Makine Öğrenmesi

Boyut Küçültme ve Düzenlileştirme

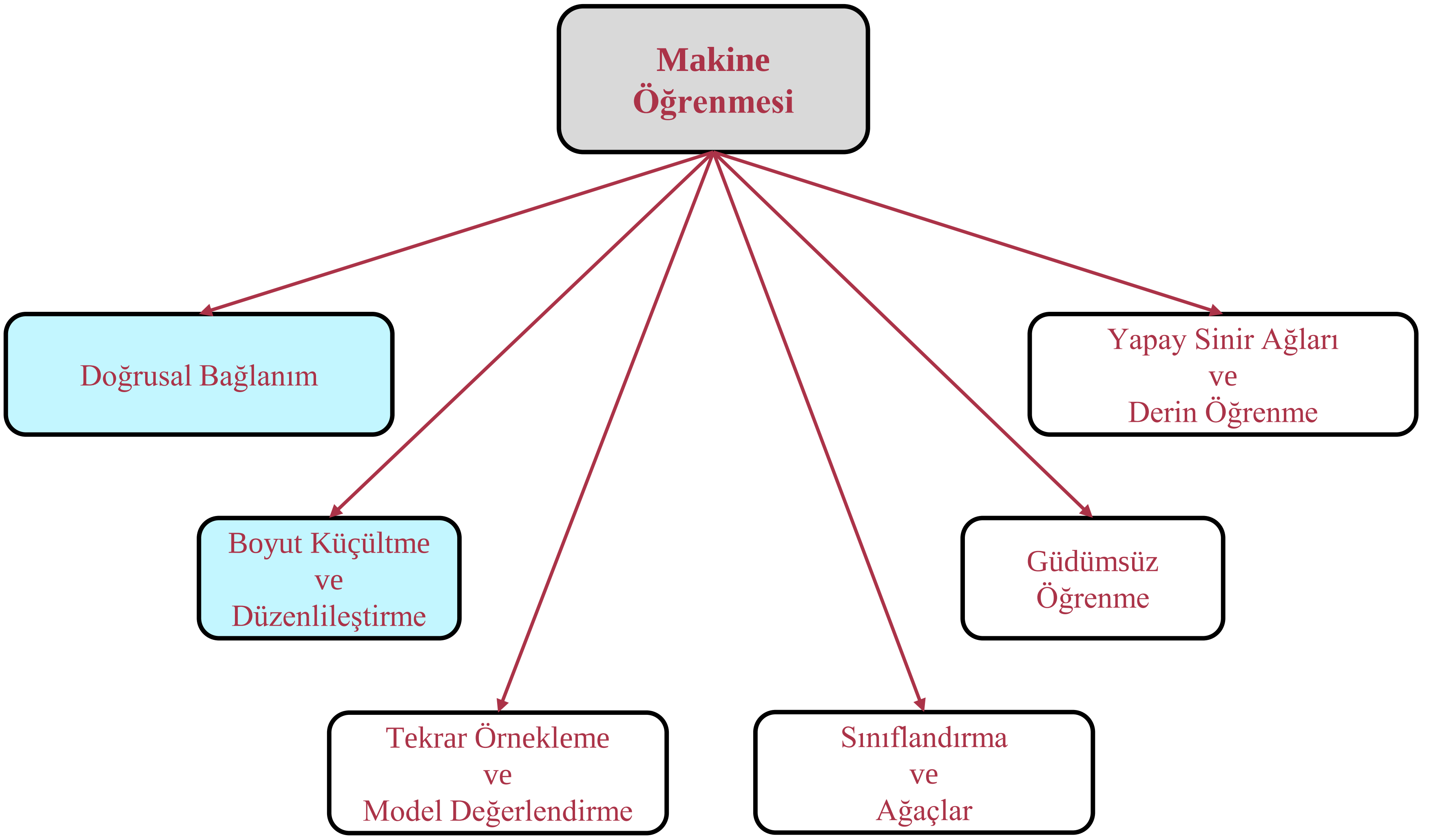
İlker Birbil ve Utku Karaca

Erasmus Üniversitesi Rotterdam

İstanbul'da Makine Öğrenmesi

27 Ocak – 2 Şubat, 2020

EUR
D4C
Bol Bilim
Veri Defteri
@sibirbil



Makine Öğrenmesi
Doğrusal Bağlanım
Yapay Sinir Ağları ve Derin Öğrenme
Boyut Küçültme ve Düzenlileştirme
Güdümsüz Öğrenme
Tekrar Örnekleme ve Model Değerlendirme
Sınıflandırma ve Ağaçlar

Öznitelikler
(Değişkenler)

$$X_1, \cancel{X_2}, X_3, \ldots, \cancel{X_{p-1}}, X_p$$

$$Y \approx \beta_0 + \beta_1 X_1 + \underbrace{\cancel{\beta_2}}_{0} X_2 + \cdots + \underbrace{\cancel{\beta_{p-1}}}_{0} X_{p-1} + \beta_p X_p$$

$\boxed{\textbf{Tahmin İsabetinin Artması}}$ $\boxed{\textbf{Model Yorumunun Kolaylaşması}}$

$n >> p$ $\longrightarrow$ EKK (tek çözüm) $\longrightarrow$ Düşük tahmin varyansı

$p > n$ $\longrightarrow$ EKK (pek çok çözüm) $\longrightarrow$ Tahmin varyansı $\uparrow \infty$

EKK: En Küçük Kareler

Alt Küme Seçimi (Subset Selection)

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

$$X_1, X_2, X_3, \ldots, X_{p-1}, X_p$$

$$\{\ \} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_0 = \{\ \}$$

$$\{1\}, \{2\}, \ldots, \{k\}, \ldots, \{p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_1 = \{k\}$$

$$\{1,2\}, \{1,3\}, \ldots, \{u,v\}, \ldots, \{p-1,p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_2 = \{u,v\}$$

$$\vdots$$

$$\{1,2,\ldots,p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_p = \{1,2,\ldots,p\}$$

2^p alt küme!

KKT: Kalıntı Kareler Toplamı

İleri Adımlı (Forward Stepwise) Alt Küme Seçimi

$$\{\ \} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_0 = \{\ \}$$

$$\{1\}, \{2\}, \ldots, \{\ell\}, \ldots, \{p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_1 = \{\ell\}$$

$$\{\ell, 2\}, \{\ell, 3\}, \ldots, \{\ell, w\}, \ldots, \{\ell, p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_2 = \{\ell, w\}$$

$$\vdots$$

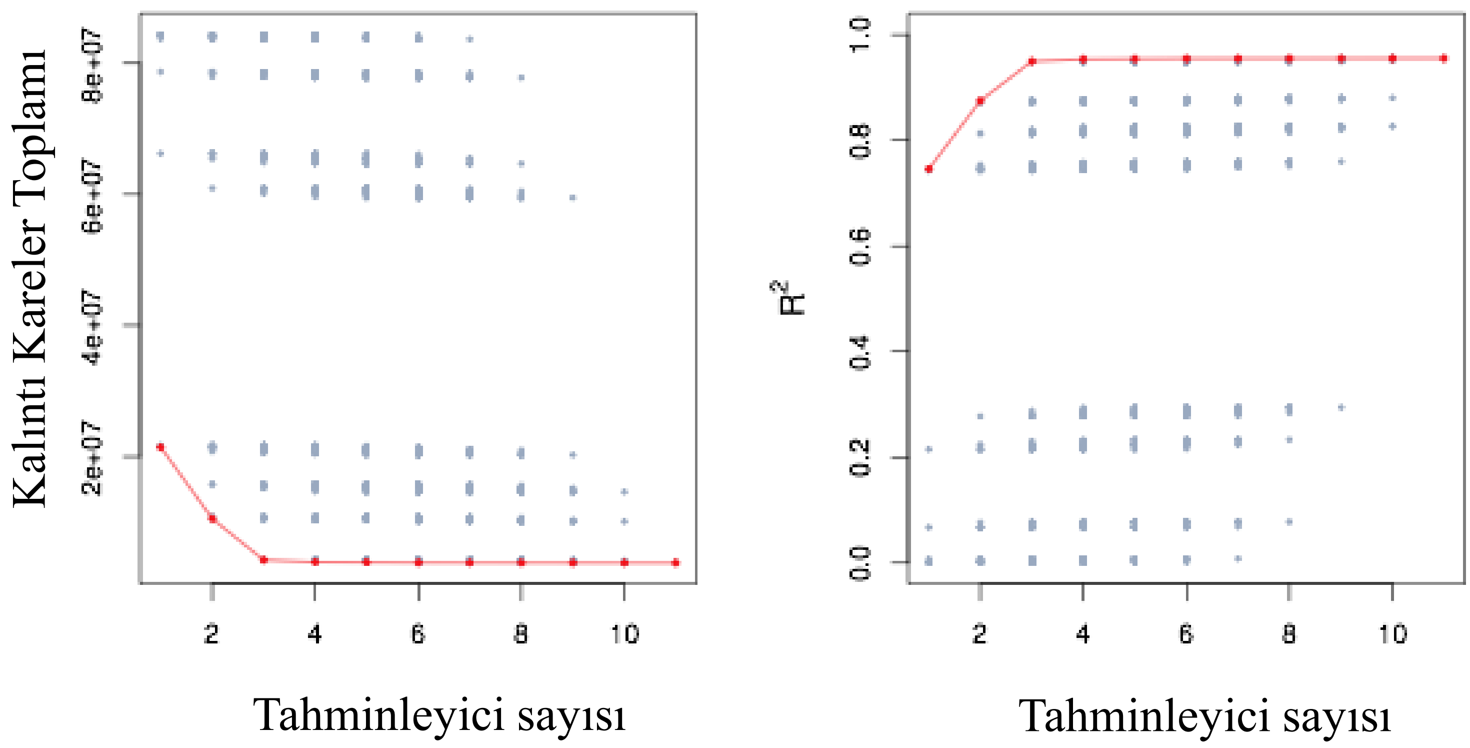
Geri Adımlı (Backward Stepwise) Alt Küme Seçimi

