

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
İnşaat Mühendisliği Anabilim Dalı



KESİKLİ TERCİH MODELLERİ

Discrete Choice Models

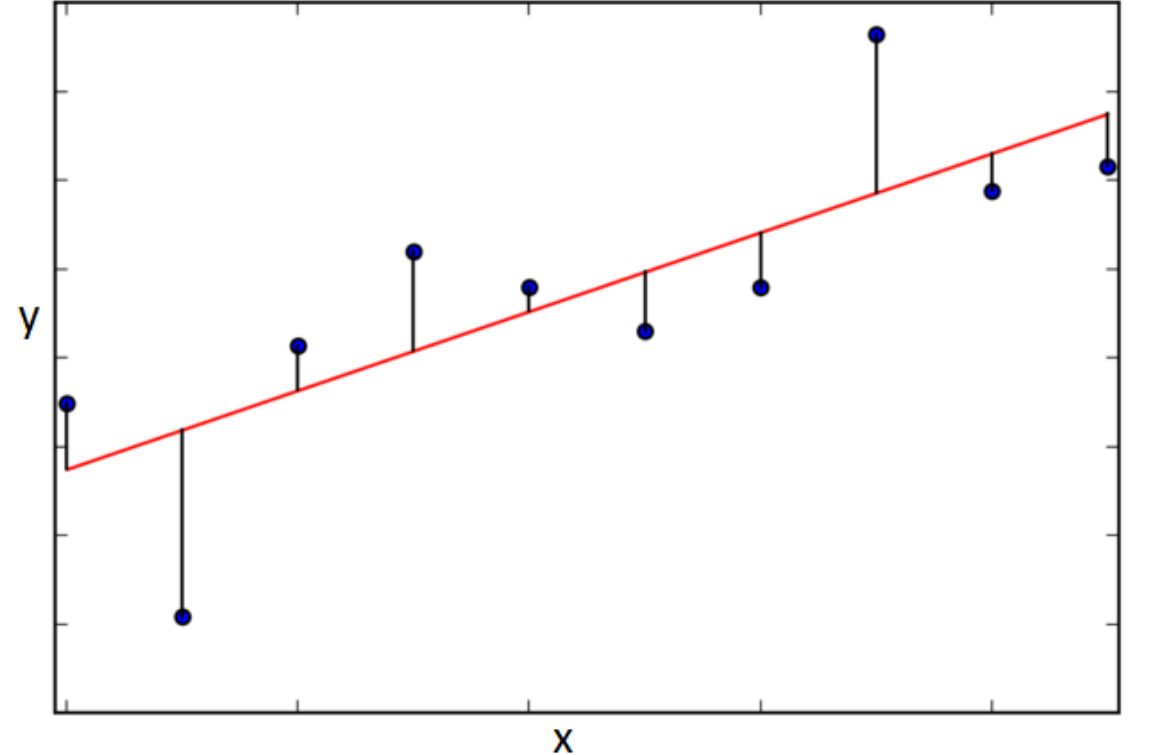
Dr. Kadir Berkhan AKALIN

4

REGRESYON ANALİZİ

Bağımsız (açıklayıcı) deęişken(ler)in deęerlerine dayalı olarak baęımlı (cevap) bir deęişkenin deęerlerini tahmin edebilen istatistiksel bir model geliştirilmesi amacıyla kullanılan yaygın bir yöntemdir.

Baęımlı ve baęımsız deęişkenlerin
«sürekli» deęişkenler olduęu
varsayımı geçerlidir.



REGRESYON ANALİZİ

BASİT REGRESYON: Bir bağımlı değişkenin (y) bir bağımsız değişken (x) yardımıyla tahmin edildiği model yaklaşımıdır.

