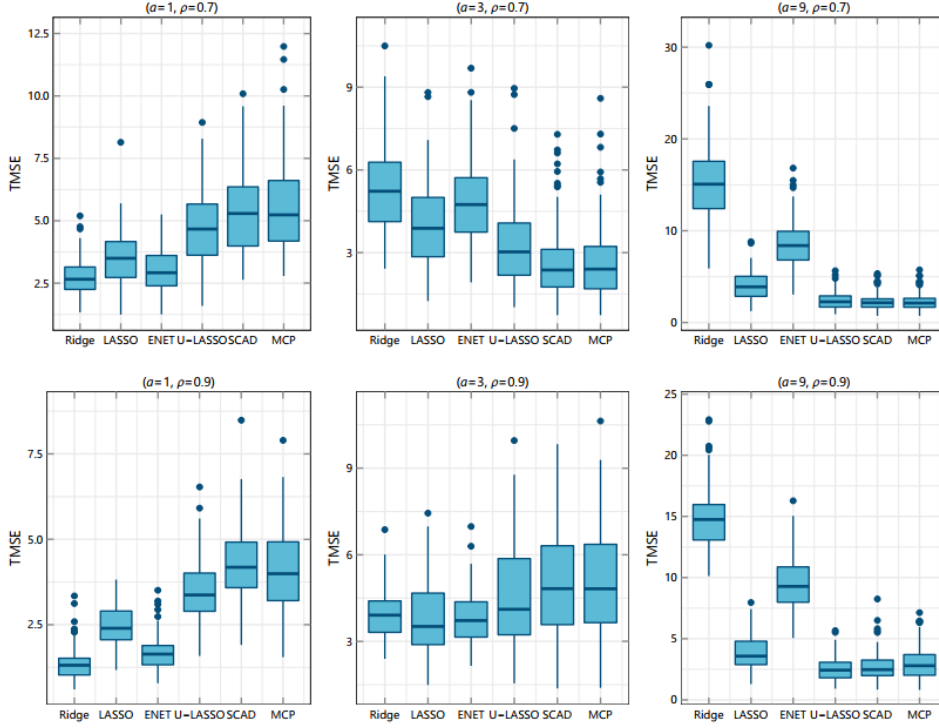
Tablo 4. S2 grubu simülasyon çalışmalarında tahmin edicilerin kalite ölçüleri ($k = 9$).

a	ρ	Yöntem	Medyan TMSE	Standart Sapma	Yanlış Sınıflama Oranı	Yanlış Pozitif Oranı	Aktif Küme Büyüklüğü
1	0.7	Ridge	1.5891	0.03	0.90	1.00	40
		LASSO	1.1860	0.07	0.25	0.25	13
		ENET	1.2246	0.05	0.35	0.39	18
		U-LASSO	1.4559	0.08	0.20	0.17	9
		SCAD	1.4318	0.07	0.15	0.17	9
		MCP	1.3318	0.11	0.10	0.08	6
1	0.9	Ridge	0.9749	0.04	0.90	1.00	40
		LASSO	1.0032	0.06	0.23	0.22	11
		ENET	0.9453	0.05	0.35	0.39	17.5
		U-LASSO	1.2402	0.07	0.20	0.17	8
		SCAD	1.4601	0.09	0.13	0.08	6
		MCP	1.5363	0.06	0.10	0.08	4
3	0.7	Ridge	3.5903	0.10	0.90	1.00	40
		LASSO	1.2542	0.07	0.23	0.25	13
		ENET	2.4609	0.11	0.63	0.69	29
		U-LASSO	0.5620	0.06	0.05	0.06	6
		SCAD	0.3893	0.03	0.05	0.06	6
		MCP	0.4291	0.04	0.03	0.03	5
3	0.9	Ridge	2.4880	0.08	0.90	1.00	40
		LASSO	1.1827	0.06	0.20	0.22	12
		ENET	1.9369	0.08	0.55	0.61	26
		U-LASSO	0.8650	0.07	0.09	0.08	7
		SCAD	0.4661	0.04	0.03	0.03	5
		MCP	0.6660	0.09	0.03	0.03	5
9	0.7	Ridge	5.9307	0.23	0.90	1.00	40
		LASSO	1.2536	0.07	0.23	0.25	13
		ENET	4.1913	0.22	0.81	0.90	36.5
		U-LASSO	0.3579	0.06	0.03	0.03	5
		SCAD	0.3433	0.03	0.00	0.00	4
		MCP	0.3564	0.04	0.00	0.00	4
9	0.9	Ridge	5.3943	0.17	0.90	1.00	40
		LASSO	1.1881	0.06	0.20	0.22	12
		ENET	3.8397	0.15	0.78	0.86	35
		U-LASSO	0.4080	0.02	0.03	0.03	5
		SCAD	0.3440	0.03	0.00	0.00	4
		MCP	0.3492	0.03	0.00	0.00	4