$$\{1, 2, \ldots, p\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_p = \{1, 2, \ldots, p\}$$

$$\{2, 3, \ldots, p\}, \{1, 3, \ldots, p\}, \ldots \{1, 2, \ldots, p-1\} \xrightarrow{\text{KKT} \downarrow} \mathcal{M}_{p-1} = \{1, 3, \ldots, p\}$$

$$\vdots$$

$$\frac{p(p+1)}{2} + 1 \text{ alt küme}$$


Kalıntı Kareler Toplamı
8e+07
6e+07
4e+07
2e+07
2
4
6
8
10
Tahminleyici sayısı
R^2
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
Tahminleyici sayısı

$$\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_p \ ?$$

Değişken Sayısı $\uparrow$

Eğitim Hatası $\downarrow$

Test Hatası $\downarrow \uparrow$?

... aşırı öğrenme ?

EYY ile doğrusal model öğrenme

$$C_p = \frac{1}{n}(\text{KKT} + 2d\hat{\sigma}^2)$$

(Akaike Information Criterion)

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}(\text{KKT} + 2d\hat{\sigma}^2)$$

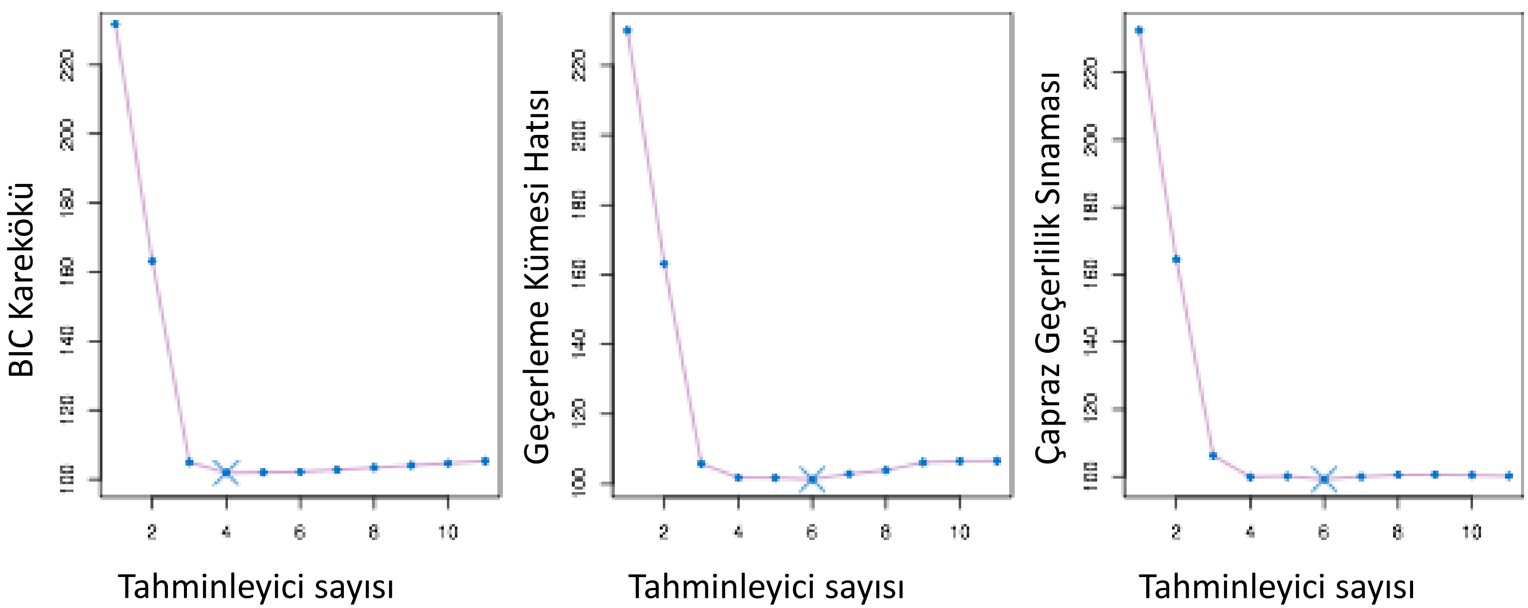
(Bayeasian Information Criterion)

$$\text{BIC} = \frac{1}{n\hat{\sigma}^2}(\text{KKT} + \log(n)d\hat{\sigma}^2)$$

$d = 1, 2, \ldots, p$

en düşük değeri veren *d* $\longrightarrow \mathcal{M}_d$

Çapraz Geçerlilik Sınaması (Cross Validation)


BIC Karekökü
Geçerleme Kümesi Hatısı
Çapraz Geçerlilik Sınaması
Tahminleyici sayısı
Tahminleyici sayısı
Tahminleyici sayısı

Hatırlatma: Doğrusal Bağlanım

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^\intercal \\ 1 & x_2^\intercal \\ \vdots & \vdots \\ 1 & x_n^\intercal \end{bmatrix}_{n\times(p+1)}$$

$$\mathbf{y}^\intercal = (y_1, \ldots, y_n)$$

$$\boldsymbol{\beta}^\intercal = (\beta_0, \beta_1, \ldots, \beta_p)$$

$$\hat{\boldsymbol{\beta}}_{\text{LS}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Tam kerte (full rank) varsayımı ile

$$\hat{\boldsymbol{\beta}}_{\text{LS}} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$$

$$\boxed{\hat{y}_0 = (1\ x_0^\intercal)\hat{\boldsymbol{\beta}}_{\text{LS}}}$$

Düzenlileştirme (Regularization)

$$Y \approx \beta_0 + \beta_1 X_1 + \underset{0}{\cancel{\beta_2}} X_2 + \cdots + \underset{0}{\cancel{\beta_{p-1}}} X_{p-1} + \beta_p X_p$$

Çıkıntı Bağlanımı (Ridge Regression)

