Doğrusal Regresyon ve Kuzenleri

by Sefa Isci

Doğrusal Regresyon ve Kuzenleri

- Basit Doğrusal Regresyon
- Çoklu Doğrusal Regresyon
- Temel Bileşen Regresyonu
- Kısmı En Küçük Kareler Regresyonu
- Ridge Regresyon
- Lasso Regresyon
- Elastic Net Regresyonu
- Her Model İçin:
  - Model
  - Tahmin
  - Model Optimizasyonu

Basit Doğrusal Regresyon

Temel amaç, bağımlı ve bağımsız değişken arasındaki ilişkiyi ifade eden doğrusal fonksiyonu bulmaktır.

Basit Doğrusal Regresyon

Anakitle teorik gösterim: $Y = \beta_0 + \beta_1 X + \varepsilon$

Örneklem gerçek değerler: $y_i = b_0 + b_1 x_i + e_i$

Tahmin modeli: $\hat{y}_i = b_0 + b_1 x_i$

β_0 = Doğrunun y eksenini kestiği nokta

β_1 = Doğrunun eğimi

ε = Hata terimi

Basit Doğrusal Regresyon

Örneklem teorik gösterim:

$$y_i = b_0 + b_1 x_i + e_i$$

Tahmin modeli:

$$\hat{y}_i = b_0 + b_1 x_i$$

Hatalar/artıklar:

$$e_i = y_i - \hat{y}_i$$

$$e_i = y_i - b_0 + b_1 x_i$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SSE = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

Basit Doğrusal Regresyon

$\hat{y}_i$

$$SSE = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2$$

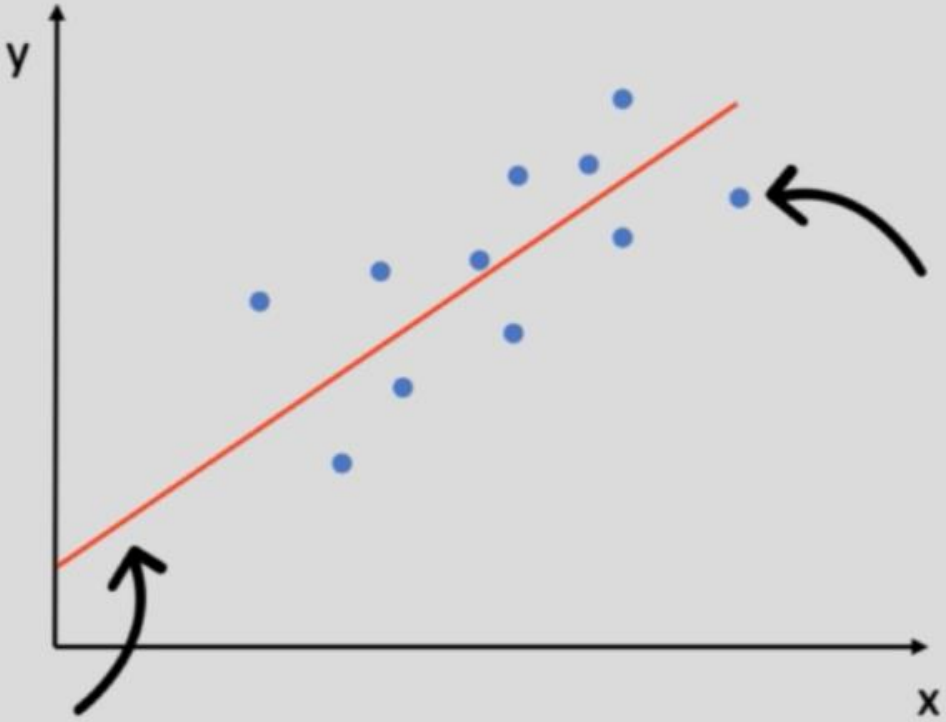
$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$b_0 = \overline{y} - b_1\overline{x}$$