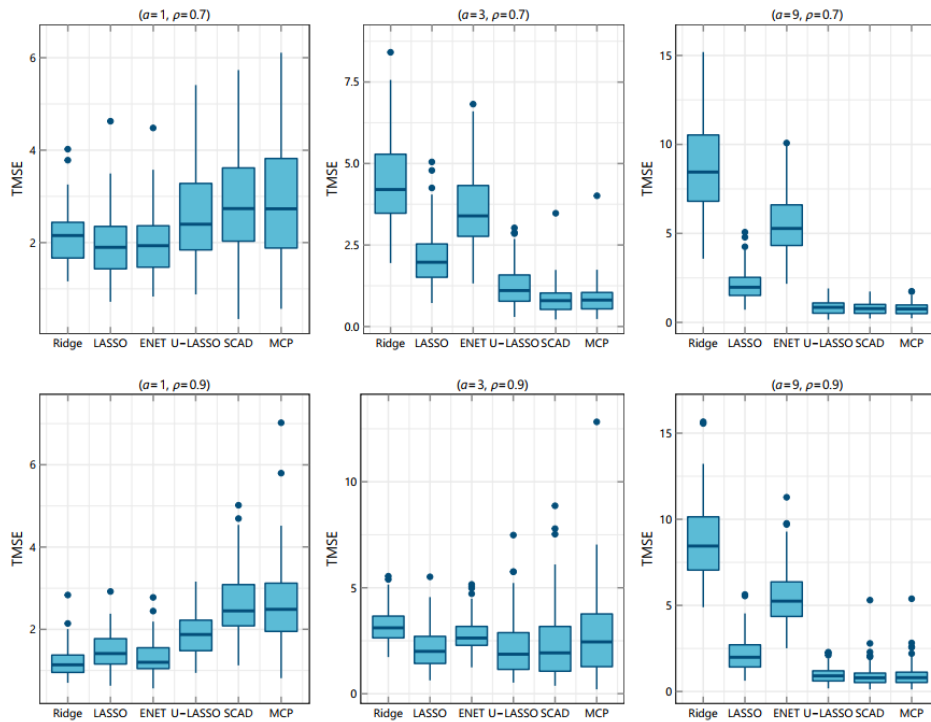
S2 grubundaki simülasyon çalışmalarında Ridge, LASSO, ENET, U-LASSO, SCAD ve MCP'den elde edilen TMSE değerlerinin kutu grafikleri Şekil 2,3,4'te verilmiştir. Kutu grafikleri, Tablo 2,3,4'te verilen analiz sonuçları ile uyumludur.

Bu çalışmanın bulguları ridge, LASSO, ENET, U-LASSO, SCAD ve MCP cezalı regresyon yöntemlerine dayalı olarak elde edilmiştir. Literatürde, katsayı vektörüne dair

