Makine Öğrenmesi

Giriş

İlker Birbil ve Utku Karaca

Erasmus Üniversitesi Rotterdam

İstanbul'da Makine Öğrenmesi

27 Ocak – 2 Şubat, 2020

EUR
D4C
Bol Bilim
Veri Defteri
@sibirbil

Veri Tabanı

Veri Analizi

Veri Analitiği

Veri İşleme

Büyük Veri

Veri Bilimi

İstatistik – Analiz – Optimizasyon – Bilgisayar Bilimleri - Programlama

Makine Öğrenmesi - Yapay Öğrenme

Kitap

"An Introduction to Statistical Learning – with Applications in R," G. James, D. Witten, T. Hastie, R. Tibshirani. Yedinci baskı, Springer, New York, 2013. (www)

Altyapı

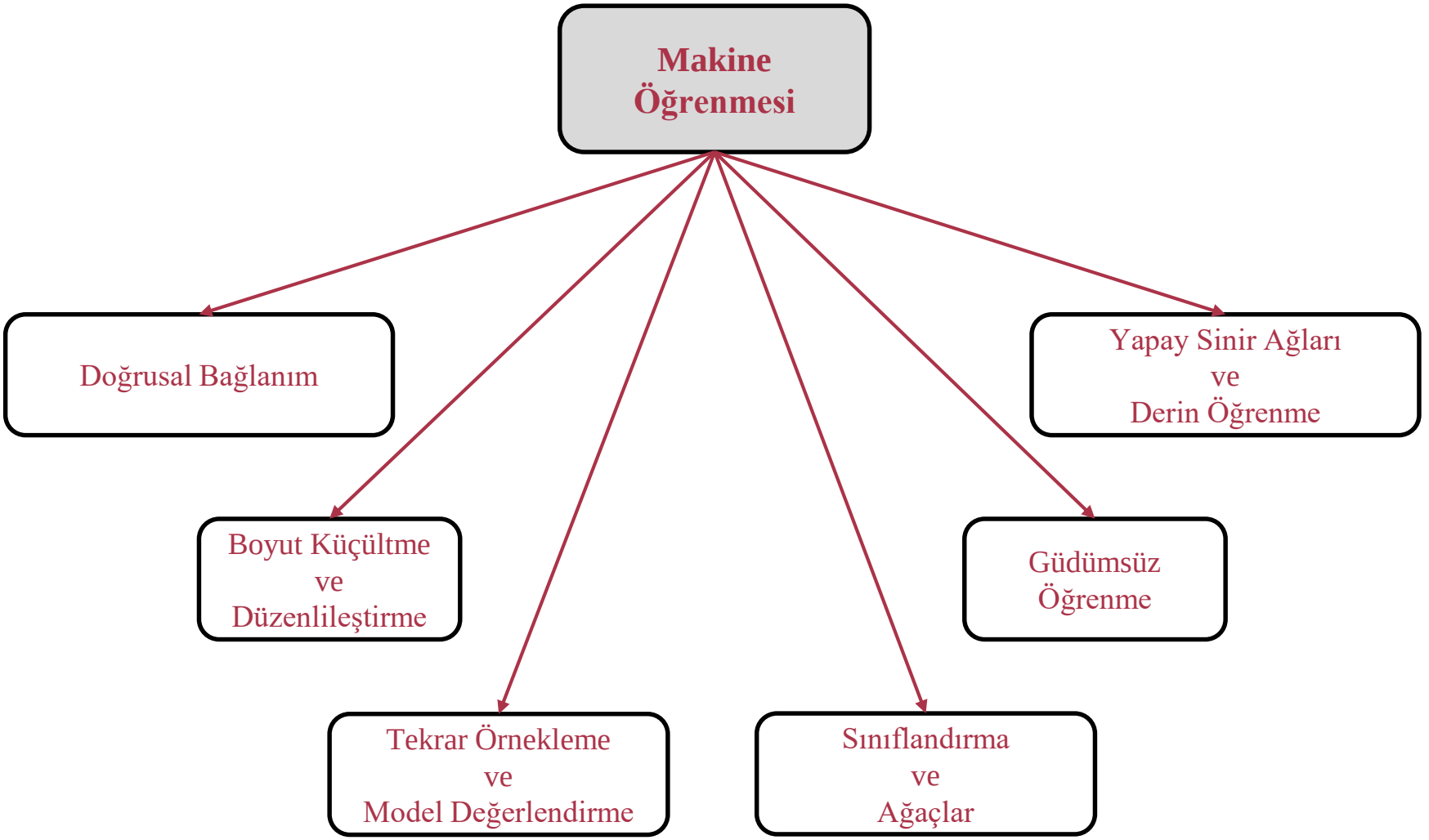
Lineer cebir: Vektörler, matrisler, ters alma, özdeğerler , vb.

Olasılık: Rassal değişkenler, beklenen değer, varyans, vb.

Analiz: Türev ve integral alma, minimum-maksimum vb.