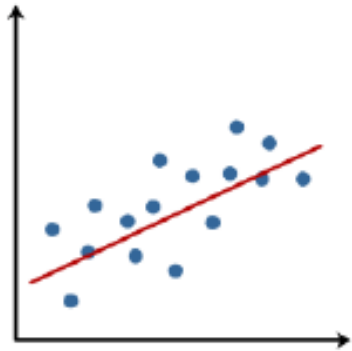
$$y = b_0 + b_1x + \varepsilon$$

ÇOKLU REGRESYON: Bir bağımlı değişkenin (y) birden fazla bağımsız değişken (x_1, \dots, x_n) yardımıyla tahmin edildiği model yaklaşımıdır.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon$$

REGRESYON ANALİZİ (REGRESSION ANALYSIS)

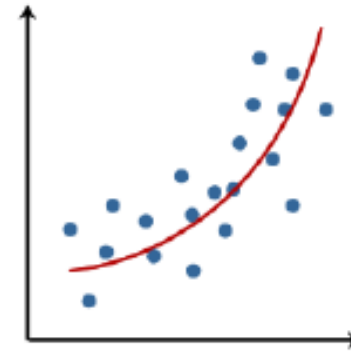
DOĞRUSAL (LINEAR) REGRESYON



Doğrusal ilişki

Örneğin;
 $y = b_0 + b_1x$

DOĞRUSAL OLMAYAN (NON-LINEAR) REGRESYON

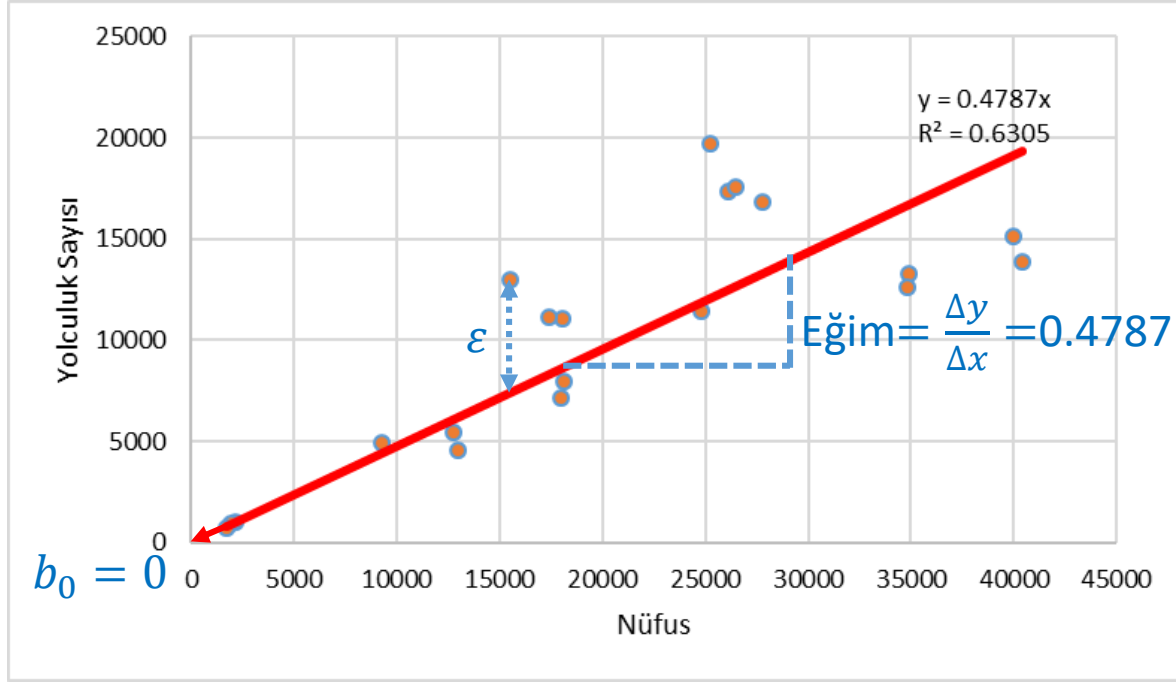


**Doğrusal olmayan ilişki:
Logaritmik, üstel vb.**

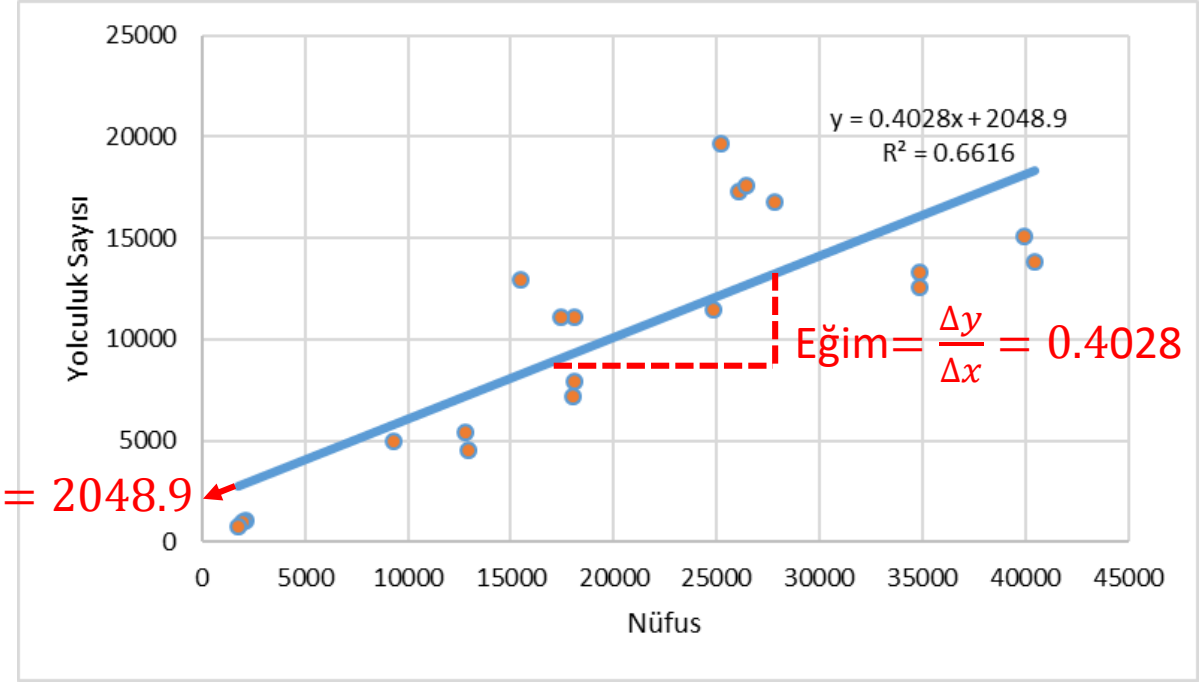
Örneğin;
 $y = b_0 + b_1\ln(x)$

DOĞRUSAL REGRESYON ANALİZİ

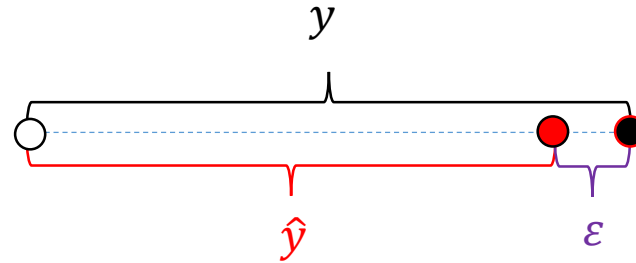
$$\hat{y} = b_0 + b_1x + \varepsilon$$



$$\hat{y} = 0.4787x$$



$$\hat{y} = 2048.9 + 0.4028x$$



KATSAYILARIN TAHMİNİ

En Küçük Kareler Yöntemi (EKK, Method of Least Squares –MLS): Gerçek ve tahmin edilen değerler arasındaki farkların mümkün olduğunca küçük olduğu doğrusal modelin geliştirilmesini amaçlamaktadır.

Regresyon fonksiyonu: $y = b_0 + b_1x$

Eğim : $b_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$

Kesişim : $b_0 = \bar{y}(b_1\bar{x})$

KORELASYON (BELİRLEME) KATSAYISI

Regresyon ile açıklanan toplam varyasyon oranı, korelasyon veya belirleme katsayısı olarak bilinir ve genellikle R^2 olarak anılır. Bağımlı değişken (y) ile bağımsız değişkenler (x) arasındaki korelasyonu birlikte ölçer.

R^2 'nin değeri 0 ile 1 arasında değişebilir ve değeri ne kadar yüksek olursa, regresyon modelinin o kadar gerçeği yansıttığı varsayılır.

Düzeltilmiş R^2 , bir veri setindeki değişken sayısını hesaba katar ve modele uymayan noktaları cezalandırır.

