

Araştırma Makalesi

Mersin Univ Sağlık Bilim Derg 2021;14(1):105-112

doi: 10.26559/mersinsbd.816561

Etkili değişkenlerin cezalı regresyon yöntemleri ile belirlenmesi: Diyabet veri kümesi üzerine bir uygulama

 Didem Derici Yıldırım¹,  Ali Türker Çiftçi¹

¹ Mersin Üniversitesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Öz

Amaç: Bu çalışmada etkili değişkenlerin bulunması amacıyla uygulanan klasik regresyon analizine alternatif olarak kullanılması önerilen ve son yıllarda sağlık verilerinde oldukça popüler hale gelen cezalı regresyon yöntemlerinden En Küçük Açık regresyonu (LARS) ve En Küçük Mutlak Küçülme ve Seçim Operatörü (LASSO) yöntemleri ele alınmıştır. Diyabet veri kümesi üzerine etkili değişkenlerin belirlenmesinde cezalı regresyon yöntemleri ve klasik regresyon analizi sonuçlarının hata kareler ortalaması (HKO) ve belirtme katsayıları (R^2) bakımından karşılaştırılması amaçlanmıştır. **Yöntem:** Diyabet tanısı almış 442 hastaya ait veri kümesine En Küçük Açık regresyonu, En Küçük Mutlak Küçülme ve Seçim Operatörü ve çoklu doğrusal regresyon yöntemleri uygulanmıştır. **Bulgular:** En Küçük Açık regresyonu ve En Küçük Mutlak Küçülme ve Seçim Operatörü regresyon yöntemleri aynı değişkenleri seçerek model tahmini yapmıştır. Cezalı regresyon yöntemleri, belirtme katsayıları ve hata kareler ortalamaları dikkate alındığında çoklu doğrusal regresyondan daha iyi sonuçlar vermiştir. **Sonuç:** Diyabet veri seti için cezalı regresyon yöntemleri ile en az sayıda ve modeli en iyi açıklayan değişkenler elde edilmiştir. Daha az sayıda değişkenle anlamlı modeller oluşturulmak istendiğinde tercih edilebilir yöntemlerdir.

Anahtar Kelimeler: LARS, LASSO, cezalı regresyon, diyabet

Başvuru Tarihi: 26.10.2020

Kabul Tarihi: 09.12.2020

Sorumlu Yazar: Didem Derici Yıldırım, Mersin Üniversitesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Tlf:05053917768, E-posta: didemderici@hotmail.com

Determining the effective variables by penalized regression methods: An application on diabetes data set

Abstract

Aim: Least Angle Regression (LARS) and Least Absolute Shrinkage Selection Operator (LASSO) methods, which have become quite popular in recent years, were discussed as an alternative to classical regression analysis in this study. It is aimed to compare the results of classical regression analysis with these penalized regression methods for determination the effective variables on diabetes dataset in terms of mean square error (MSE) and coefficient of determination (R^2). **Methods:** Least Angle Regression, Least Absolute Shrinkage and Selection Operator and multiple regression methods were applied to data set of 442 patients diagnosed with diabetes. **Results:** Least Angle Regression and Least Absolute Shrinkage Selection Operator methods predict the model by selecting the same variables. However, these methods were better than multiple regression in terms of coefficient of determination and mean square error. **Conclusion:** Penalized regression methods constituted the best model for the diabetes data set with the least number of independent variables. These methods should be preferable to obtain significant models with fewer variables.

Keywords: LARS, LASSO, penalized regression, diabetes

Giriş

Diyabet dünya genelinde en yaygın görülen hastalıklardan biridir.¹ Dünya Sağlık Örgütü'nün raporuna göre 2030 yılına kadar diyabet tanısı almış hasta sayısı yaklaşık yedi katına çıkacaktır.² Bu durum nedeniyle diyabet tahminini yapabilecek modeller oluşturulması oldukça önemli hale gelmektedir.³ Bu tahmini en doğru şekilde yapabilmek için de regresyon analizinden yararlanılmaktadır.