Ders Malzemeleri

https://github.com/sibirbil/IMO2020

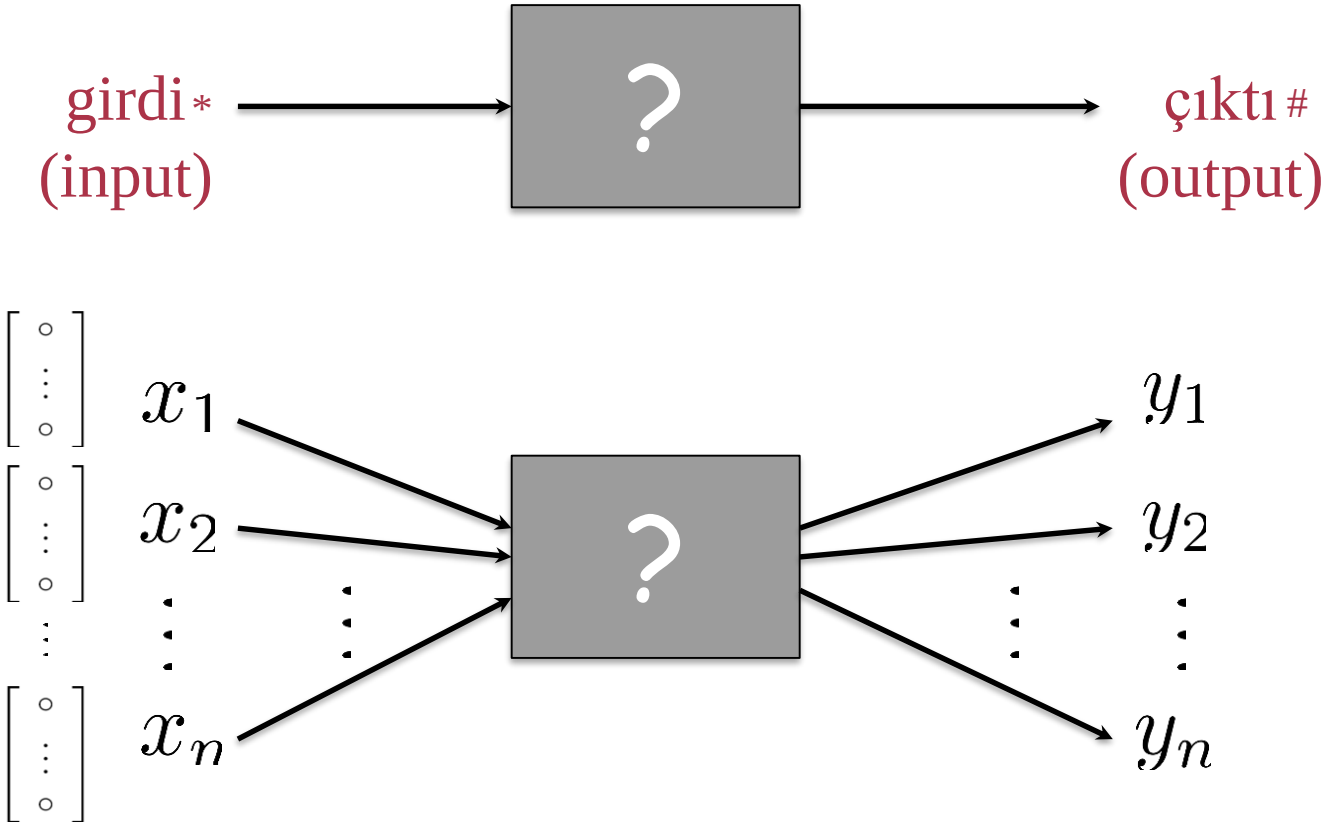

Makine
Öğrenmesi
Doğrusal Bağlanım
Yapay Sinir Ağları
ve
Derin Öğrenme
Boyut Küçültme
ve
Düzenlileştirme
Güdümsüz
Öğrenme
Tekrar Örnekleme
ve
Model Değerlendirme
Sınıflandırma
ve
Ağaçlar

Dersin sonunda

- Temel yöntemlerin kavranması
- Model kurma ve çözmenin anlaşılması
- Sonuçların incelenmesi ve yorumlanması