KORELASYON (BELİRLEME) KATSAYISI

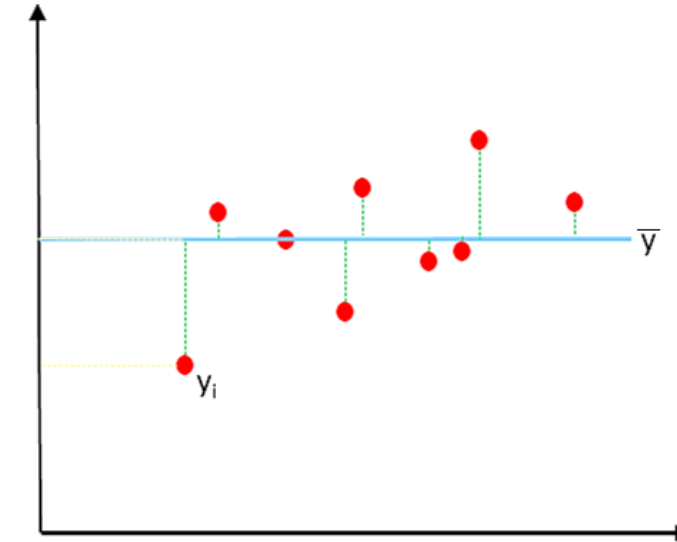
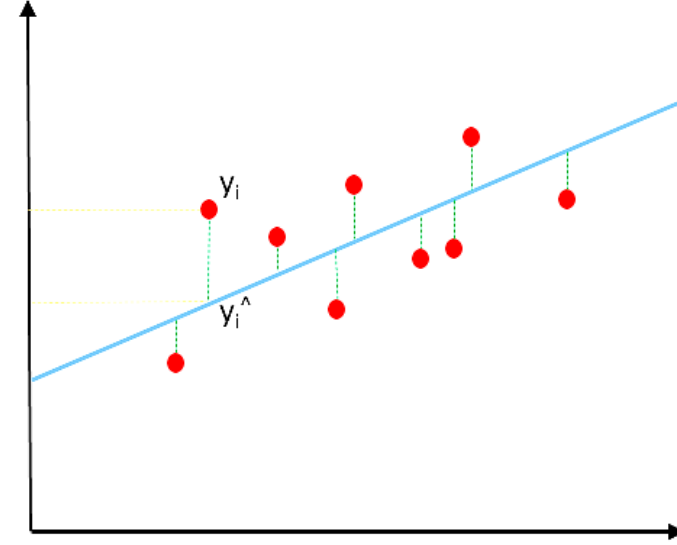
$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Düzeltilmiş (Adjusted) Belirleme Katsayısı

$$Adj_R^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

n ; Örneklem büyüklüğü,

p ; Bağımsız değişken sayısı.



R-Project

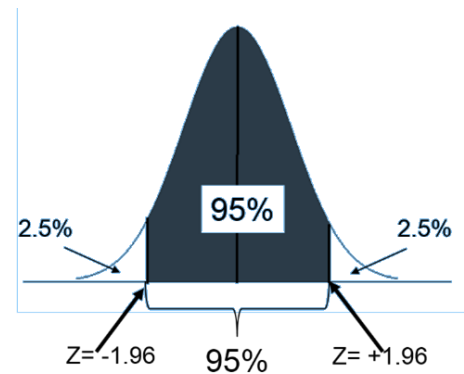
İstatistiksel hesaplama ve grafikler için bir programlama dilidir.



RStudio

İstatistiksel hesaplama ve grafikler için bir programlama dili olan R için entegre açık kaynak kodlu bir geliştirme ortamıdır.

Anlamlılık Seviyesi (Significance Level)	Güven Aralığı (Confidence Interval)	z-değeri
$\alpha = 0.01 = \%1$	$1 - \alpha = \%99$	± 2.576
$\alpha = 0.05 = \%5$	$1 - \alpha = \%95$	± 1.960
$\alpha = 0.10 = \%10$	$1 - \alpha = \%90$	± 1.645
$\alpha = 0.20 = \%20$	$1 - \alpha = \%80$	± 1.282
$\alpha = 0.30 = \%30$	$1 - \alpha = \%70$	± 1.036



DÜZENLİLEŞTİRME (REGULARIZATION) YÖNTEMLERİ

İstatistik, matematik ve bilgisayar bilimlerinde, özellikle makine öğrenimi problemlerinde, değişken sayısının artmasıyla meydana gelebilecek aşırı öğrenme (overfitting) sorunları, hangi değişkenlerin kullanılacağına tespit edilmesinde yaşanan zorluklar, yorumlanması zor olan modellerin üretilmesi gibi problemlerin üstesinden gelmek için kullanılan yöntemlerdir.