Regresyon analizi, bir bağımlı değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişkinin modellenmesinde kullanılmaktadır. Bağımsız değişkenlerin bağımlı değişkendeki değişimi ne oranda açıkladıkları ve açıklama başarılarına göre sıralanmaları oldukça önemli hale gelmektedir. Çünkü regresyon analizinin en temel amacı, bağımlı değişkeni en iyi açıklayan, en uygun sayıda bağımsız değişkenle tahmin modelinin elde edilmesidir.⁴ Bağımsız değişken sayısının fazla olduğu ve bağımsız değişkenler arasında ilişki olan veri setlerinde klasik regresyon analizi ile oluşturulan modellerin tahmin başarısı düşük olmaktadır. Çalışmada ele alınan cezalı regresyon yöntemleri, klasik

regresyon yöntemlerine alternatif olarak geliştirilmiştir.^{5,6}

Cezalı regresyon yöntemleri, doğrusal regresyondaki En Küçük Kareler (EKK) tahmin edicilerinin varyansının büyük olması sebebiyle önerilen sapmalı tahmin edici yöntemlerdir. Bu yöntemlerin amacı EKK'ya göre daha küçük varyansla katsayıların tahmin edilmesidir. Bu yöntemler, bir ceza parametresi kullanır ve bu parametreden dolayı artık kareler toplamının minimize edilmesi sağlanır. En Küçük Mutlak Küçülme ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator, LASSO) yöntemi ilk olarak Robert Tibshirani (1996) tarafından jeofizik alanında kullanılmıştır.⁵ LASSO yöntemi, kullandığı ceza parametresinden dolayı bazı değişken katsayılarını küçültürken, bazı değişkenleri de sıfıra indirerek hem katsayıları tahmin etmek hem de değişken seçimi yapmak amacıyla kullanılan bir yöntemdir. Bir diğer yöntem olan En Küçük Açılı Regresyon (Least Angle Regression, LARS) ise ilk olarak Efron ve ark. (2004) tarafından öne sürülmüştür.⁷ Bu yöntem, her iterasyonda artıklarla en yüksek ilişkiye (korelasyona) sahip değişkenleri modele ekleyerek bir seçim prosedürü oluşturur. Bu çalışmada, diyabet tanısı almış

442 hastaya ait veri kümesine, cezalı regresyon yöntemlerinden LARS ve LASSO uygulanarak, elde edilen sonuçların çoklu doğrusal regresyon ile hata kareler ortalaması (HKO) ve belirtme katsayıları (R^2) bakımından karşılaştırılması amaçlanmıştır.

Gereç ve Yöntem

Çoklu doğrusal regresyon: En sık kullanılan regresyon yöntemlerinden biridir. Bir bağımlı değişken ve birden fazla bağımsız değişken olduğu durumda kullanılmaktadır. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi matematiksel modellerle ifade etmeye çalışır. Y_i bağımlı değişkeni, β_i regresyon katsayılarını ve X_i 'ler bağımsız değişkenleri göstermek üzere tüm doğrusal regresyon modelleri Eşitlik 1'de gösterildiği gibidir.⁴

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \quad i=1,2,\dots,p \quad (1)$$

En Küçük Mutlak Shrinkage Seçim Operatörü yöntemi (Least Absolute Shrinkage and Selection Operator, LASSO): En küçük kareler tahmin edicisinin verideki ufak

değişiklerden çok etkilenmesi ve varyansın şişmesinden kaynaklı modele yanlış değişken seçimi gibi nedenlerden dolayı, Tibshirani (1996) tarafından alternatif yöntem olarak En Küçük Mutlak Shrinkage Seçim Operatör yöntemi(LASSO) önerilmiştir.⁵ Bu yöntemin iki ana görevi düzenleme ve değişken seçimidir. Regresyon katsayılarının mutlak değerinin toplamını sabit bir değerden daha az olmaya zorlayarak, hem değişken seçimini yapmış olur, hem de modeli daha iyi yorumlanabilir hale getirir.⁶ LASSO yöntemi, bazı katsayıları sıfıra doğru daraltan ve bazılarını da tamamen sıfır yapan bir yöntemdir. Ancak değişkenler arasında yüksek ilişki söz konusu olduğunda ilişkili olan değişkenlerden bir tanesini modele alır ve diğer ilişkili değişkenleri modelden çıkartır. Bu özellik de değişken seçimi olarak bilinmektedir.⁴ Bu yöntemin en büyük avantajı, çok iyi tahminler yapabmesidir. Çünkü katsayıların sıfır olması veya sıfıra doğru daralması varyansı azaltabilmektedir. Özellikle geleneksel yöntemlerin başarılı olmadığı bir durum olan örnek genişliği küçük ve bağımsız değişken sayısı fazla olduğu durumda kullanışlı bir yöntemdir.^{6,8} Lasso tahmincisi Eşitlik 2'de tanımlanmıştır.⁶