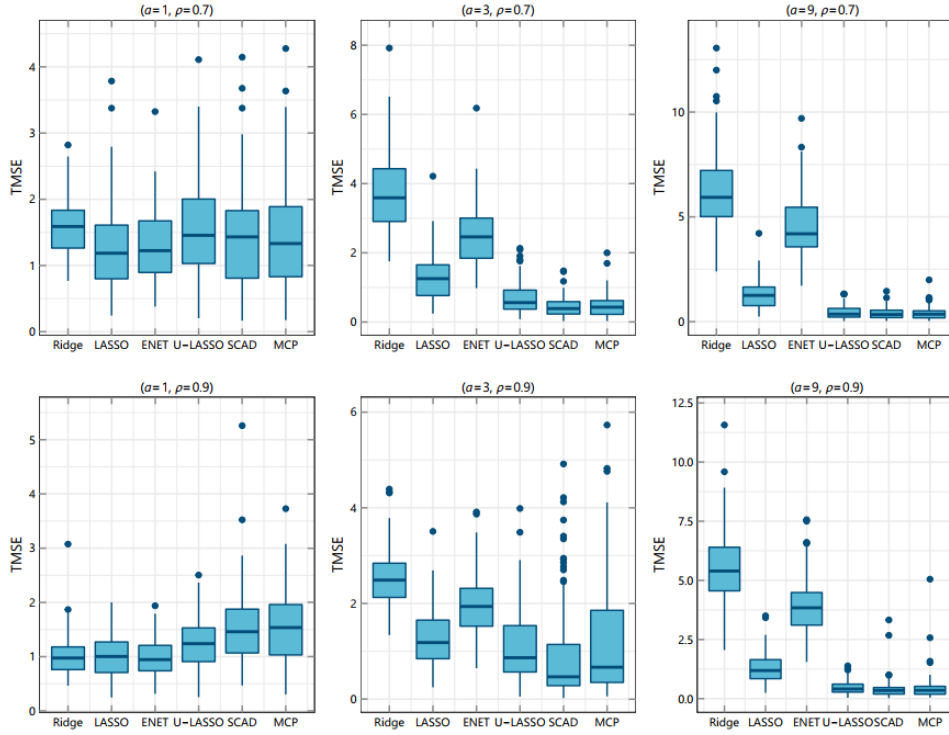
önsel bilginin kullanımını içeren Genç-Özkale [25], shift-ridge, shift-lasso [26] ve önsel LASSO [27] gibi LASSO'nun genelleştirmeleri olan tahmin ediciler de bulunmaktadır. Ayrıca veri kümesinin yapısına bağlı olarak ön tahminde başarılı sonuçlar veren grup LASSO [28] ve kaynaşmış LASSO [29] tahmin edicileri mevcuttur. Söz konusu tahmin edicilerin bu çalışmada incelenen tahmin edicilere göre performanslarının karşılaştırılması ayrı bir çalışmanın konusudur.



Şekil 2. S2 grubu simülasyon çalışmalarında TMSE değerlerinin kutu grafikleri ($k = 1$).



Şekil 3. S2 grubu simülasyon çalışmalarında TMSE değerlerinin kutu grafikleri ($k = 4$).



Şekil 4. S2 grubu simülasyon çalışmalarında TMSE değerlerinin kutu grafikleri ($k = 9$).

4. Sonuçlar

Bu çalışmada, gerçek katsayı vektörünün özelliklerine bağlı olarak konveks ve konveks olmayan cezalı regresyon yöntemlerinin doğrusal regresyon modelinin performansı üzerindeki etkisi araştırılmıştır. Model katsayılarının büyüklüğü, konumu ve sinyal değerleri göz önüne alınarak; ridge, LASSO, ENET ve U-LASSO konveks cezalı regresyon yöntemleri ile SCAD ve MCP konveks olmayan cezalı regresyon yöntemlerinin ayrıntılı olarak bir karşılaştırması yapılmıştır. TMSE karşılaştırma kriterinden elde edilen sonuçlara göre modelin gerçek katsayı vektörünün yapısına bağlı olarak konveks ve konveks olmayan cezalı regresyon yöntemlerinin ön tahmin performansı gerçek katsayı vektörünün yapısından büyük ölçüde etkilenmektedir. Yanlış sınıflama oranı, yanlış pozitif oranı ve aktif küme büyüklüğü ölçülerine göre gerçek katsayı vektörü ve veri kümesinin özellikleri, konveks ve konveks olmayan cezalı regresyon yöntemlerinin değişken seçimindeki başarısı üzerinde etkilidir. Gerçekleştirilen simülasyon çalışmaları ile seyreklik ve sinyal büyüklüğü bakımından farklı yapıda olan modellerde, tahminleme ve değişken seçimi konusunda konveks ve konveks olmayan cezalı regresyon yöntemlerinin güçlü ve zayıf yönleri ortaya konmuştur.

Kaynaklar

- [1] Hoerl, A.E., Kennard, R.W., Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 1970;12(1):55-67.
- [2] Rao, C.R., Toutenburg, H., Linear models: Springer; 1995.