Ridge, Lasso ve Elastic Net teknikleri en yaygın kullanılan tekniklerdir.

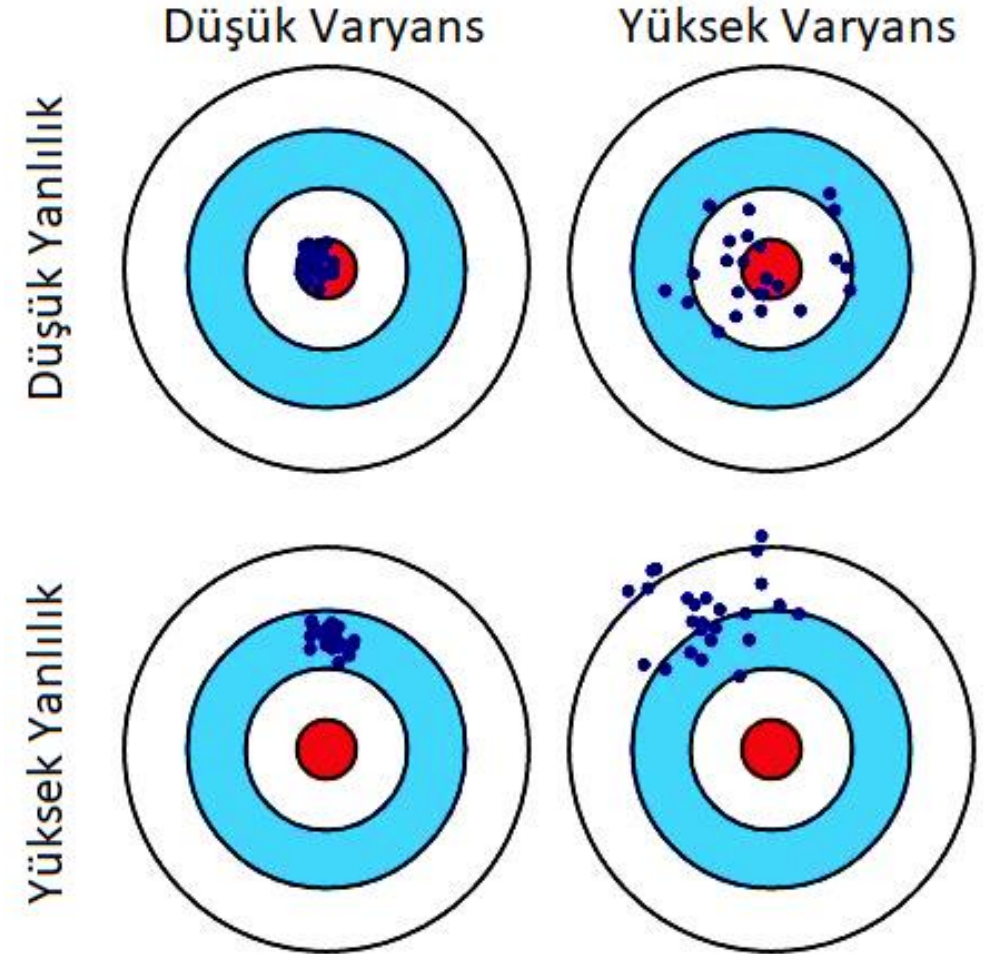
DÜZENLİLEŞTİRME (REGULARIZATION) YÖNTEMLERİ

Doğrusal Regresyon için;

$$L_{EKK}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 = \|y - X\hat{\beta}\|^2$$