$$\hat{\beta}^{Lasso} = \operatorname{argmin}_{\beta} \left(\frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (2)$$

En Küçük Açık Regresyon Yöntemi (Least Angle Regression, LARS): En küçük açı regresyon yöntemi, LASSO yöntemi gibi cezalı regresyon yöntemidir. Bu iki yöntem birbirleriyle bağlantılıdır.^{7,9} Ancak LARS yöntemi, LASSO yönteminin yaptığı işlemleri daha hızlı ve tek bir seferde yapmaktadır.¹⁰ Bu yöntem geometrik temele dayanan ve geometrik düzlemdeki değişken doğrultularının arasında oluşan açının ikiye bölünmesini ifade eden bir sürece dayanmaktadır. Bu geometrik işlemler sayesinde çok büyük boyuttaki verilerde LASSO yöntemine göre daha hızlı ve daha kolay sonuçlar elde edilmektedir.¹¹

LARS yöntemi, bütün katsayıların sıfıra eşit olduğu durumla başlar. İlk adımda bağımlı değişken ile en yüksek ilişkiye

(korelasyona) sahip x_{j1} bağımsız değişkenini tespit eder. Bu x_{j1} değişkeni artıklarla ($y - \hat{y}$), en yüksek korelasyonlu değişken olmak şartıyla; b_{j1} katsayısı hesaplanır. Artıklarla, x_{j1} gibi en yüksek korelasyonlu ikinci değişken x_{j2} 'yi bulduğu zaman süreç durur. İkinci değişkeni aktif modele dâhil eder. LARS prosedürü, bu noktada diğer yöntemlerden farklılık göstermektedir. Algoritma, birinci değişken olan x_{j1} doğrultusunda devam etmek yerine, üçüncü en yüksek korelasyona sahip değişkeni modele alana kadar x_{j1} ve x_{j2} değişkenleri arasında oluşan eşit açı doğrultusunda devam eder.⁷ Bu işlem, diğer bir değişken olan x_{j3} değişkeni modele alındıktan sonra x_{j1}, x_{j2}, x_{j3} arasında geometrik düzlemde eşit olarak yani eşit açı yönü boyunca başka bir değişken bulunana kadar devam eder. Bu

süreç, modelde bütün değişkenler yer alana kadar sürer. Analiz sonucunda yöntemlerin karşılaştırılması amacıyla hesaplamaları, Eşitlik 3 ve 4'de verilen hata kareler ortalaması ve belirtme katsayısı değerleri kullanılmıştır.

$$HKO = \frac{Y'Y - \hat{\beta}'X'Y}{n-p-1} \quad (3)$$

$$R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \quad (4)$$

Uygulama Verisi

Bu çalışmada literatürde yer alan diyabet veri kümesi ele alınmış olup, veri kümesinde bağımlı değişken hastalığın başlangıcından bir yıl sonraya kadar geçen sürede hastalık seyrini gösteren nicel bir ölçüm (Y), bağımsız değişkenler ise aşağıda belirtildiği gibidir.⁷

X₁: Yaş (Age)

X₂: Cinsiyet (Sex)

X₃: Vücut Kitle İndeksi (BMI)

X₄: Ortalama Kan Basıncı (MAP)

X₅: Kan transfüzyon değeri (TC)

X₆: Düşük yoğunluklu kolesterol (LDL)

X₇: Yüksek yoğunluklu kolesterol (HDL)

X₈: Toplam kolesterol (TCH)

X₉: Karaciğer Trigliserit değeri (LTG)

X₁₀: Glukoz (GLU)

Bu çalışmanın analizi için R v.3.6.3 yazılımında "lars" paketinden yararlanılmıştır.