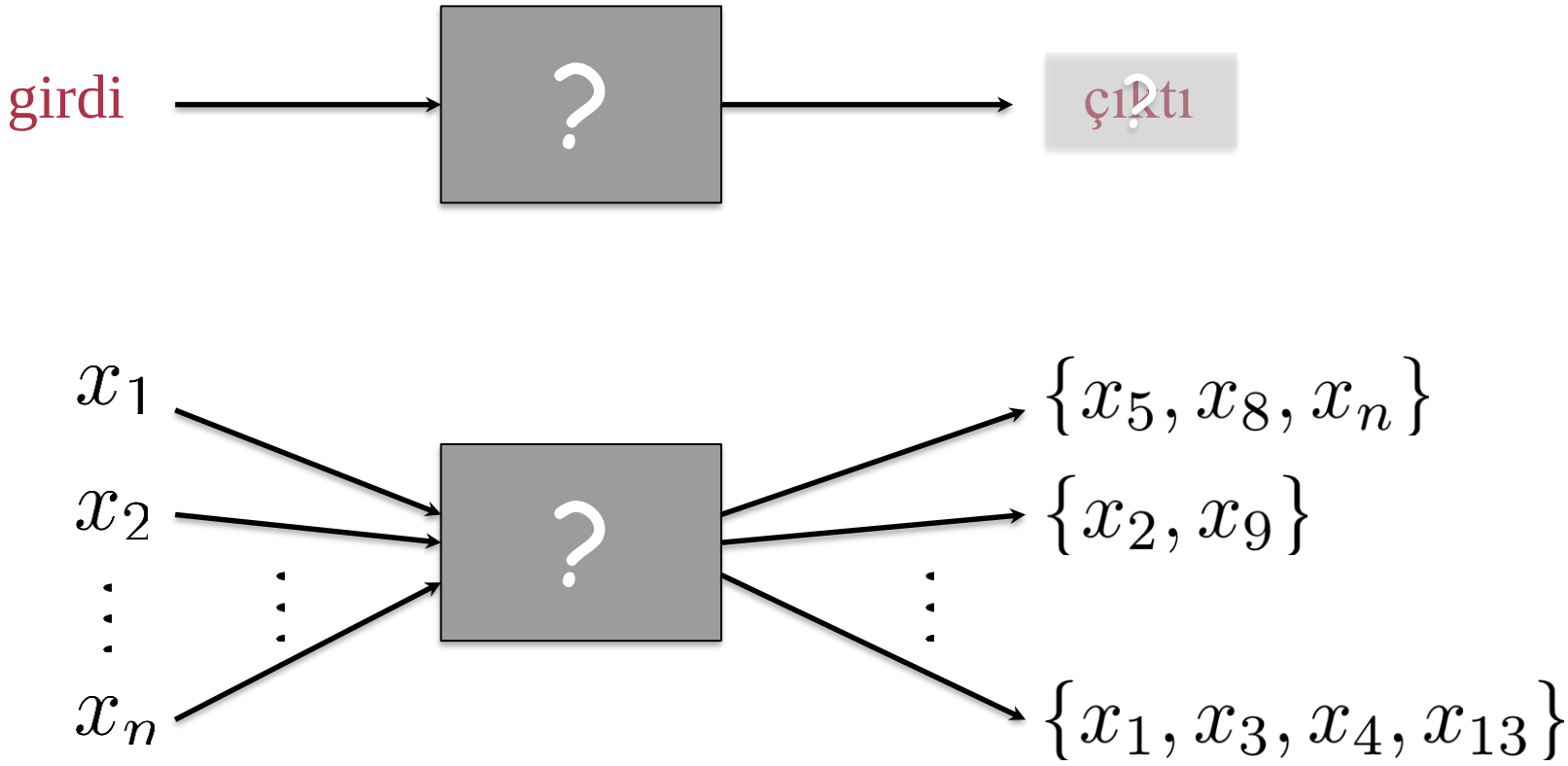
Ders sonrası

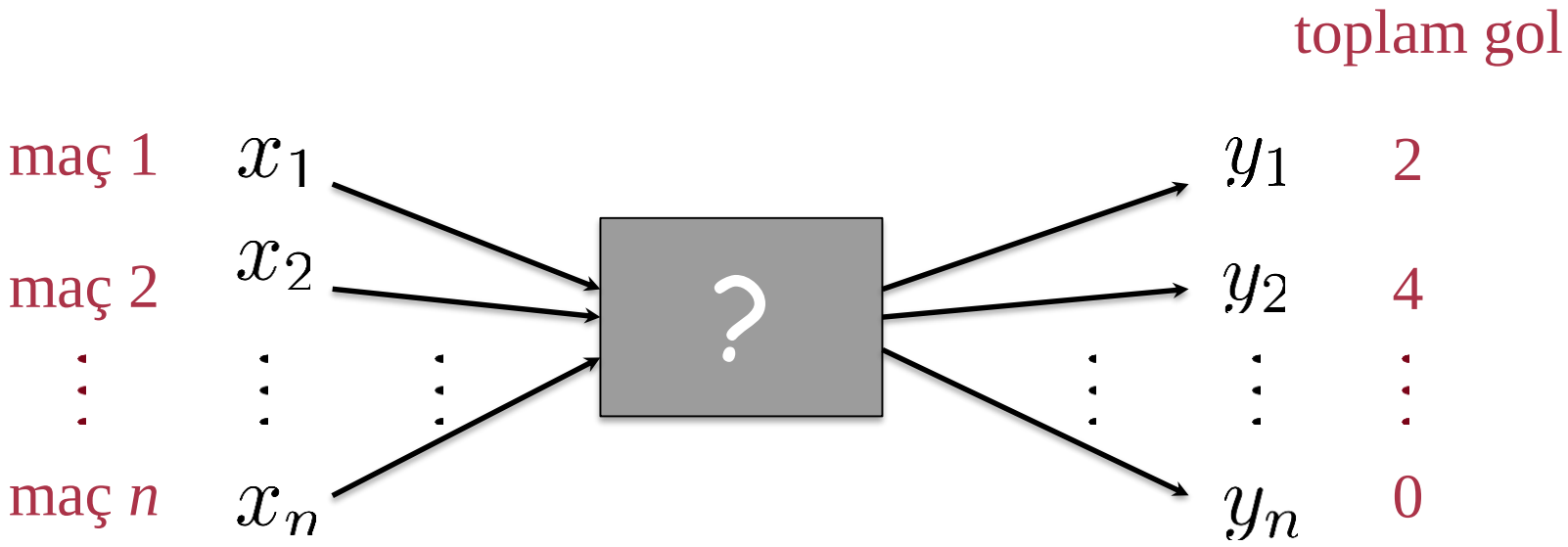
- Kuramsal çalışmalarda derinleşme
- Farklı uygulamalar ile boğuşma
- Eksik konular: pekiştirmeli öğrenme (reinforcement learning), yarı-güdümlü öğrenme (semi-supervised learning), Bayesci öğrenme, …



* bağımsız değişken (independent variable); kestirici (predictor)

bağımlı değişken (dependent variable); hedef değer (target value)


girdi
?
çıktı
?
x_1
x_2
x_n
?
$\{x_5, x_8, x_n\}$
$\{x_2, x_9\}$
$\{x_1, x_3, x_4, x_{13}\}$