Bağımsız değişkenin ortalaması

Bağımlı değişkenin ortalaması

Basit Doğrusal Regresyon Geometrik Gösterim



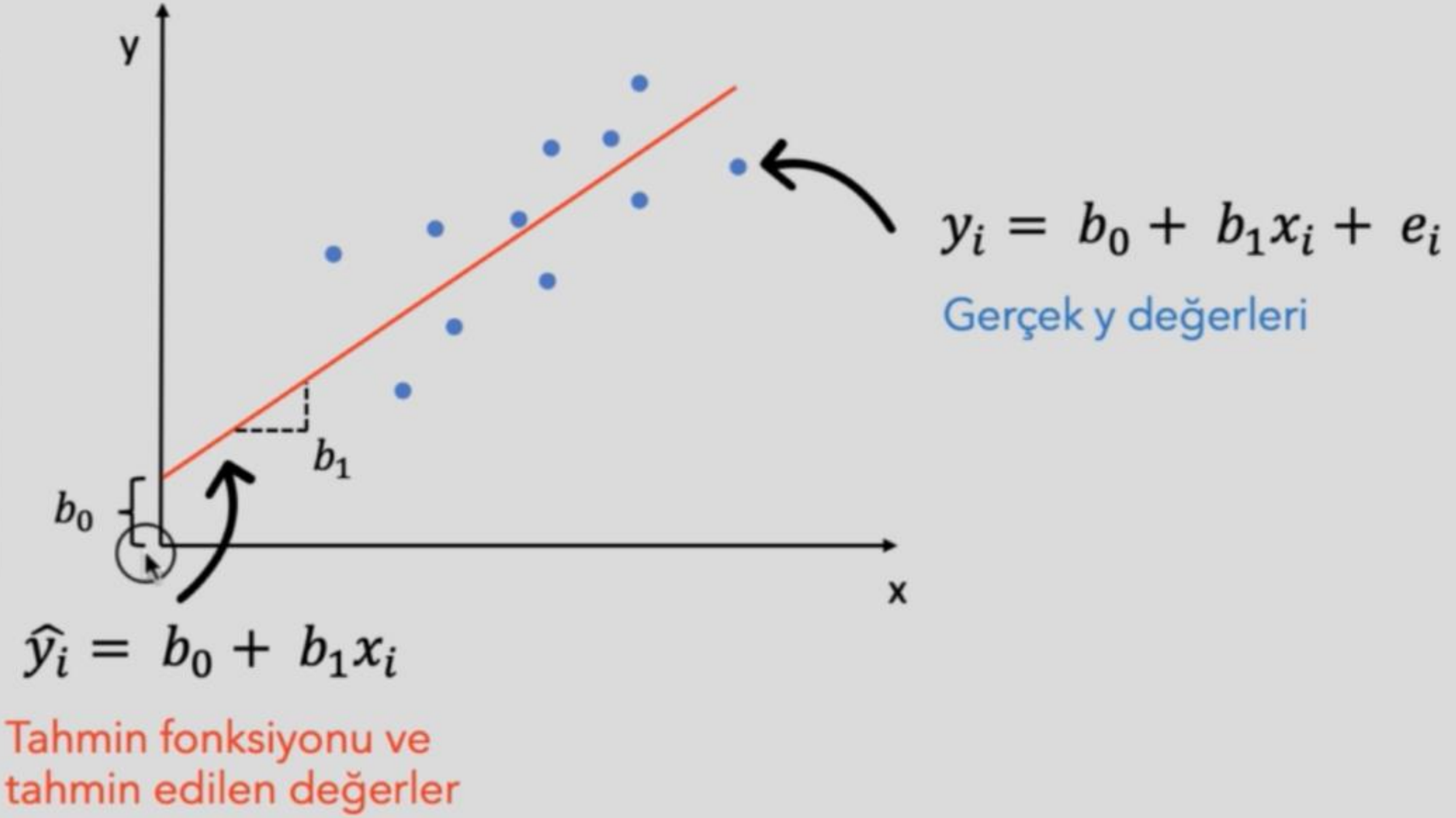
$$y_i = b_0 + b_1 x_i + e_i$$

Gerçek y değerleri

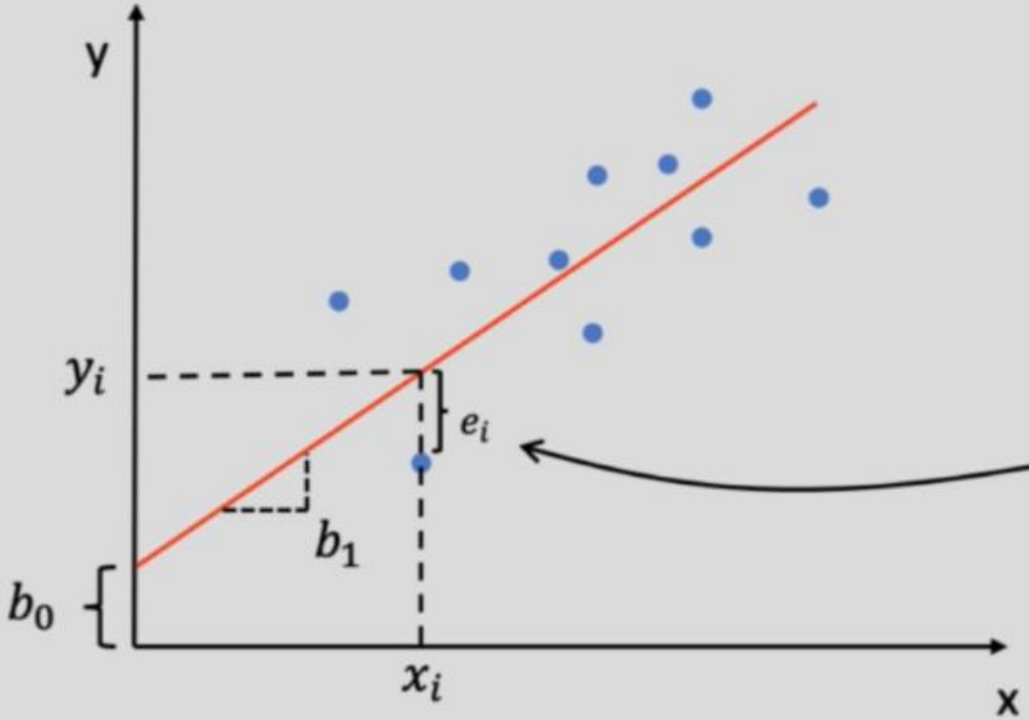
$$\hat{y}_i = b_0 + b_1 x_i$$

Tahmin fonksiyonu ve
tahmin edilen değerler

Basit Doğrusal Regresyon Geometrik Gösterim



Basit Doğrusal Regresyon Geometrik Gösterim



$$y_i = b_0 + b_1 x_i + e_i$$

$$\hat{y}_i = b_0 + b_1 x_i$$

Çoklu Doğrusal Regresyon

Temel amaç, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ifade eden doğrusal fonksiyonu bulmaktır.

Çoklu Doğrusal Regresyon

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{\beta} = (X^T.X)^{-1} X^T.Y$$