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

λ : ayarlama (tuning) parametresi

Lasso

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

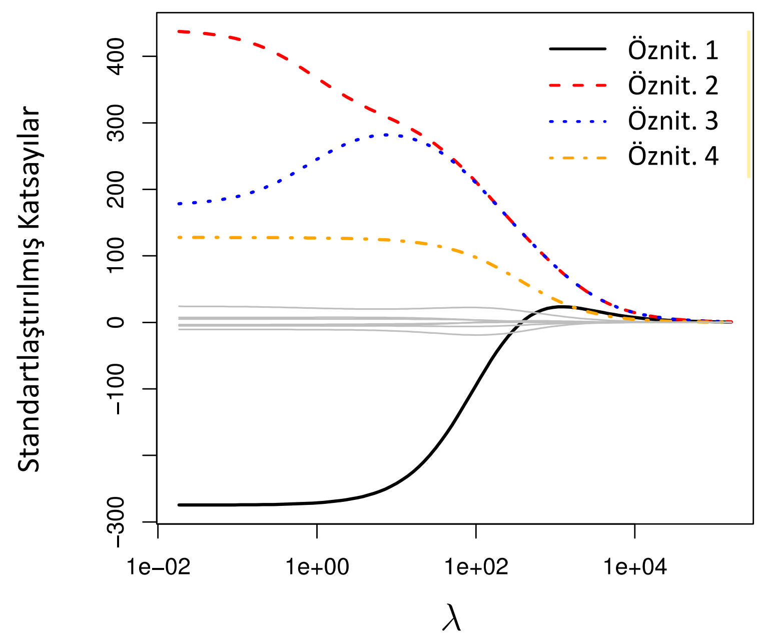
Çıkıntı Bağlanımı

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

$$\lambda \uparrow \infty \quad \longrightarrow \quad \beta_j \downarrow 0, \quad j = 1, \cdots, p$$

Standartlaştırma

$$\hat{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_{ij})^2}}$$



$$\hat{\beta}_{\mathrm{R}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\intercal \boldsymbol{\beta}$$

$$\hat{\beta}_{\mathrm{LS}} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y}$$

$$\boxed{\hat{\beta}_{\mathrm{R}} = (\mathbf{X}^\intercal \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\intercal \mathbf{y}}$$

β_0 hariç
(ortalanmış girdi)

$$\mathbf{X} = \begin{bmatrix} x_1^\intercal \\ x_2^\intercal \\ \vdots \\ x_n^\intercal \end{bmatrix}_{n \times p}$$

Tekil Değer Çözüşümü (SVD)

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\intercal$$

$$\mathbf{U} = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_p]$$

$$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$$

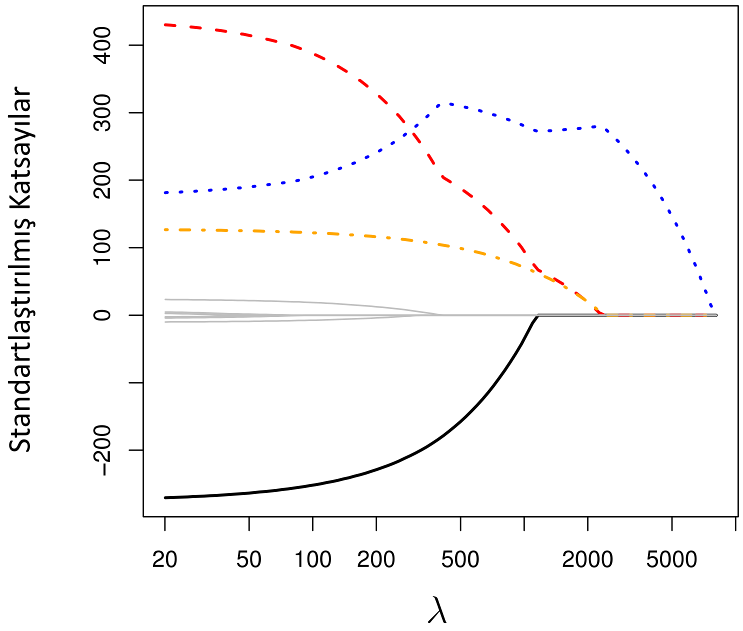
$$\begin{aligned} \mathbf{X}\hat{\beta}_{\mathrm{LS}} &= \mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y} \\ &= \mathbf{U}\mathbf{U}^\intercal \mathbf{y} \\ &= \sum_{j=1}^{p} \mathbf{u}_j \mathbf{u}_j^\intercal \mathbf{y} \end{aligned}$$

$$\begin{aligned} \mathbf{X}\hat{\beta}_{\mathrm{R}} &= \mathbf{X}(\mathbf{X}^\intercal \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\intercal \mathbf{y} \\ &= \mathbf{U}\mathbf{D}(\mathbf{D}^2 + \lambda \mathbf{I})^{-1} \mathbf{D}\mathbf{U}^\intercal \mathbf{y} \\ &= \sum_{j=1}^{p} \mathbf{u}_j \frac{d_j^2}{d_j^2 + \lambda} \mathbf{u}_j^\intercal \mathbf{y} \end{aligned}$$

Lasso

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|$$

$$\lambda\uparrow\infty \quad\longrightarrow\quad \beta_j\downarrow 0,\quad j=1,\cdots,p$$



$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = (\mathbf{X}^\intercal \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\intercal \mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_L = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1$$

β_0 hariç
(ortalanmış girdi)

$$\mathbf{X} = \begin{bmatrix} x_1^\intercal \\ x_2^\intercal \\ \vdots \\ x_n^\intercal \end{bmatrix}_{n \times p}$$

$\downarrow$

Dışbükey Eniyileme

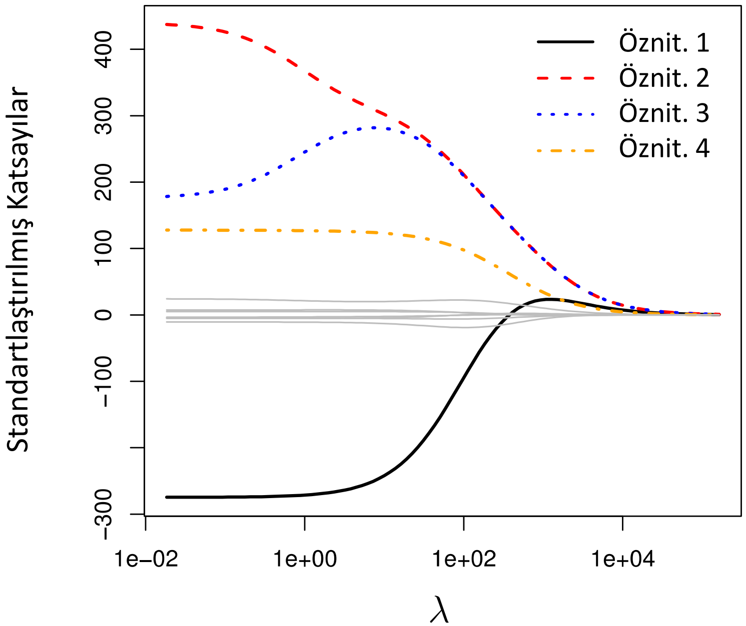
$\downarrow$

Analitik çözümü yok

(çok hızlı çözüm yöntemleri mevcut)

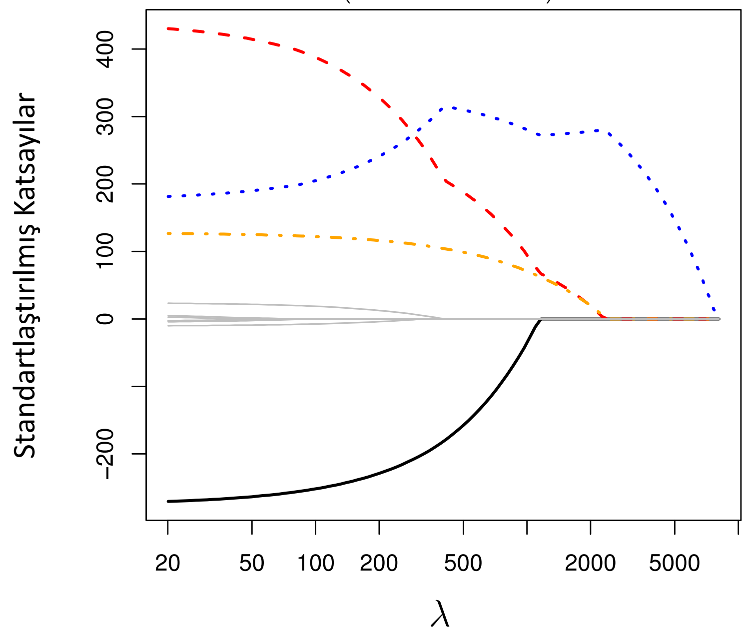
Çıkıntı Bağlanımı

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$



Lasso

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$



Örnek

$$\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix} = \begin{bmatrix}0.01\\ -0.01\\ 0.01\end{bmatrix}$$