$$\text{Bias}(\hat{\beta}_{EKK}) = E(\hat{\beta}_{EKK}) - \beta$$

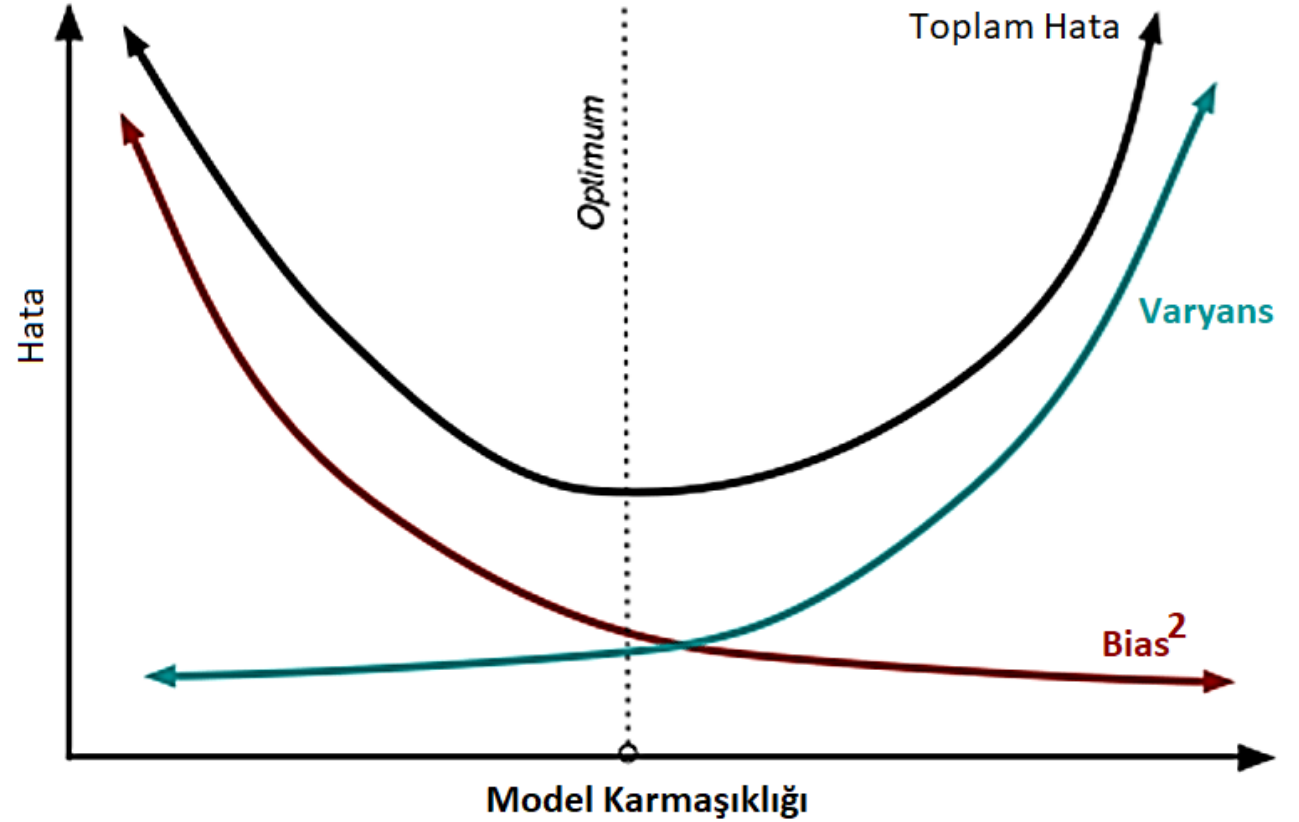
$$\text{Var}(\hat{\beta}_{EKK}) = \sigma^2 (X'X)^{-1}$$



DÜZENLİLEŞTİRME (REGULARIZATION) YÖNTEMLERİ

EKK tahmincisi, tarafsız olma konusunda istenen özelliğe sahiptir. Ancak, varyans değerleri çok yüksek çıkabilir. Özellikle tahmin değişkenleri birbirleriyle yüksek oranda ilişkili olması durumu için sıklıkla bu durumla karşılaşılır.

Genel çözüm; biraz yanlılığı göze alarak varyansı azaltmak olabilir.



RIDGE REGRESYON (L2 REGULARIZATION)

Hata kareleri toplamını en küçükleyen katsayıları bir ceza uygulayarak hesaplar. Aşırı öğrenmeye karşı dirençlidir. Bütün değişkenler için modeli kurarken ilgisiz değişkenleri modelden çıkarmasa da katsayılarını sıfıra yakınsayabilir.

$$L_{Ridge}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m \omega_j \hat{\beta}_j^2 = \|y - X\hat{\beta}\|^2 + \lambda \|\hat{\beta}\|^2$$

$$Bias(\hat{\beta}_{Ridge}) = -\lambda(X'X + \lambda I)^{-1}\beta$$

$$Var(\hat{\beta}_{Ridge}) = \sigma^2(X'X + \lambda I)^{-1}X'X(X'X + \lambda I)^{-1}$$

λ büyüdükçe varyansın azaldığını ve yanlılığın arttığını görebilirsiniz. Ancak en uygun değer?