Çoklu Doğrusal Regresyon

```
Call:
lm(formula = Sales ~ TV + Radio, data = caseStudyData)

Residuals:
    Min      1Q  Median      3Q     Max
-8.7977 -0.8752  0.2422  1.1708  2.8328

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.92110    0.29449   9.919   <2e-16 ***
TV           0.04575    0.00139  32.909   <2e-16 ***
Radio        0.18799    0.00804  23.382   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.681 on 197 degrees of freedom
Multiple R-squared:  0.8972,	Adjusted R-squared:  0.8962
F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

Çoklu Doğrusal Regresyon

Doğrusal Regresyonun Varsayımları

- Hatalar normal dağılır.
- Hatalar birbirinden bağımsızdır ve aralarında otokorelasyon yoktur.
- Her bir gözlem için hata terimleri varyansları sabittir.
- Değişkenler ile hata terimi arasında ilişki yoktur.
- Bağımsız değişkenler arasında çoklu doğrusal ilişki problemi yoktur.

Çoklu Doğrusal Regresyon

Regresyon Modellerinin Avantaj ve Dezavantajları

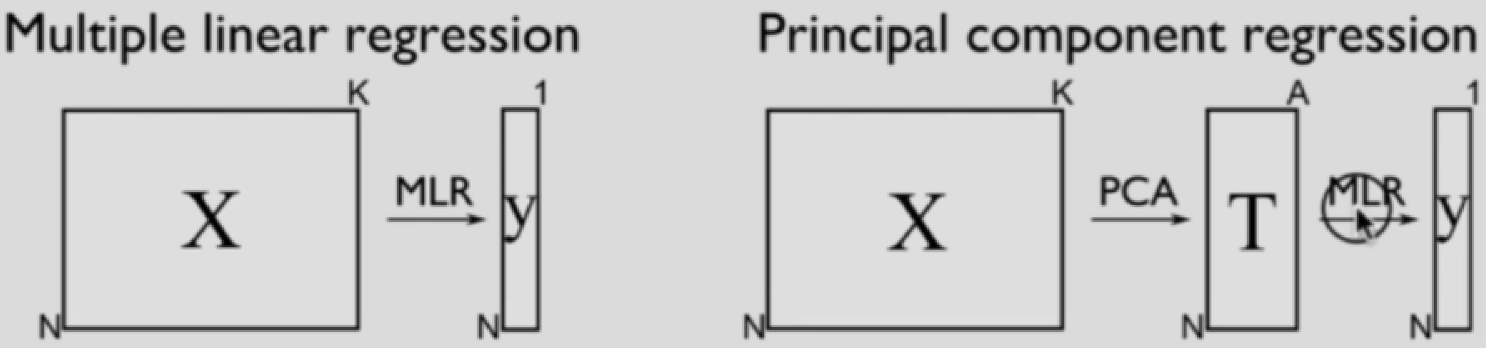
- ✓ İyi anlaşılırsa diğer tüm ML ve DL konuları çok rahat kavranır.
- ✓ Doğrusallık nedensellik yorumları yapılabilmesini sağlar, bu durum aksiyoner ve stratejik modelleme imkanı verir.
- ✓ Değişkenlerin etki düzeyleri ve anlamlılıkları değerlendirilebilir.
- ✓ Bağımlı değişkendeki değişkenliğin açıklanma başarısı ölçülebilir.
- ✓ Model anlamlılığı değerlendirilebilir.
- ❖ Varsayımları vardır.
- ❖ Aykırı gözlemlere duyarlıdır.

PCR – Temel Bileşen Regresyonu

Değişkenlere boyut indirgeme uygulandıktan sonra çıkan bileşenlere regresyon modeli kurulması fikrine dayanır.

(Massy 1965)

PCR - Temel Bileşen Regresyonu

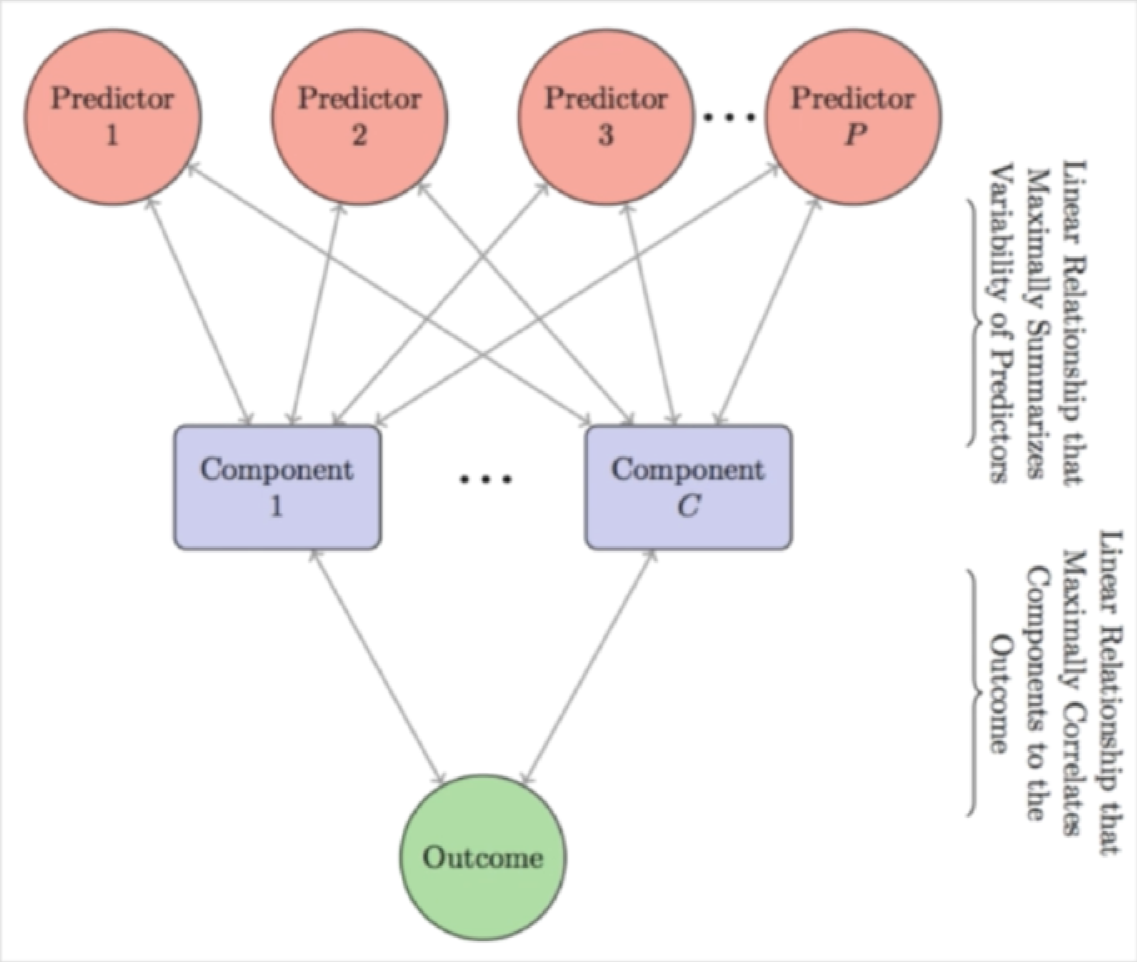


PLS - Kısmi En Küçük Kareler Regresyonu

Değişkenlerin daha az sayıda ve aralarında çoklu doğ. bağlantı problemi olmayan bileşenlere indirgenip regresyon modeli kurulması fikrine dayanır.

Herman Wold (1966, 1982)