$$x_i = \begin{bmatrix} 34 \\ 27 \\ 1 \\ 0 \end{bmatrix}$$

öznitelikler

attıkları gol

yedikleri gol

hava durumu: (0) yağmurlu, (1) güneşli

moral: (-2) kötü, (0) normal, (2) bomba

$$x_i = \begin{bmatrix} 34 \\ 27 \\ 1 \\ 0 \end{bmatrix}$$

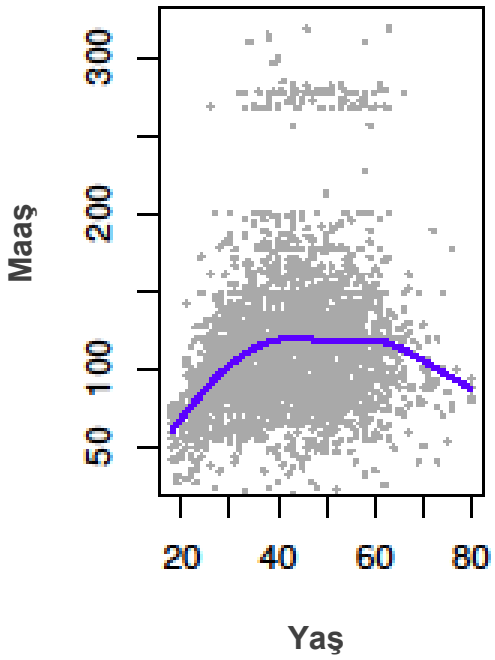
attıkları gol
yedikleri gol

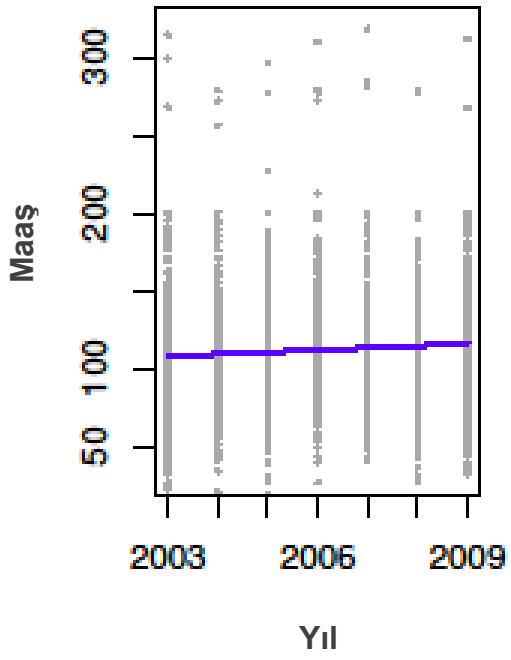
sayısal değişkenler
(numerical variables)

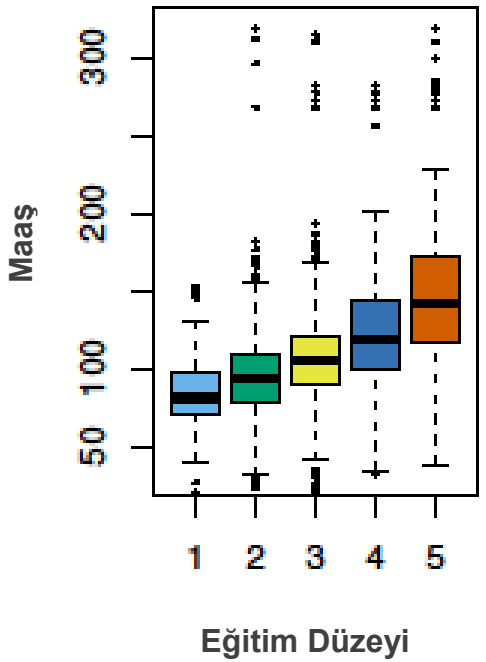
yağmurlu, güneşli
kötü, normal, bomba

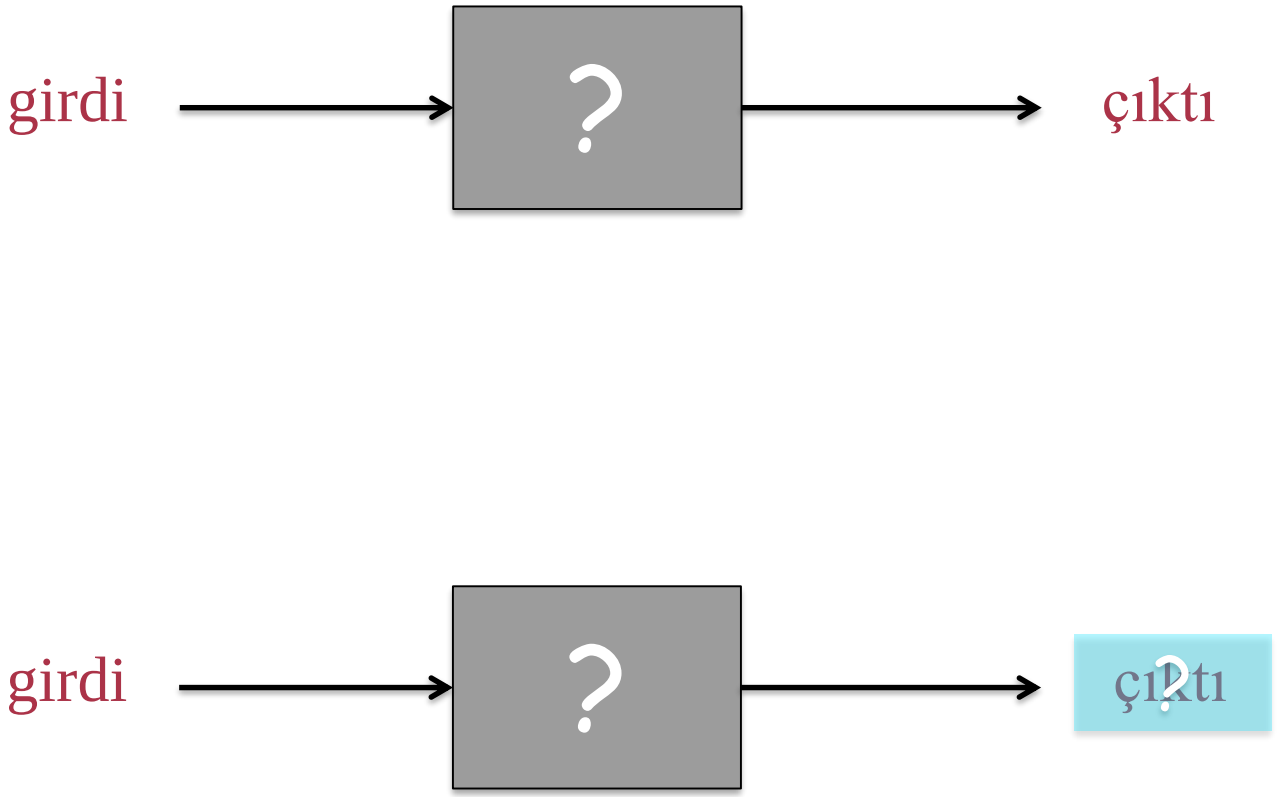
kategorik değişkenler*
(categorical variables)