$$\sum_{j=1}^{3}\beta_j^2 = 3(0.01)^2 = 0.0003$$

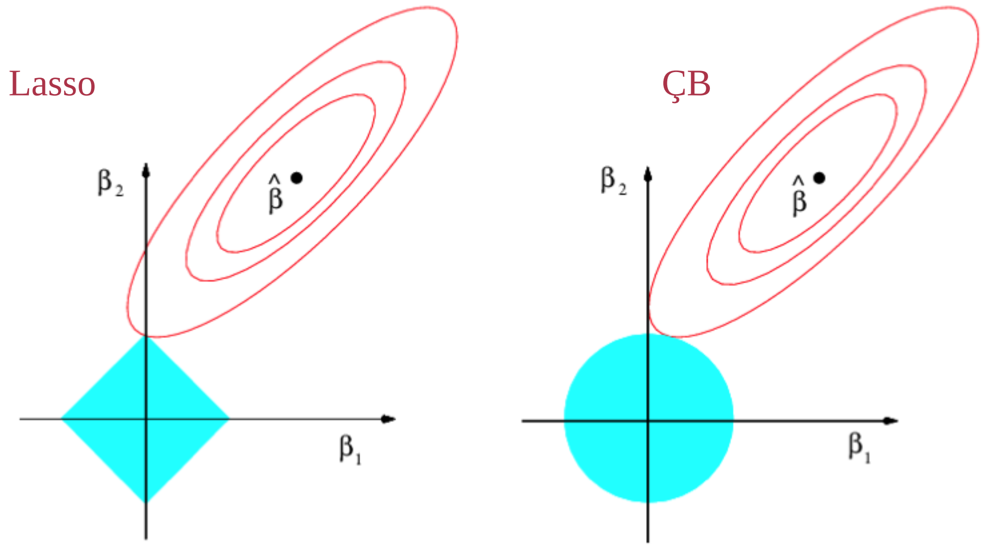
$$\sum_{j=1}^{3}|\beta_j| = 3(0.01) = 0.03$$

Çıkıntı Bağlanımı (ÇB)

$$\min_{\boldsymbol{\beta}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \ : \ \|\boldsymbol{\beta}\|_2^2 \leq \Delta \right\}$$

Lasso

$$\min_{\boldsymbol{\beta}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \ : \ \|\boldsymbol{\beta}\|_1 \leq \Delta \right\}$$



En Küçük Kareler Çözümü: $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$

Çıkıntı Bağlanımı – Lasso

- Seyrek (bol sıfırlı) çözüm: Lasso
- Çok ve birbirlerine yakın parametreler: Çıkıntı bağlanımı
- Yorumlama kolaylığı: Lasso

- Ayarlama parametresi (λ) seçimi:

Izgara (grid) arama **ve** çapraz geçerlilik sınaması

λ ↑ Model Karmaşıklığı ↓ Varyans ↓ Yanlılık ↑

Olasılıksal Çıkarsama (Probabilistic Inference)

$y_i \sim N(\hat{y}_i, \sigma^2), \;\; i = 1, \ldots, n$ ve b.ö.d.

$$\boxed{\mathbb{P}(\mathbf{y}|\boldsymbol{\beta}) = \prod_{i=1}^{n} P(y_i|\boldsymbol{\beta})}$$

$$\mathbb{P}(\boldsymbol{\beta}|\mathbf{y}) = \frac{\mathbb{P}(\mathbf{y}|\boldsymbol{\beta})\mathbb{P}(\boldsymbol{\beta})}{\mathbb{P}(\mathbf{y})}$$

$$\mathbb{P}(\boldsymbol{\beta}|\mathbf{y}) \propto \mathbb{P}(\mathbf{y}|\boldsymbol{\beta})\mathbb{P}(\boldsymbol{\beta})$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{MAP}} &= \arg\max_{\boldsymbol{\beta}} \;\; \mathbb{P}(\mathbf{y}|\boldsymbol{\beta})\mathbb{P}(\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \;\; \log \mathbb{P}(\mathbf{y}|\boldsymbol{\beta}) + \log \mathbb{P}(\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \;\; \sum_{i=1}^{n} \log \mathbb{P}(y_i|\boldsymbol{\beta}) + \log \mathbb{P}(\boldsymbol{\beta})
\end{aligned}$$

sabit önsel , $\mathbb{P}(\boldsymbol{\beta})$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\bullet} &= \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log \mathbb{P}(y_i|\boldsymbol{\beta}) + \log \mathbb{P}(\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log \mathbb{P}(y_i|\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2}{2\sigma^2}} \right) \\
&= \arg\max_{\boldsymbol{\beta}} \; -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 \\
&= \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2
\end{aligned}$$

$$\Longrightarrow \; \hat{\boldsymbol{\beta}}_{\bullet} = \hat{\boldsymbol{\beta}}_{\mathrm{LS}}$$

normal dağılmış önsel , $\beta_j \sim N(0, \psi^2),\ j = 1, \ldots, p$ ve b.ö.d.

$$\mathbb{P}(\boldsymbol{\beta}) = \prod_{j=1}^{p} P(\beta_j)$$

β_0 hariç
(ortalanmış veri)

$$\mathbf{X} = \begin{bmatrix} x_1^{\intercal} \\ x_2^{\intercal} \\ \vdots \\ x_n^{\intercal} \end{bmatrix}_{n \times p}$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\bullet} &= \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log \mathbb{P}(y_i | \boldsymbol{\beta}) + \log \mathbb{P}(\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2}{2\sigma^2}} \right) + \sum_{j=1}^{p} \log \left(\frac{1}{\psi\sqrt{2\pi}} e^{-\frac{\beta_j^2}{2\psi^2}} \right) \\
&= \arg\min_{\boldsymbol{\beta}} \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \frac{\sigma^2}{\psi^2} \sum_{j=1}^{p} \beta_j^2 \right) \\
&= \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\intercal} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{\sigma^2}{\psi^2} \boldsymbol{\beta}^{\intercal} \boldsymbol{\beta}
\end{aligned}$$

$$\implies \hat{\boldsymbol{\beta}}_{\bullet} = \hat{\boldsymbol{\beta}}_{\mathrm{R}} \text{ with } \lambda = \frac{\sigma^2}{\psi^2}$$

b.ö.d.:bağımsız özdeşçe dağılmış

Laplace dağılmış önsel $\beta_j \sim \text{Laplace}(0, \phi), \;\; j = 1, \ldots, p$ ve b.ö.d.

$$\hat{\boldsymbol{\beta}}_{\bullet} = \arg\max_{\boldsymbol{\beta}} \;\; \sum_{i=1}^{n} \log \mathbb{P}(y_i | \boldsymbol{\beta}) + \log \mathbb{P}(\boldsymbol{\beta})$$

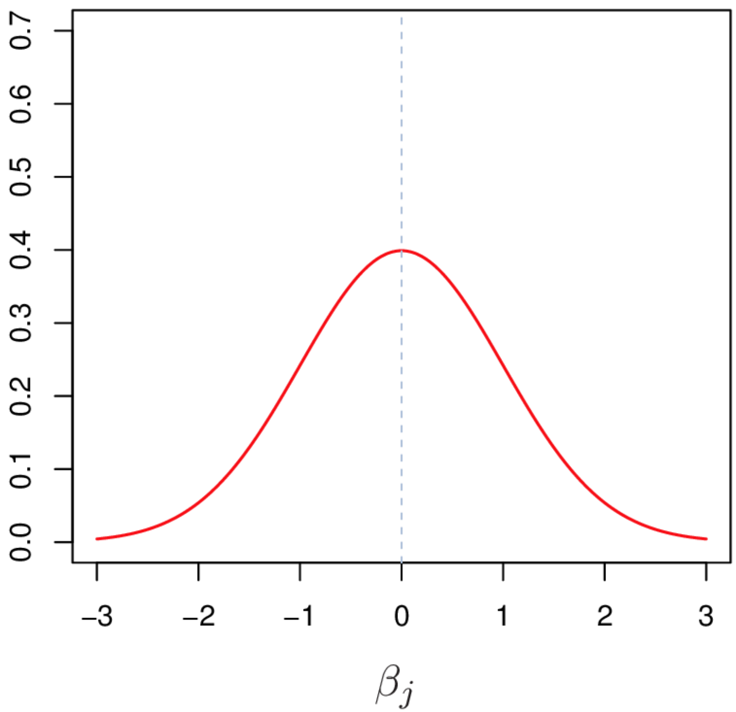
$$= \arg\max_{\boldsymbol{\beta}} \;\; \sum_{i=1}^{n} \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2}{2\sigma^2}} \right) + \sum_{j=1}^{p} \log \left(\frac{1}{2\phi} e^{-\frac{|\beta_j|}{2\phi}} \right)$$

$$= \arg\min_{\boldsymbol{\beta}} \;\; \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \frac{\sigma^2}{\phi} \sum_{j=1}^{p} |\beta_j| \right)$$

$$= \arg\min_{\boldsymbol{\beta}} \;\; (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{\sigma^2}{\phi} \|\boldsymbol{\beta}\|_1$$