Bulgular

Regresyon analizine dâhil edilen 442 hastanın %46.8 erkek, %53.2 kadın olup, diğer değişkenlere dair tanımlayıcı istatistikler Tablo 1'de verilmiştir.

Tablo 1. Diyabet veri kümesine ait tanımlayıcı istatistikler

	Ort.±S S	Min-Max
Yaş(X ₁)	48.52±13.11	19-79
BMI(X ₃)	36.38±4.42	18-42.20
MAP(X ₄)	94.64±13.83	62-133
TC(X ₅)	189.14±34.61	97-301
LDL(X ₆)	115.44±30.41	41.60-242.40
HDL(X ₇)	49.79±12.93	22-99
TCH(X ₈)	4.08±1.29	2-9.09
LTG(X ₉)	4.64±0.52	3.26-6.11
GLU(X ₁₀)	91.26±11.50	58-124

Çalışmadaki bağımsız değişken sayısının fazla olmasından kaynaklı olarak bu değişkenler arasındaki ilişkinin de mutlaka test edilmesi gerekmektedir. Bağımsız değişkenler arasındaki ilişkiyi (çoklu bağlantı) incelemek amacıyla hesaplanan istatistikler Tablo 2'de

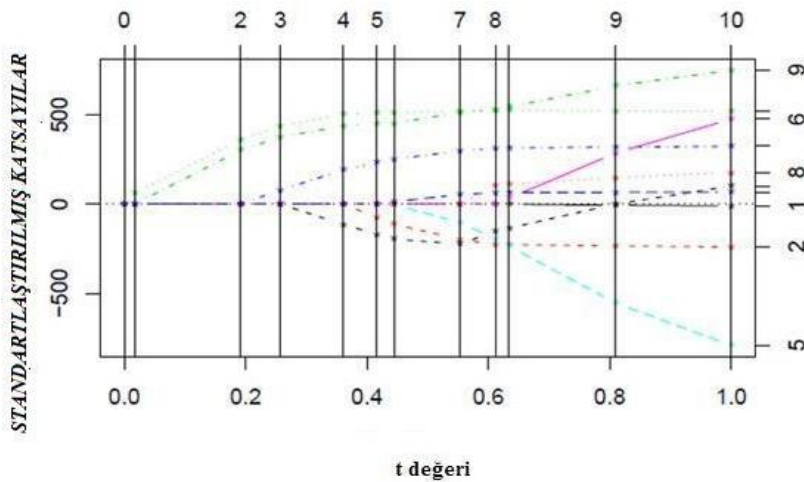
verilmiştir. Tolerans değerlerinin 0'a yakın olması ve varyans şişme değerlerinin (VIF) 10'dan büyük olması güçlü çoklu bağlantının varlığını göstermektedir. Tablo 2'de koyu renk ile belirtilen değerlere bakıldığında diyabet veri kümesinde çoklu bağlantı sorunu olduğu görülmektedir.

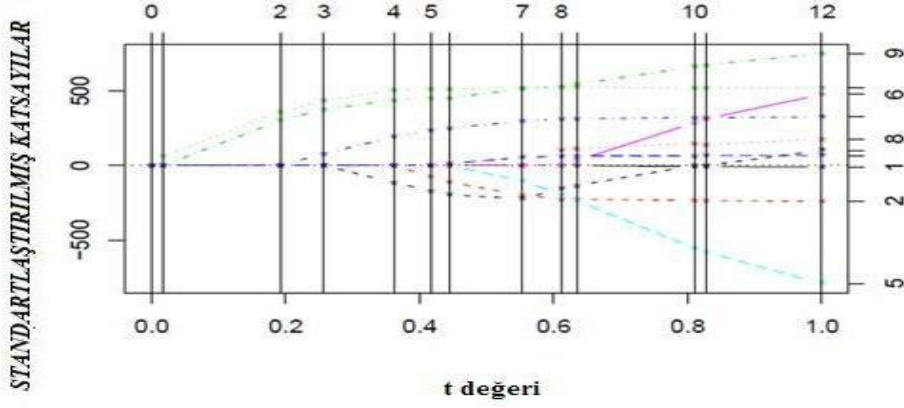
Tablo 2. Bağımsız değişkenlere ilişkin Tolerans ve VIF değerleri