* faktörler (factors)

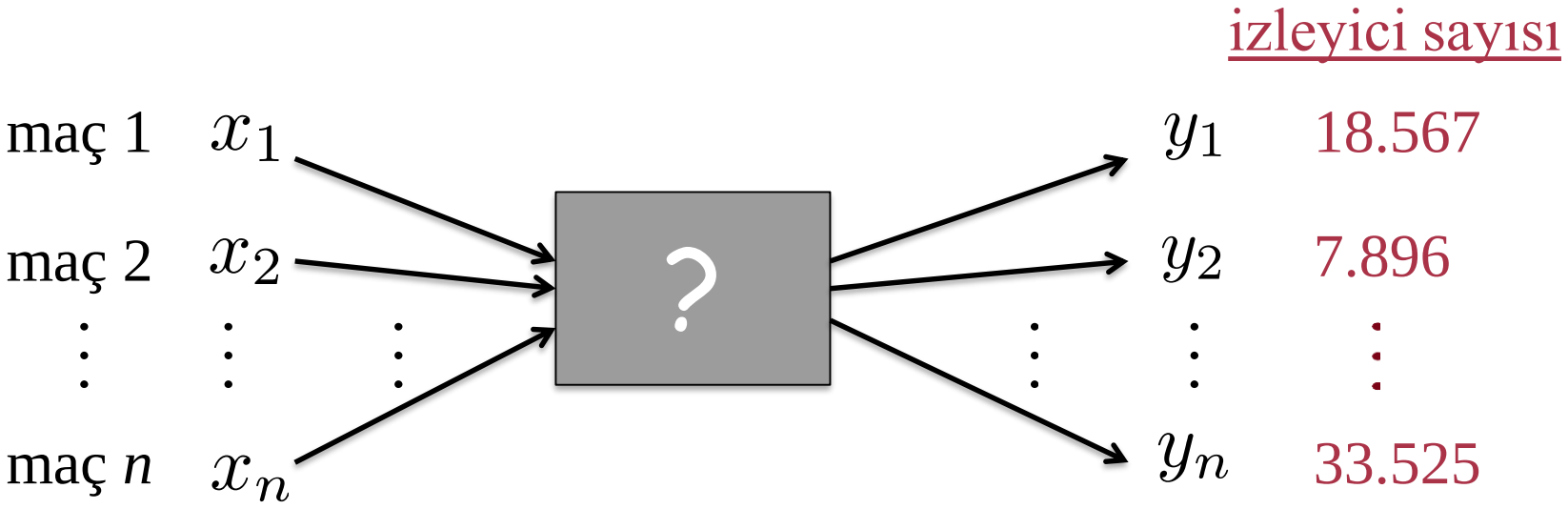

Maaş
50
100
200
300
20
40
60
80
Yaş


Maaş
50
100
200
300
2003
2006
2009
Yıl


Maaş
50
100
200
300
1
2
3
4
5
Eğitim Düzeyi


girdi
?
çıktı
girdi
?
çıktı
?

Güdümlü Öğrenme (Supervised Learning)

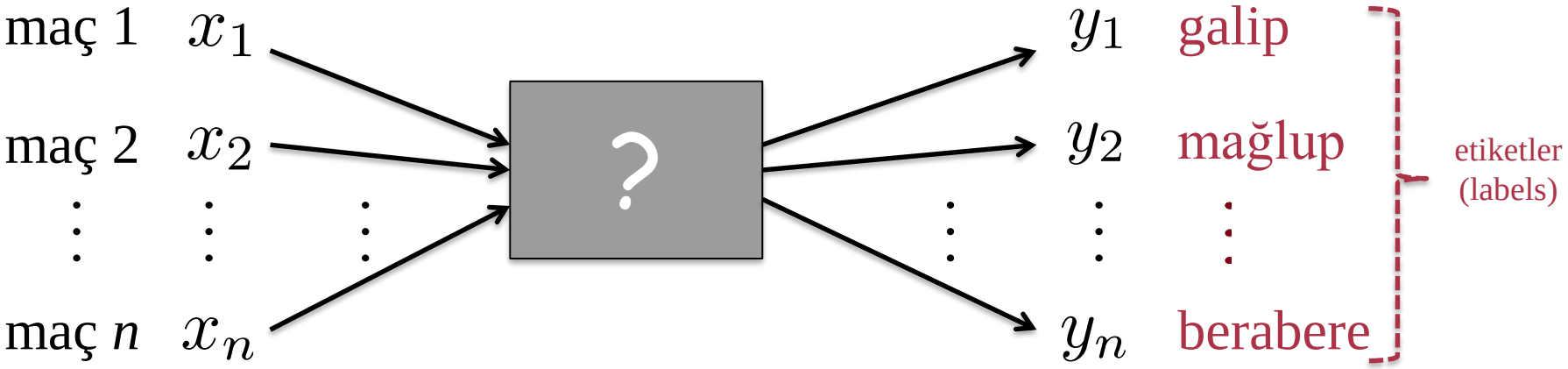


bağlanım (regression)

eğitim verisi

$$\{(x_i, y_i) : 1, \ldots, n\}$$

Güdümlü Öğrenme (Supervised Learning)