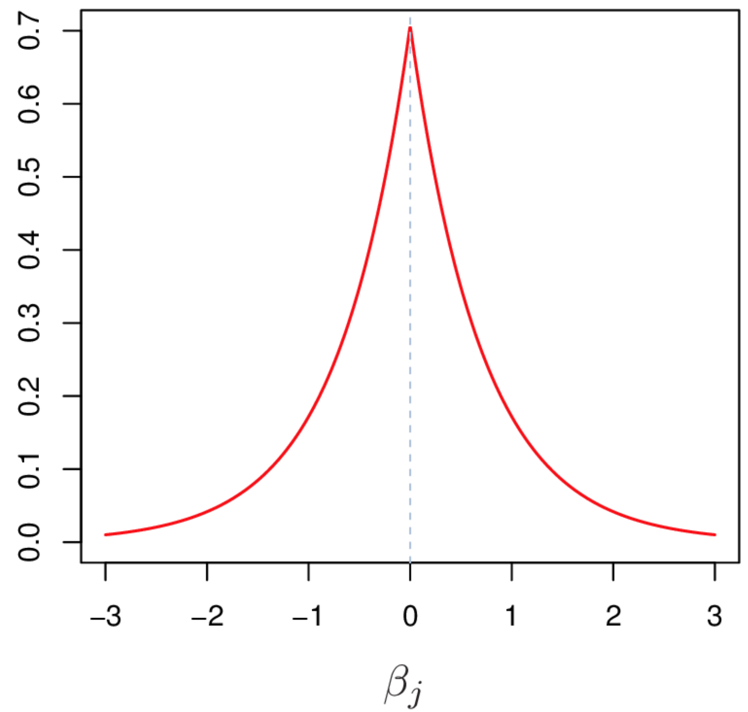
$$\implies \hat{\boldsymbol{\beta}}_{\bullet} = \hat{\boldsymbol{\beta}}_{\mathrm{L}} \text{ with } \lambda = \frac{\sigma^2}{\phi}$$

$$\beta_j \sim N(0, \psi^2)$$



$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta}$$

$$\beta_j \sim \mathrm{Laplace}(0, \phi)$$

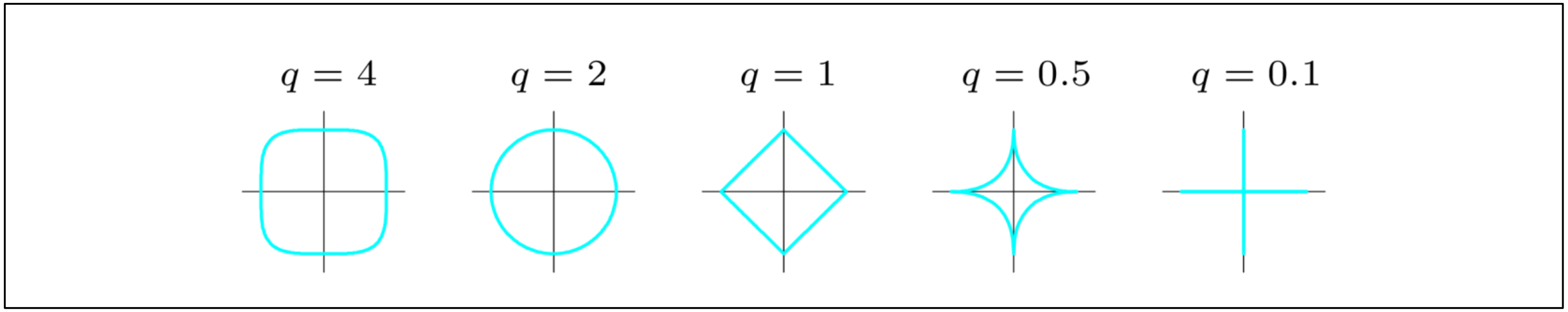


$$\hat{\boldsymbol{\beta}}_{L} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathsf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1$$

Elastic Net

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q$$

$$q > 0$$



*

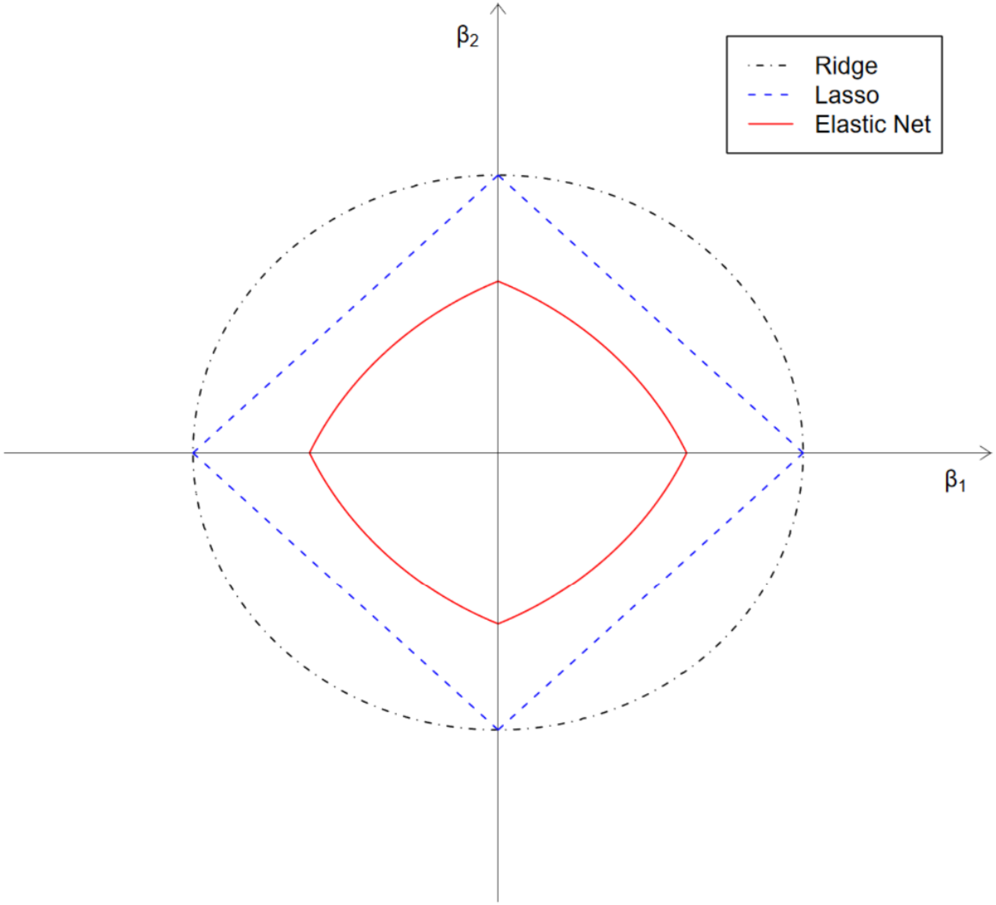
$$\hat{\boldsymbol{\beta}}_{\mathrm{EN}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1-\alpha)|\beta_j|)$$

$$0 \leq \alpha \leq 1$$

* *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, T. Hastie, R. Tibshirani, J. Friedman, Second Edition, Springer, 2009, pg 72.