	Tolerans Değerleri	VIF Değerleri
Yaş	0.821	1.217
Cinsiyet	0.782	1.278
BMI	0.662	1.509
MAP	0.685	1.459
TC	0.017	59.203
LDL	0.026	39.193
HDL	0.065	15.402
TCH	0.112	8.891
LTG	0.099	10.076
GLU	0.674	1.485

Önerilen yöntemlerin klasik regresyondan bir diğer farkı ise tüm değişkenleri modele almaktansa belirli prosedürlere göre sıralama yaparak belli bir adımda modele almasıdır. LASSO yöntemi değişken seçimini LAR yöntemine göre daha fazla adımda yapmıştır ve toplamda 12 adım sayısı uygulamıştır. Her iki yöntemde modele sırasıyla BMI(X_3), LTG(X_9), MAP(X_4), HDL(X_7), Cinsiyet (X_2), GLU(X_{10}), TC(X_5) değişkenlerini almıştır. Her bir yöntemin

değişken seçim sıralamasını gösteren grafikler Şekil 1 ve 2'de özetlenmiştir. Burada sütunda standartlaştırılmış katsayılar, satırda ise 't' noktasında o değişkene ait katsayının maksimum katsayıya oranı olarak belirtilen değer yer almaktadır. Grafiğin en üstünde adım sayıları yer almakta olup, her adımda eklenen değişkenler belirtilmiştir. Her renk bir değişkeni temsil etmektedir.

**Şekil 1.** LARS yöntemi için değişken sıralaması



Şekil 2. LASSO yöntemi için değişken sıralaması

Yöntemlere göre modelde yer alan değişkenler, katsayı tahminleri ve yöntem performanslarının karşılaştırılması için

hesaplanan HKO ve R^2 değerleri Tablo 3'de verilmiştir.

Tablo 3. Yöntemlere ait katsayı tahminleri, HKO ve R^2 sonuçları

Değişkenler	LARS	LASSO	Çoklu Doğrusal Regresyon
Yaş(X_1)	0.000	0.000	-10.012
Cinsiyet(X_2)	-197.757	-197.757	-239.819
BMI(X_3)	522.265	522.265	519.839
MAP(X_4)	297.160	297.160	324.390
TC(X_5)	-103.946	-103.946	-792.184
LDL(X_6)	0.000	0.000	476.745
HDL(X_7)	-223.926	-223.926	101.044
TCH(X_8)	0,000	0.000	177.064
LTG(X_9)	514.749	514.749	751.28
GLU(X_{10})	54.768	54.768	67.625
HKO	2859.690	2861.940	2930.690
R^2	0.513	0.513	0.507
HKO_çapraz geçerlilik	2790.163	3231.394	3302.085
R^2 _çapraz geçerlilik	0.496	0.495	0.488

LARS yöntemi ile LASSO yönteminin değişken katsayıları tamamen aynı çıkmıştır. En yüksek HKO değerine sahip yöntem çoklu doğrusal regresyon yöntemi iken, en düşük olan yöntem ise LARS yöntemi olmuştur. R^2

değerleri açısından ise LARS ve LASSO yöntemi tamamen aynı değeri verirken çoklu doğrusal regresyon için daha küçük bir değer elde edilmiştir. Katsayı tahminlerinde ise çoklu doğrusal regresyon yöntemi ile

diğerlerine göre daha büyük katsayılar elde edilmiş olup, cezalı regresyon yöntemleri ile daha az sayıda değişkenle model açıklama yüzdesi daha yüksek modeller elde edilmiştir.