- [3] Sarkar, N., A new estimator combining the ridge regression and the restricted least squares methods of estimation. **Communications in Statistics-Theory and Methods**, 1992;21(7):1987-2000.
- [4] Miller, A., **Subset Selection in Regression**: CRC Press; 2002.
- [5] Breiman, L., Better subset regression using the nonnegative garrote. **Technometrics**, 1995;37(4):373-384.
- [6] Frank, L.E., Friedman, J.H., A statistical view of some chemometrics regression tools. **Technometrics**. 1993;35(2):109-135.
- [7] Tibshirani, R., Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**. 1996;58(1):267-288.
- [8] Zou, H., Hastie, T., Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: series B (Methodological)**. 2005;67(2):301-320.
- [9] Zou, H., The adaptive lasso and its oracle properties. **Journal of the American Statistical Association**. 2006;101(476):1418-1429.
- [10] Fan, J., Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties. **Journal of the American Statistical Association**. 2001;96(456):1348-1360.
- [11] Meinshausen, N., Yu, B., Lasso-type recovery of sparse representations for high-dimensional data. **The Annals of Statistics**. 2009;37(1):246-270.
- [12] Johnstone, I.M., Titterton, D.M., Statistical challenges of high-dimensional data. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**. 2009;367(1906):4237-4253.
- [13] Shahriari, S., Faria, S., Gonçalves A.M., Variable selection methods in high-dimensional regression—A simulation study. **Communications in Statistics-Simulation and Computation**. 2015;44(10):2548-2561.
- [14] Ahmed, S.E., Kim, H., Yıldırım, G., Yüzbaşı, B., High-Dimensional Regression Under Correlated Design: An Extensive Simulation Study. **International Workshop on Matrices and Statistics**, Springer. 2016:145-175.
- [15] Genç, M., Bir Simülasyon Çalışması ile Cezalı Regresyon Yöntemlerinin Karşılaştırılması. **Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi**. 2022;9(1):80-91.
- [16] Genç, M., Özbilen, Ö., The Effect of the Second Stage Estimator on Model Performance in Post-LASSO Method. **Turkish Journal of Science and Technology**. 2023;18(2):319-330.
- [17] Hastie, T., Tibshirani, R., Friedman, J., The elements of statistical learning. New York: Springer series in statistics; 2001.
- [18] Efron, B., Hastie, T., Johnstone, I., Tibshirani R., Least angle regression. **The Annals of Statistics**. 2004;32(2):407-499.
- [19] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., Distributed optimization and statistical learning via the alternating direction method of multipliers. **Foundations and Trends in Machine learning**. 2011;3(1):1-122.
- [20] Friedman, J., Hastie, T., Tibshirani, R., Regularization paths for generalized linear models via coordinate descent. **Journal of Statistical Software**. 2010;33(1):1-22.

- [21] Zhang, C.H., Nearly unbiased variable selection under minimax concave penalty. **The Annals of Statistics**. 2010;38(2):894-942.
- [22] Zou, H., Li, R., One-step sparse estimates in nonconcave penalized likelihood models. **The Annals of Statistics**. 2008;36(4):1509-1533.
- [23] Breheny, P., Huang, J., Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. **The Annals of Applied Statistics**. 2011;5(1):232-253.
- [24] Hussami, N., Tibshirani, R., A component lasso. **Canadian Journal of Statistics**. 2015;43(4):624-646.
- [25] Genç, M., Özkale, M. R., Regularization and variable selection with triple shrinkage in linear regression: a generalization of lasso. **Communications in Statistics-Simulation and Computation**. 2023.
- [26] Genç, M., Özkale, M. R., Usage of the GO estimator in high dimensional linear models. **Computational Statistics**. 2021;36(1):217-239.
- [27] Jiang, Y., Variable selection with prior information for generalized linear models via the prior lasso method. **Journal of the American Statistical Association**. 2016;111(513):355-376.
- [28] Yuan, M., Lin, Y., Model selection and estimation in regression with grouped variables. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**. 2006;68(1):49-67.
- [29] Tibshirani, M., Saunders, M., Rosset, S., Zhu, J., Knight, K., Sparsity and smoothness via the fused lasso. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**. 2005;67(1): 91-108.