Elastic Net maliyet fonksiyonunun geometrisi

*



Regularization and variable selection via the elastic net, H. Zou ve T. Hastie, 2005, sf. 5.
(www, son erişim tarihi 20 Ağustos 2019)

Pratikte

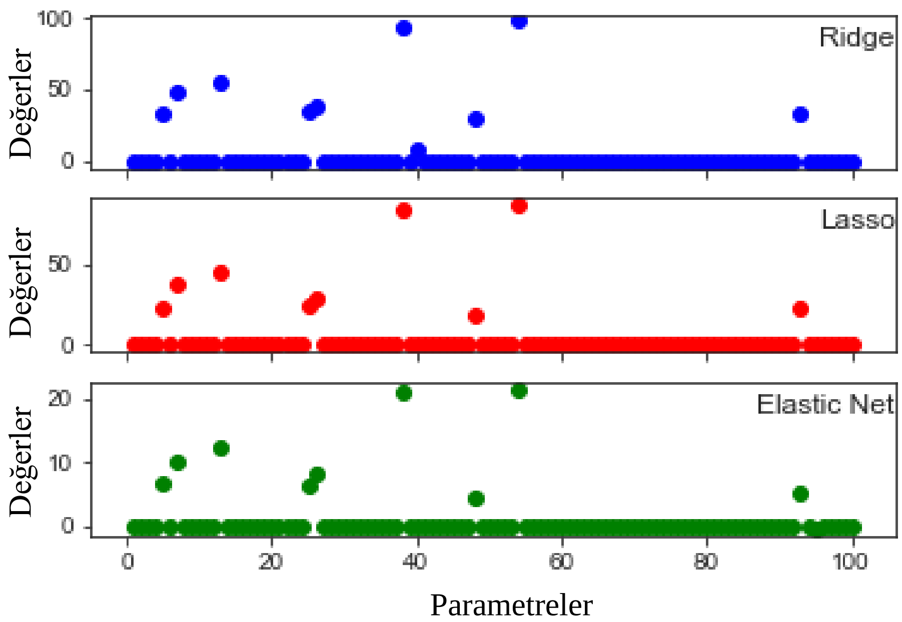
```python
from sklearn.datasets.samples_generator import make_regression
from sklearn.linear_model import Lasso
from sklearn.linear_model import Ridge
from sklearn.linear_model import ElasticNet

X, y = make_regression(n_samples=2000, n_features=100, random_state=0)

rdg = Ridge(alpha=10)
lss = Lasso(alpha=10)
eln = ElasticNet(alpha=10, l1_ratio=0.7)

rdg.fit(X, y)
lss.fit(X, y)
eln.fit(X, y)
```

paketler ve fonksiyonlar

sentetik problem üretimi

üç modelin hazırlanışı

üç modelin uydurumu



Pratikte

```python
import numpy as np
from sklearn.datasets.samples_generator import make_regression
from sklearn.linear_model import Lasso
from sklearn.linear_model import Ridge
from sklearn.linear_model import ElasticNet
```

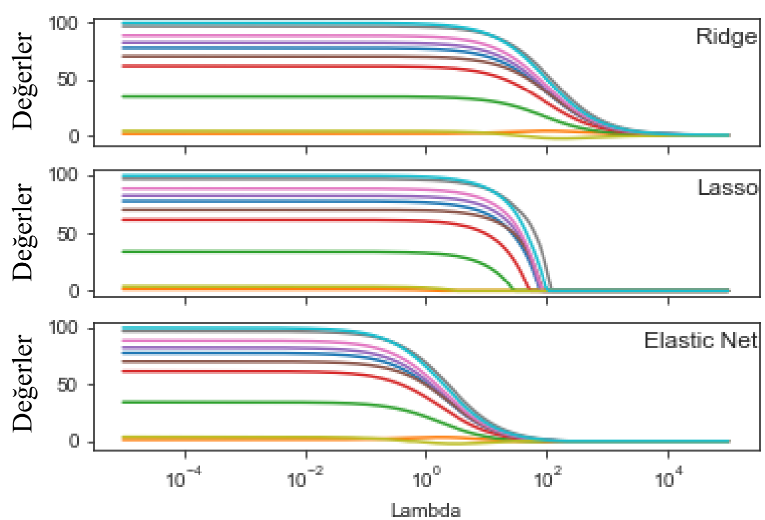
paketler ve fonksiyonlar

```python
X, y = make_regression(n_samples=100, n_features=10, random_state=0, bias=3.5)

rdg = Ridge()
lss = Lasso()
eln = ElasticNet()
```

üç modelin hazırlanışı

```python
rdg_coefs = []; lss_coefs = []; eln_coefs = [];
alphas = np.logspace(-5, 5, 200)
for a in alphas:
    rdg.set_params(alpha=a)
    lss.set_params(alpha=a)
    eln.set_params(alpha=a)
    rdg.fit(X, y); lss.fit(X, y); eln.fit(X, y);
    rdg_coefs.append(rdg.coef_)
    lss_coefs.append(lss.coef_)
    eln_coefs.append(eln.coef_)
```

değişken maliyet fonksiyonu parametreleri



Boyut Küçültme

$$X_1, X_2, \ldots, X_p$$

$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j, \; m = 1, 2, \ldots, M \qquad (M < p)$$

$$Y \approx \theta_0 + \theta_1 Z_1 + \theta_2 Z_2 + \cdots + \theta_M Z_M$$

Örnek

$$p = 4, M = 2$$

$$Z_1 = \phi_{11} X_1 + \phi_{21} X_2 + \phi_{31} X_3 + \phi_{41} X_4$$

$$Z_2 = \phi_{12} X_1 + \phi_{22} X_2 + \phi_{32} X_3 + \phi_{42} X_4$$

$$Y \approx \theta_0 + \theta_1 Z_1 + \theta_2 Z_2$$

İzdüşüm (Projection)

$$X_1, X_2, ..., X_p$$

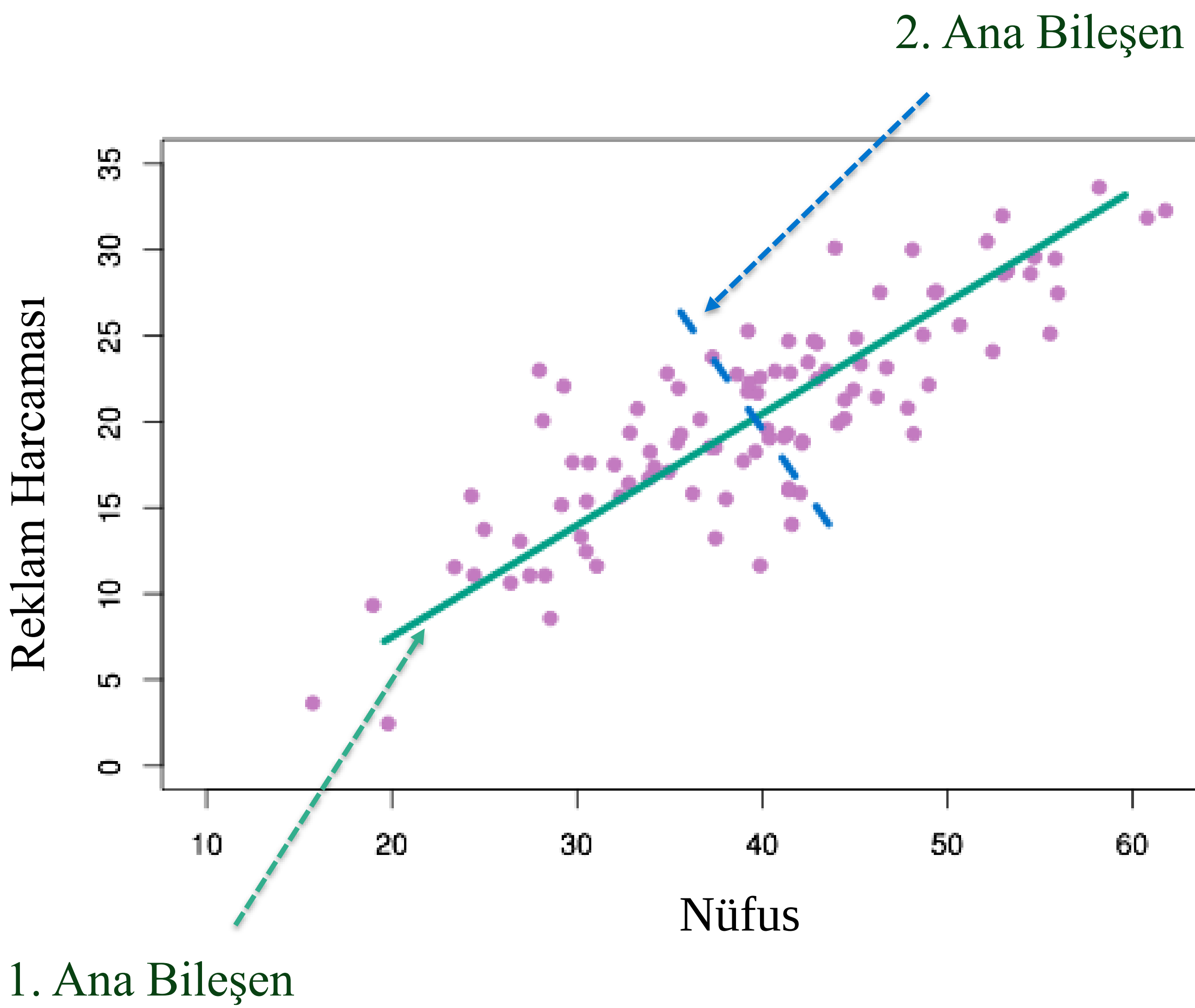
$$\Big\downarrow ?$$

$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j, \;\; m = 1, 2, \ldots, M$$