LASSO ve LARS yönteminde cezayı kontrol eden bir λ değeri bulunmaktadır. $\lambda = 0$ olduğunda EKK ile aynı sonuçlar elde edilmektedir. Uygun λ değeri 10 katlı çapraz geçerlilik ile elde edilmiştir. Yöntemlerin karşılaştırılmasında sonuçların daha güvenilir olması adına, verinin %67'si ile model oluşturup, %33'ü üzerinde denenmiştir. Bu işlem 1000 kez tekrarlanarak elde edilen model performansını gösteren HKO ve R^2 sonuçları Tablo 3'de HKO_çapraz geçerlilik ve R^2 _çapraz geçerlilik olarak eklenmiştir. Bu sonuçlara göre de LASSO ve LARS yöntemleri yine çoklu doğrusal regresyondan daha başarılı bulunmuş olup, yöntemler arasındaki farklılıklar biraz daha belirginleşmiştir.

Tartışma

Cezalı regresyon yöntemleri, klasik regresyondan farklı olarak hatayı en aza indirmesi ve tahmin gücünün yüksek olmasından kaynaklı son yıllarda oldukça popüler hale gelmiştir. Bu çalışmada kullanılan yöntemler HKO ve R^2 değerleri bakımından karşılaştırılmıştır. Bu sonuçlara göre bağımsız değişken sayısı fazla olduğu ve özellikle bağımsız değişkenler arasında ilişki olduğu durumda çoklu doğrusal regresyon analizi, LASSO ve LARS gibi cezalı regresyon yöntemlerine göre daha başarısız bulunmuştur.

Değişken seçimi elde edilen modelin başarısı açısından oldukça önemlidir. Günümüzde verilerin boyutu artmakta olup, oldukça fazla olan bağımsız değişkenler içerisinde sonuç değişkenini en iyi açıklayan değişkenlerin tespit edilmesi oldukça önemli bir hale gelmiştir.³ Özellikle karmaşık yapıya sahip hastalıklar için en uygun değişken seçim yönteminin kullanılması daha doğru bilgiler elde edilmesini sağlayacaktır. Regresyon analizi için kullanılan geleneksel değişken seçim yöntemleri ileriye yönelik, geriye yönelik ve

adimsal yöntemler olarak sıralanabilir. Ancak bu yöntemlerin en önemli kısıtlılığı değişkenler arası ilişkiyi dikkate almamasıdır. Aynı zamanda katsayı tahminleri yanıltıcı olabilmektedir. Bu probleme çözüm bulmak amacıyla LASSO ve LARS yöntemlerinden sıklıkla yararlanılmaktadır.

Sağlık alanında LASSO regresyon bir çalışmada kronik subdural hematoma için ilişkili biyomarkerler içerisinde hastalığı en iyi açıklayanı tespit etmek için kullanılırken¹², başka bir çalışmada birbiriyle ilişkili genlerin hastalıkla ilişkisini ortaya koymak için kullanılmıştır.¹³ Bazı çalışmalarda da cezalı regresyon yöntemleri bizim çalışmamızda olduğu gibi klasik yöntemlerle kıyaslanmıştır. Literatürde Prostat Kanseri verisinin kullanıldığı bir çalışmada ise LASSO ve çoklu doğrusal regresyon yöntemlerinin de aralarında olduğu yöntem karşılaştırması sonucunda LASSO yönteminin daha iyi model tahmini yaptığı sonucuna varılmıştır.¹⁴ Bir başka çalışmada ise LASSO ve Adimsal Regresyon yöntemi karşılaştırılmasında, LASSO yöntemi için R^2 değeri 0.968; Adimsal Regresyonda ise R^2 değeri 0.786 olarak bulunmuş ve LASSO'nun Adimsal Regresyona göre daha iyi tahmin ettiği görülmüştür.¹⁵ Zaman serileri analizinde değişken seçimi ile model başarısının değerlendirilmesini ele alan bir çalışmada ise LASSO, LARS ve en küçük kareler tahminleri karşılaştırılmış, en başarılı tahminin LARS yöntemi ile gerçekleştiği sonucuna varılmıştır.¹⁶