PLS - Kısmi En Küçük Kareler Regresyonu



PLS - Kısmi En Küçük Kareler Regresyonu

- Çok boyutluluk laneti $p > n$
- Çoklu doğrusal bağlantı problemi
- PLS de PCR gibi bağımsız değişkenlerin doğrusal kombinasyonlarını bulur. Bu doğrusal kombinasyonlar bileşen ya da latent değişken olarak adlandırılır.
- PLS NIPALS'in özel bir halidir, iteratif olarak bağımlı değişken ile
  yüksek korelasyona sahip değişenler arasındaki gizil (latent) ilişkiyi bulmaya çalışır.

PCR vs PLS

- PCR'da doğrusal kombinasyonlar yani bileşenler **bağımsız değişken uzağındaki değişkenliği** maksimum şekilde özetleyecek şekilde oluşturulur.
- Bu durum bağımlı değişkeni açıklama yeteneği olmamasına sebep olmakta.
- PLS'te ise **bileşenler bağımlı değişken ile olan kovaryansı** maksimum şekilde özetleyecek şekilde oluşturulur.
- Değişkenler atılmak istenmiyorsa ve açıklanabilirlik aranıyorsa: PLS
- PLS, gözetimli boyut indirgeme prosedürü, PCR gözetimsiz boyut indirgeme prosedürü olarak görülebilir.
- İki yöntemin de bir tunning parametresi vardır o da bileşen sayısıdır.
- Optimum bileşen sayısını belirlemek için CV yöntemi kullanılır.

Ridge Regresyon

Amaç hata kareler toplamını minimize eden katsayıları bu katsayılara bir ceza uygulayarak bulmaktır

Hoerl & Kennard 1970

Ridge Regresyon

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Ridge Regresyon

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

$$SSE_{L_2} = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \boxed{\lambda} \underbrace{\sum_{j=1}^{P} {\beta_j}^2}_{\text{Ceza Terimi}}$$

Ayar Parametresi Lambda

Ridge Regresyon

- Aşırı öğrenmeye karşı dirençli.
- Yanlıdır fakat varyansı düşüktür. (Bazen yanlı modelleri daha çok tercih ederiz.)
- Çok fazla paremetre olduğunda EKK'ya göre daha iyidir.
- Çok boyutluluk lanetine karşı çözüm sunar.
- Çoklu doğrusal bağlantı problemi olduğunda etkilidir.