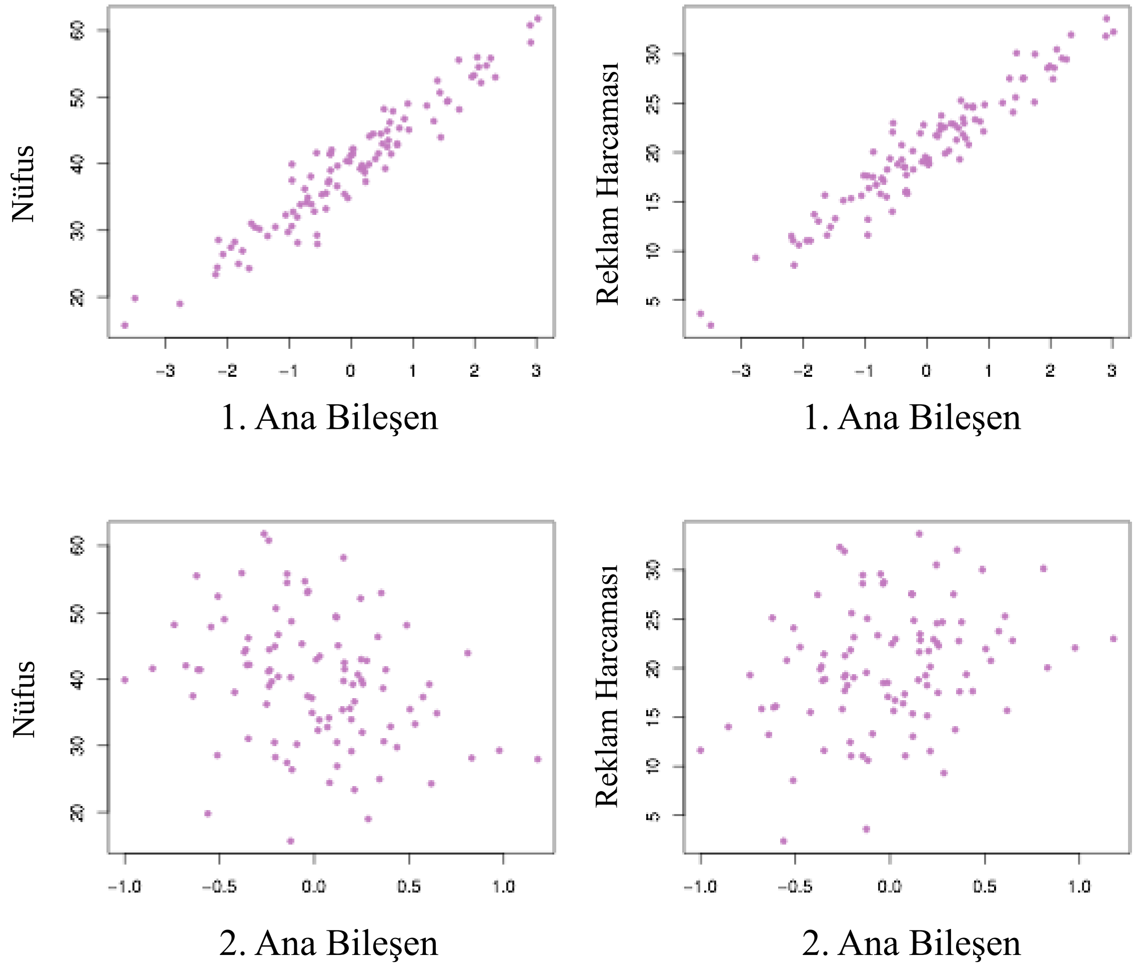
Ana Bileşenler Analizi (Principal Component Analysis)

Öznitelik seçmek için kullanılmaz!

Görsel olarak Ana Bileşenler Analizi



Görsel olarak Ana Bileşenler Analizi



Ana Bileşenler Bağlanımı (ABB)

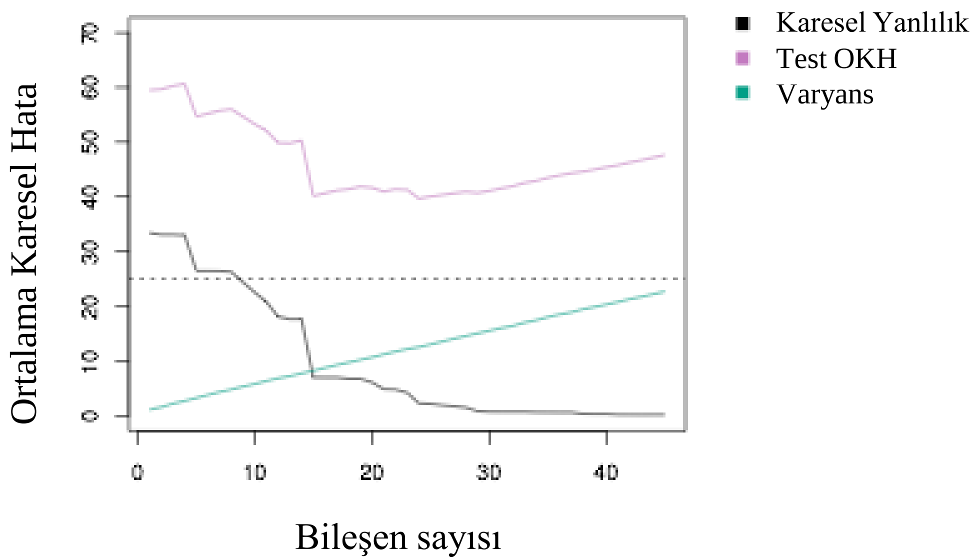
ABB Uygulaması

$$Y \approx \theta_0 + \theta_1 Z_1 + \theta_2 Z_2 + \cdots + \theta_M Z_M$$

EKK Uygulaması

Standartlaştırma uygulaması (çıkıntı bağlanımı gibi)

M seçimi için çapraz geçerlilik sınaması uygulanabilir



Pratikte

```python
import numpy as np
from sklearn.datasets import load_diabetes
from sklearn.decomposition import PCA
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import scale
```

paketler ve fonksiyonlar

```python
X, y = load_diabetes(return_X_y=True)
```

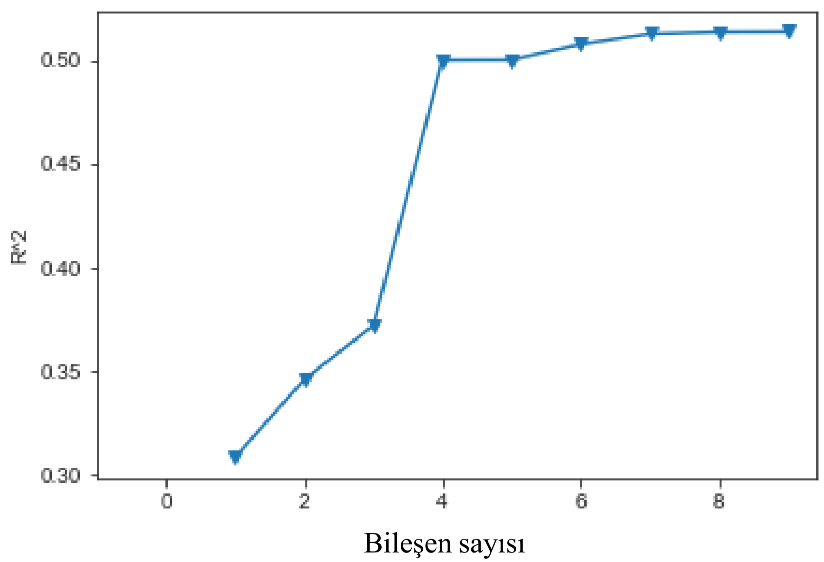
diabetes verisinin yüklenmesi

```python
pca = PCA()

Z = pca.fit_transform(scale(X))
```