Sonuç ve Öneriler

Çalışmada ele alınan cezalı regresyon yöntemleri birbirlerine avantaj ve dezavantajlarına sahip olmasına rağmen genel olarak bakıldığında yakın sonuçlar vermiştir. Elde edilmesi zor klinik verilere geleneksel yöntemler dışında yeni geliştirilen yöntemlerin uygulanması ile bağımsız değişkenler arasındaki ilişkiler dikkate alınarak, anlamlılığı en doğru şekilde ortaya koyulmuş modeller elde edilmektedir. Son yıllarda kod ağırlıklı programların kullanımının artmasına paralel olarak bu yöntemlerin kullanımı daha kolay hale gelmiştir. Bu çalışma ile LASSO ve LARS gibi

cezalı regresyon yöntemlerinin klinik araştırmacılara tanıtılması ve sağlık alanında kullanılabilirliğini ortaya koymak amaçlanmıştır.

Yazar katkısı: Tüm yazarlar araştırmanın planlama, uygulama, çözümleme ve yazım aşamalarında görev almıştır.

Çıkar çatışması/Mali destek: Yazarların çıkar çatışması bulunmamaktadır./Bu çalışma için herhangi bir mali destek alınmamıştır.

Kaynaklar

1. Png ME, Yoong J, Tan CS, Chia KS. Excess Hospitalization Expenses Attributable to Type 2 Diabetes Mellitus in Singapore. *Value in health regional issues*. 2018; 15:106-111.
2. 3. Organization WH. Global report on diabetes. 2016.
3. Farbahari A, Dehesh T, Gozashti MH. The Usage Of Lasso, Ridge and Linear Regression to Explore The Most Influential Metabolic Variables That Affect Fasting Blood Sugar In Type 2 Diabetes Patients. *Rom J Diabetes Nutr Metab Dis*. 2020; 26(4) :371-379.
4. Alpar R. Çoklu Doğrusal Regresyon. İçinde: Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 5. Baskı. Ankara: Detay Yayıncılık; 2017: 399-400.
5. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B*. 1996; 58(1):267-288.
6. Fonti V, Belitser E. Feature Selection Using Lasso. Research Paper In Business Analytics. 2017. https://beta.vu.nl/nl/images/werkstuk-fonti_tcm235-836234.pdf.
7. Efron B, Hastie T, Johnstone I, Tibshirani R. Least Angle Regression. *The Annals of Statistics. Institute of Mathematical Statistics*. 2004; 32(2):407-499.
8. Kayanan M, Wijekoon P. Performance of Lasso And Elastic Net Estimators in Misspecified Linear Regression Model. *Ceylon Journal Of Science*. 2019; 48(3):293-299.
9. Hastie TJ, Tibshirani R, Friedman J. Linear Methods For Regression. İçinde: Hastie TJ, Tibshirani R, Friedman J. The Elements of Statistical Learning Data Mining, Inference and Prediction. 2nd ed. New York: Springer; 2008: 61-73.
10. Khan JA, Van Aelst S, Zamar R.H. Robust Linear Model Selection Based On Least Angle Regression. *Journal of the American Statistical Association*. 2007; 102(480):1289-1299.
11. Januaviani Adelheid MT, Gusriani N, Joebaedi K, Supian S, Subiyanto. The Best Model Of LASSO With The LARS (Least Angle Regression and Shrinkage) Algorithm Using Mallows's C_p . *An International Scientific Journal*. 2019; 116:245-252.
12. Pripp AH, Stanisic M. Association between biomarkers and clinical characteristics in chronic subdural hematoma patients assessed with lasso regression. *Plos One*. 2017; 12(11):1-15.
13. Anaraki JR, Usefi H. A Comparative Study of Feature Selection Methods on Genomic Datasets. 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS); June 5, 2019; Cordoba, Spain.
14. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J. Royal. Statis. Soc. B*. 2005; 67(2):301-320.
15. Gauthier PA, Scullion W, Berry A. Sound quality prediction based on systematic metric selection and shrinkage: Comparison of stepwise, lasso, and elastic-net algorithms and clustering preprocessing. *Journal of Sound and Vibration*. 2017; 400:134-153.
16. Iturbide E, Cerda J, Graff M. A Comparison between LARS and LASSO for Initialising the Time-Series Forecasting Auto-Regressive Equations *Procedia Technlogy*. 2013; 7:282-2.