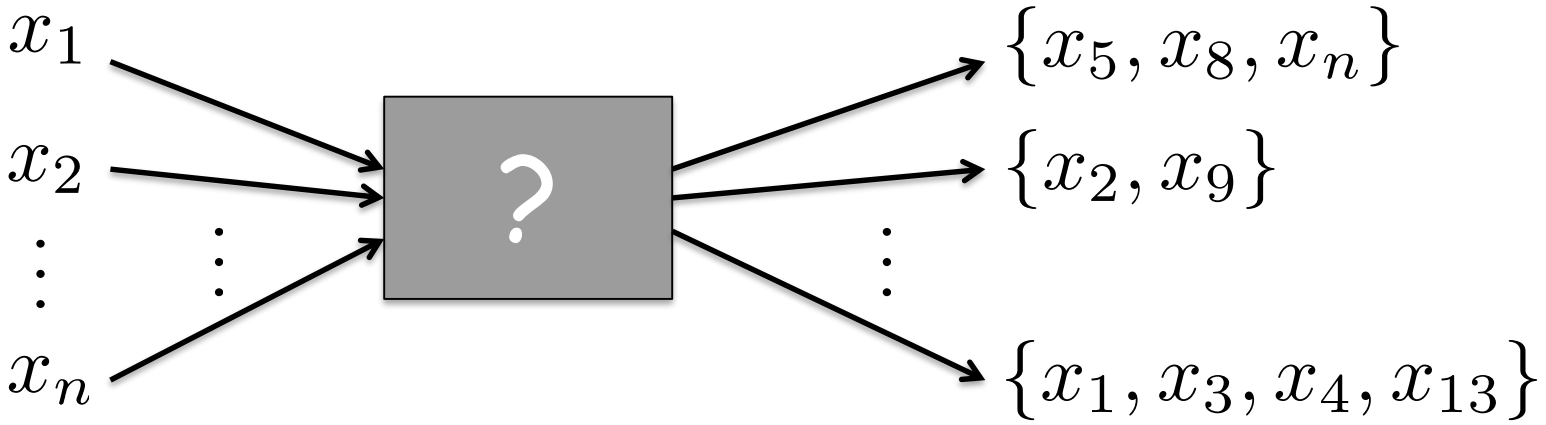


sınıflandırma (classification)

eğitim verisi

$$\{(x_i, y_i) : 1, \ldots, n\}$$

Güdümsüz Öğrenme (Unsupervised Learning)



eğitim verisi

$$\{x_i : 1, \ldots, n\}$$

Güdümlü öğrenme

$$\{(x_i, y_i) : i = 1, \ldots, n\}$$

- Resimlerden yüz tanıma
- Hasta bilgilerinden teşhis önerme
- Sayfa ziyaretlerinden harcama tahmini

... *yaklaşıklama* (*approximation*)

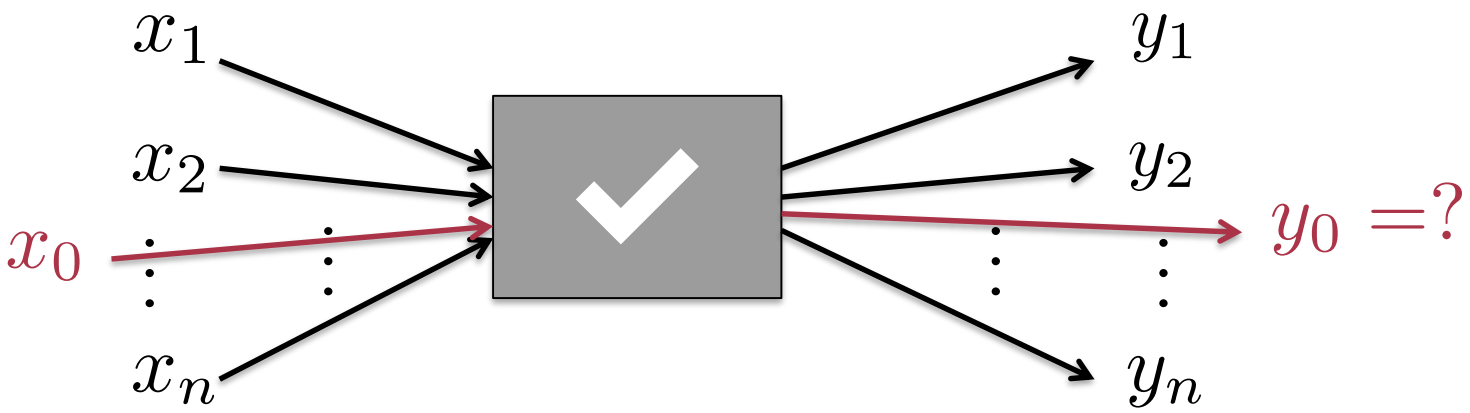
Güdümsüz öğrenme

$$\{x_i : i = 1, \ldots, n\}$$

- Alışveriş miktarlarına göre müşteri gruplama (segmentasyon)
- Hesap hareketleri ile yolsuzluk tahmini
- Öznitelik sayısını azaltma

... *betimleme* (*description*)

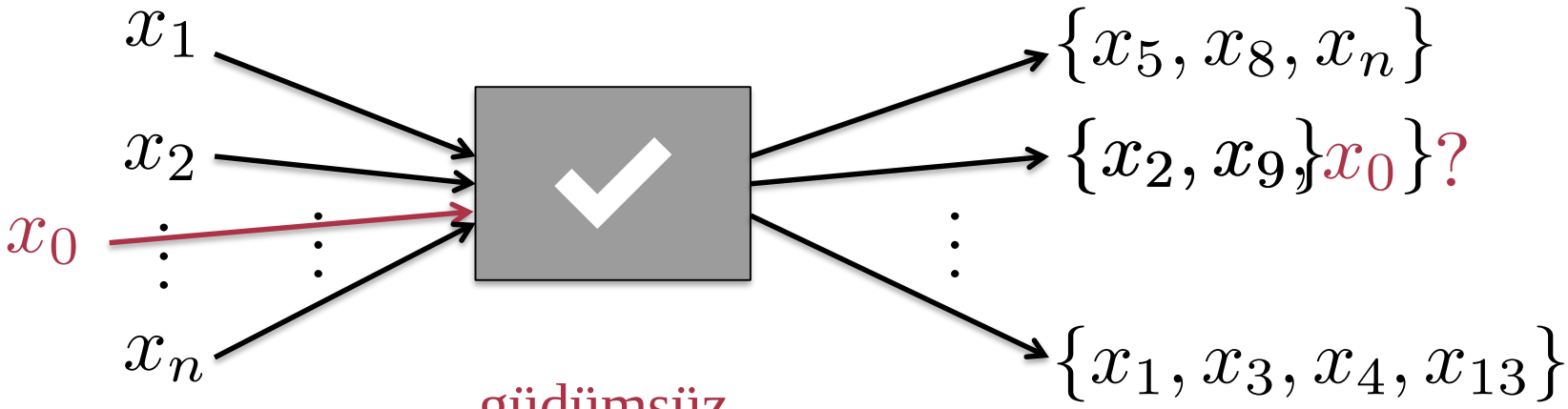
(x_0, y_0) : test verisi



güdümlü

---

x_0 : test verisi



güdümsüz

Tahmin (prediction)

$$x_0 = \begin{bmatrix} \circ \\ \vdots \\ \circ \end{bmatrix} \longrightarrow y_0 = ?$$