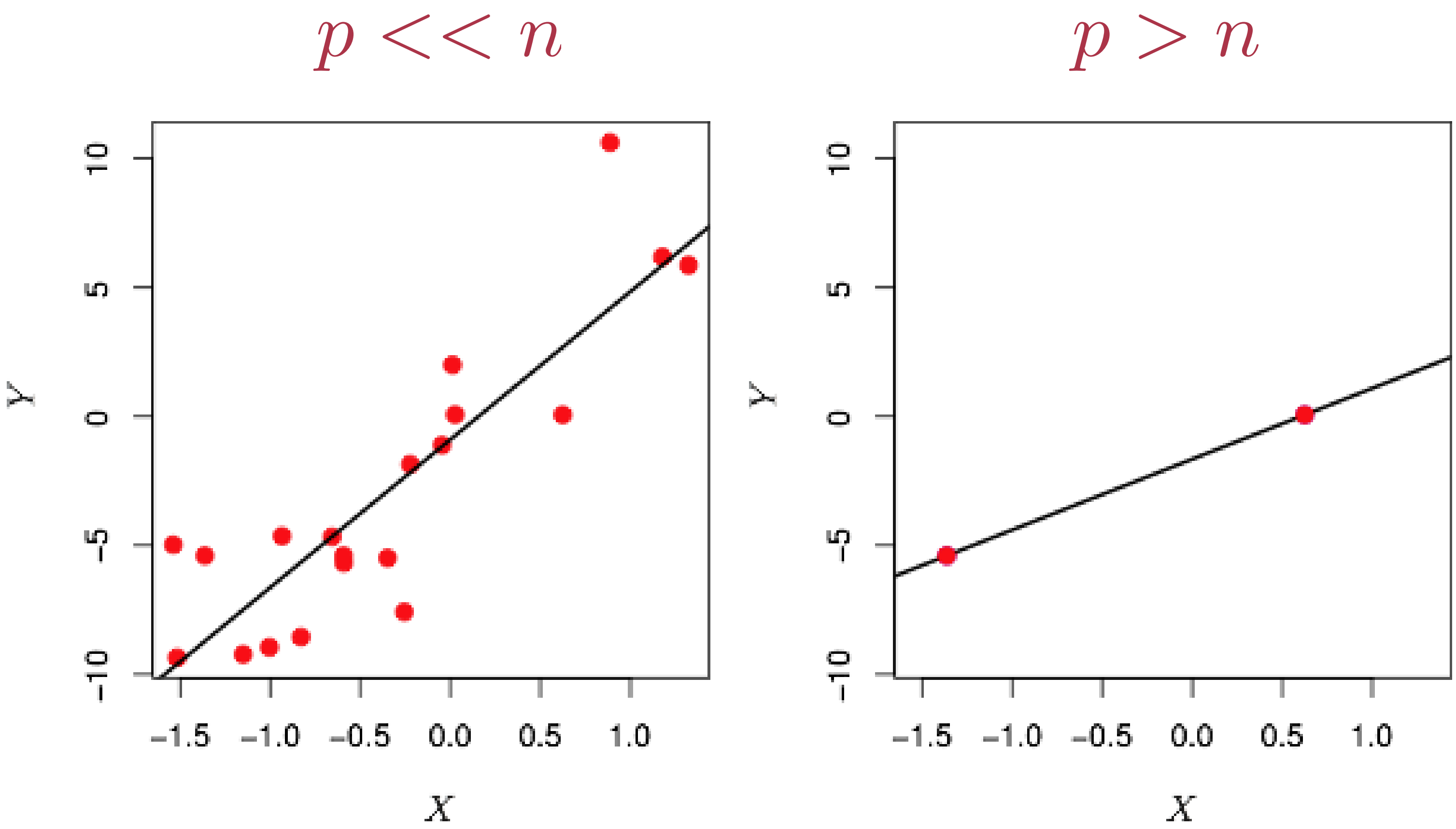
ABB (ölçekleme ile)

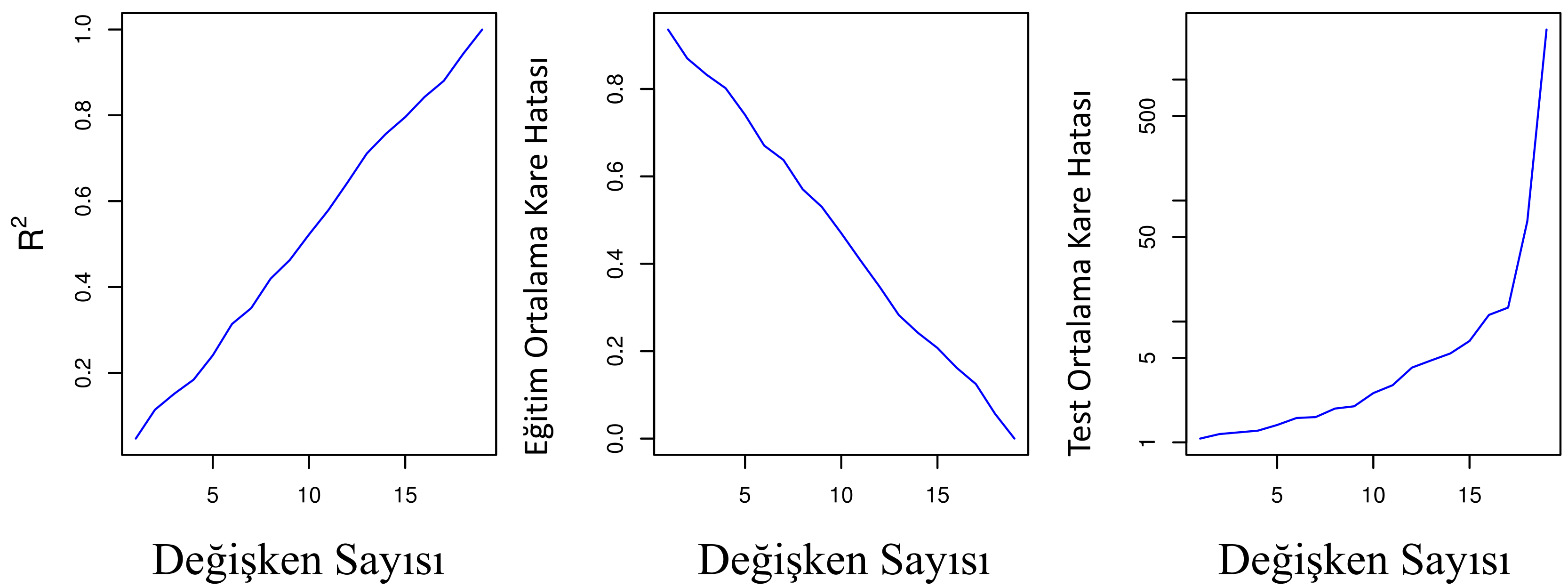
```python
regr = LinearRegression()
r2scores = []

p = len(Z[1,:])
# Fit to all components one-by-one
for i in np.arange(1, p):
    score = regr.fit(Z[:,:i], y).score(Z[:,:i], y)
    r2scores.append(score)
```

Farklı ana bileşen sayısına göre eğitim hataları



Yüksek Boyutlar





R^2
1.0
0.8
0.6
0.4
0.2
5
10
15
Değişken Sayısı
Eğitim Ortalama Kare Hatası
0.8
0.6
0.4
0.2
0.0
5
10
15
Değişken Sayısı
Test Ortalama Kare Hatası
500
50
5
1
5
10
15
Değişken Sayısı

Dikkat!

$$\boxed{p > n}$$

- Güvenilmez EKK sonuçları
- ~~R^2, p değeri, KKT~~
- Çapraz Geçerlilik Sınaması uygulaması
- Alt küme seçiminde güçlük
- Pek çok değişkenin doğrudaş (collinear) olması
- İleri adımlı seçim, çıkıntı bağlanımı, lasso, elastic net veya ABB

Özet

- Alt Küme Seçimi
  - İleri Adımlı Alt Küme Seçimi
  - Geri Adımlı Alt Küme Seçimi
- Düzenlileştirme
  - Çıkıntı Bağlanımı
  - Lasso
  - Elactic Net
- Boyut küçültme: Ana Bileşenler Bağlanımı
- Yüksek Boyutlar