AIC ve BIC

Akaike ölçütü (Akaike information criterion –AIC), bir veri kümesi için istatistiksel modelin kalitesi ile ilgili göreceli ölçüdür. AIC her modelin kalitesini, diğer modellerle göreceli olarak tahmin etmektedir: $AIC = 2k - 2\ln(\hat{L})$

Bayes ölçütü (Bayesian information criterion –BIC), sonlu bir model seti arasından model seçimi için bir kriterdir: $BIC = k \ln(n) - 2\ln(\hat{L})$

AIC ve BIC değerleri düştükçe modelin tercih edilebilirliği artmaktadır.

\hat{L} ; maksimum olasılık fonksiyonu, k ; tahmin edilen parametre sayısı, n ; veri sayısı.

LASSO REGRESYON (L1 REGULARIZATION)

Ridge'de olduğu gibi hata kareleri toplamını en küçükleyen katsayıları bir ceza uygulayarak hesaplar. Ancak modelin yorumlanabilirliğini arttırmak için hem değişken seçimi hem düzenleme yapar. İlgisiz değişkenlerin katsayılarını doğrudan sıfıra götürür.

$$L_{Lasso}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m |\hat{\beta}_j|$$

ELASTIC NET

Ridge ve Lasso regresyon tekniklerini birleştirerek hem cezalandırma hem de değişken seçimi yapılmasını sağlar. Burada hem ceza parametrelerinin (λ) hem de bu iki tekniğin karışım oranının (α) ayarlanması gerekmektedir.

$$L_{ENet}(\hat{\beta}) = \frac{\sum_{i=1}^n (y_i - x_i' \hat{\beta})^2}{2n} + \lambda \left(\frac{1 - \alpha}{2} \sum_{j=1}^m \hat{\beta}_j^2 + \alpha \sum_{j=1}^m |\hat{\beta}_j| \right)$$

$\alpha = 0$ ise Ridge, $\alpha = 1$ ise Lasso

KESİKLİ TERCİH MODELLERİ

Discrete Choice Models



TEŞEKKÜRLER
HAFTAYA GÖRÜŞMEK ÜZERE 😊

Dr. Kadir Berkhan AKALIN

KAYNAK GÖSTERME

Bu sunuma ařađıdaki gibi atıf yapabilirsiniz:

Akalın, K.B. (2023). Kesikli Tercih Modelleri Ders Notu. Eskiřehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü

You can cite this presentation as follows:

Akalın, K.B. (2023). Discrete Choice Models Lecture Notes. Eskişehir Osmangazi University Graduate School of Natural and Applied Sciences.

KAYNAKLAR

- Akalın, K.B. (2021). Yolculuk Üretim ve Çekim Modellerinin Rastgele Pişmanlık Minimizasyonu ve Rastgele Fayda Maksimizasyonu Yöntemleri ile Geliştirilmesi. Doktora Tezi. Eskişehir Osmangazi Üniversitesi.
- Ben-Akiva, M., Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. Handbook of transportation science.
- Ben-Akiva, M., Lerman, S. (1985). Discrete Choice Analysis, The MIT Press.
- De Dios Ortúzar, J., Willumsen, L.G. (2011). Modelling Transport. John Wiley & Sons.
- Frumin, M., Ben-Akiva M. (2008). Transportation Systems Analysis: Demand And Economics. MIT Open Courseware.
- Hensher, D.A., Rose, J.M., and Greene, W.H. (2005). Applied Choice Analysis: A Primer, Cambridge University Press, Cambridge, UK.
- Richardson, A.J., Ampt, E.S., and Meyburg, A.H. (2012). Survey Methods for Transport Planning, Eucalyptus Press.
- Tezcan, H.O. (2021). Discrete Choice Modelling in Transportation Lecture Notes. İstanbul Technical University.
- Tezcan, H.O. (2021). Transportation Models Lecture Notes. İstanbul Technical University.
- Train, K. (2002). Discrete Choice Methods with Simulation, Cambridge University Press.
- <https://www.datacamp.com/tutorial/tutorial-ridge-lasso-elastic-net>
- <https://towardsdatascience.com/ridge-lasso-and-elasticnet-regression-b1f9c00ea3a3>
- <https://makineogrenimi.wordpress.com/2017/05/31/duzenlilestirme-regularization/>