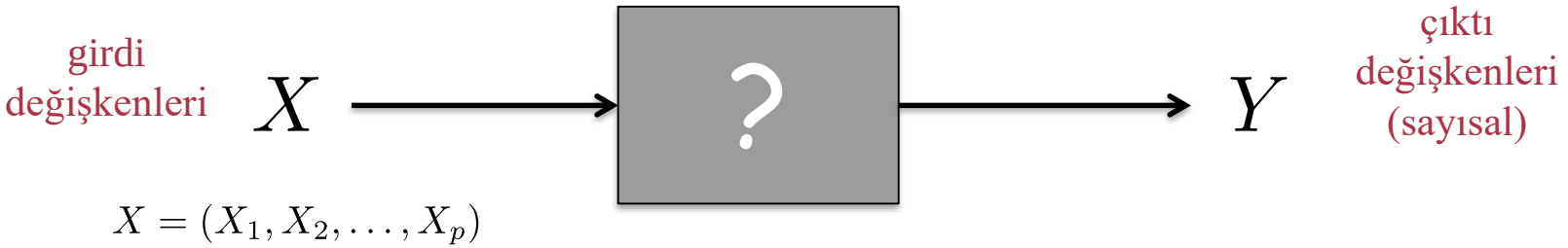
Ne?

Çıkarım (inference)

$$x_0 = \begin{bmatrix} \bullet \\ \vdots \\ \bullet \end{bmatrix} \xrightarrow{?} y_0$$

Nasıl?

BAĞLANIM



$$X = (X_1, X_2, \ldots, X_p)$$

bilinmeyen
fonksiyon

$$Y = f(X) + \epsilon$$

rassal
hata terimi
(girdiden bağımsız
ve ortalaması 0)

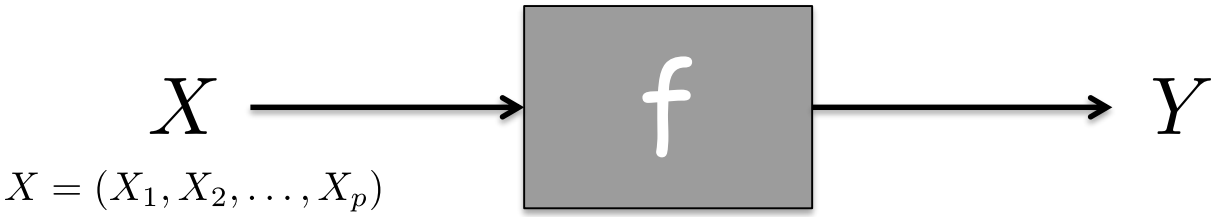
$$Y = f(X) + \epsilon \xrightarrow{\text{yaklaşık?}} \hat{Y} = \hat{f}(X)$$

$$Y - \hat{Y} = ?$$

$\hat{f}, X$ sabit

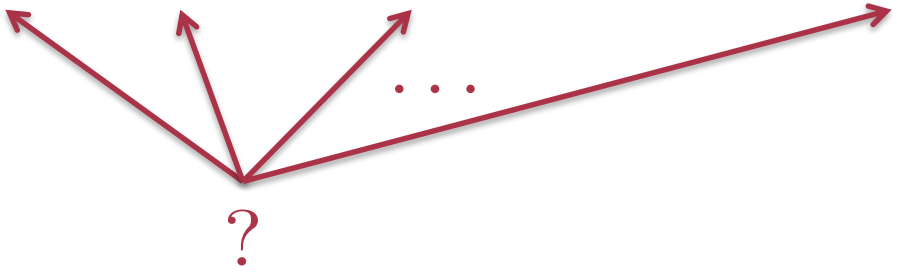
$$\mathbb{E}(Y - \hat{Y})^2 = \underbrace{\left(f(X) - \hat{f}(X)\right)^2}_{\text{bir şey yapılabilir (yöntem/model)}} + \underbrace{\text{Var}(\epsilon)}_{\text{şansımız yok}}$$

Parametrik Yöntemler (Parametric Methods)



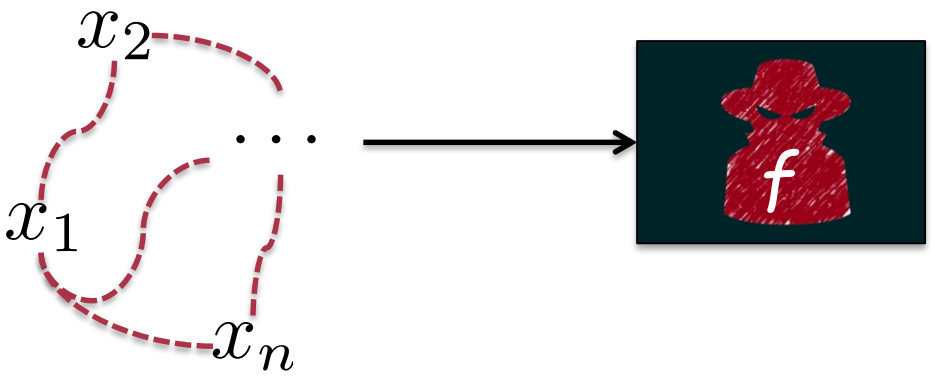
Doğrusal Model (varsayım)