- Tüm değişkenler ile model kurar. İlgisiz değişkenleri modelden çıkarmaz, katsayılarını sıfıra yaklaştırır.
- λ kritik roldedir. İki terimin (formüldeki) göreceli etkilerini kontrol etmeyi sağlar.
- λ için iyi bir değer bulunması önemlidir. Bunun için CV yöntemi kullanılır.

Lasso Regresyon

Amaç hata kareler toplamını minimize eden katsayıları bu katsayılara bir ceza uygulayarak bulmaktır

Tibshirani 1996

Lasso Regresyon

$$SSE_{L_1} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \boxed{\lambda}\underbrace{\sum_{j=1}^{P}|\beta_j|}_{\text{Ceza Terimi}}$$

Ayar Parametresi Lambda

Lasso Regresyon

- Ridge regresyonun ilgili-ilgisiz tüm değişkenleri modelde bırakma dezavantajını gidermek için önerilmiştir.
- Lasso'da katsayıları sıfıra yaklaştırır.
- Fakat L1 normu λ yeteri kadar büyük olduğunda bazı katsayıları sıfır yapar. Böylece değişken seçimi yapmış olur.
- λ'nın doğru seçilmesi çok önemlidir, burada da CV kullanılır.
- Ridge ve Lasso yöntemleri birbirinden üstün değildir.

Lasso Regresyon

λ Ayar Parametresinin Belirlenmesi

- λ'nın sıfır olduğu yer EKK'dır. HKT'yi minimum yapan λ'yı arıyoruz
- λ için belirli değerleri içeren bir küme seçilir ve her birisi için cross validation test hatası hesaplanır.
- En küçük cross validation'ı veren λ ayar parametresi olarak seçilir.
- Son olarak seçilen bu λ ile model yeniden tüm gözlemlere fit edilir.

ElasticNet Regresyonu

Amaç hata kareler toplamını minimize eden katsayıları bu katsayılara bir ceza uygulayarak bulmaktır. ElasticNet L1 ve L2 yaklaşımlarını birleştirir.

Zou & Hastie 2005

ElasticNet Regresyonu

$$SSE_{Enet} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^{P}{\beta_j}^2 + \lambda_2 \sum_{j=1}^{P}|\beta_j|$$

ElasticNet Regresyonu

L2 ve L1 Ayar Parametleri

$$SSE_{Enet} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \boxed{\lambda_1}\sum_{j=1}^{P}\beta_j{}^2 + \boxed{\lambda_2}\sum_{j=1}^{P}|\beta_j|$$

Ceza Terimleri