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

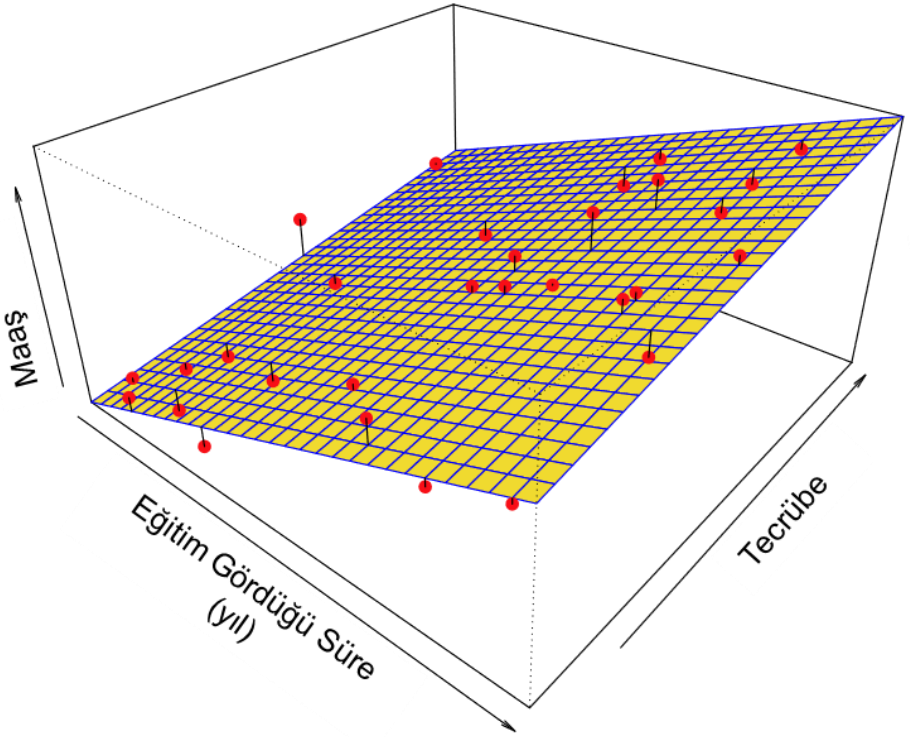


Parametrik Olmayan Yöntemler (Non-parametric Methods)

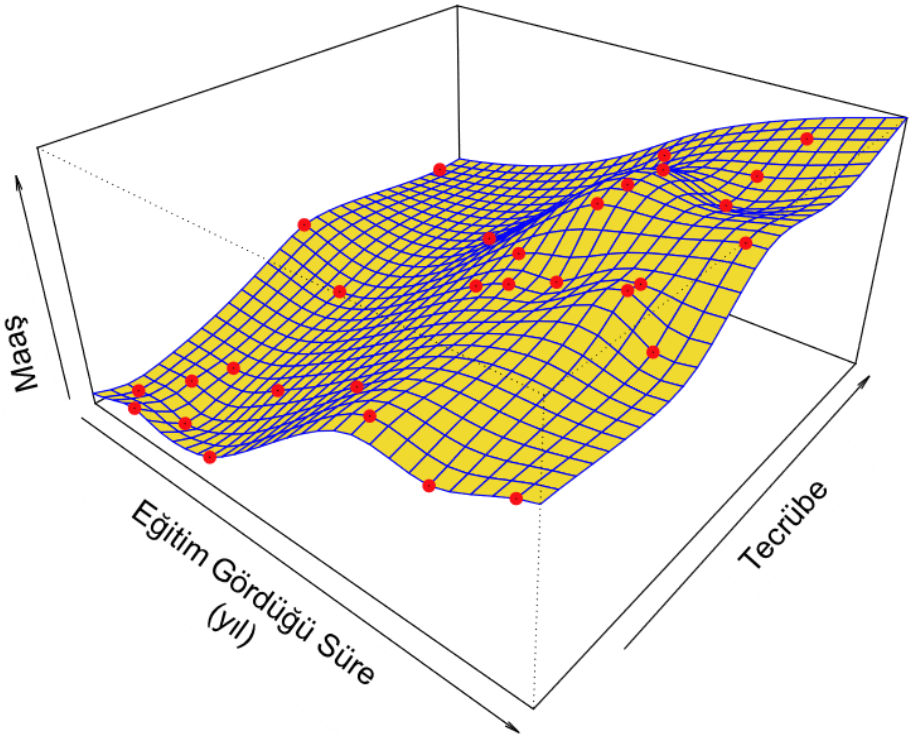
$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$$



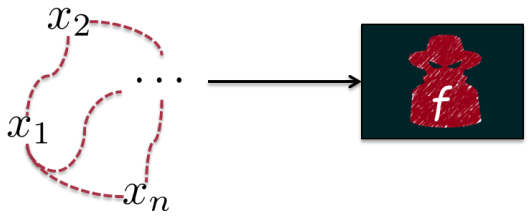
Parametrik Yöntem
(doğrusal model)



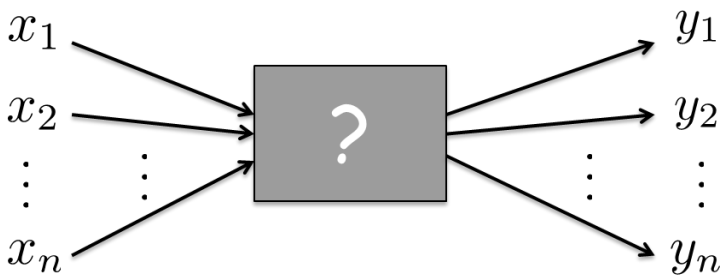
Parametrik Olmayan Yöntem
(spline)



$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$



$$Y = f(X) + \epsilon \xrightarrow{\text{yaklaşık?}} \hat{Y} = \hat{f}(X)$$



eğitim verisi

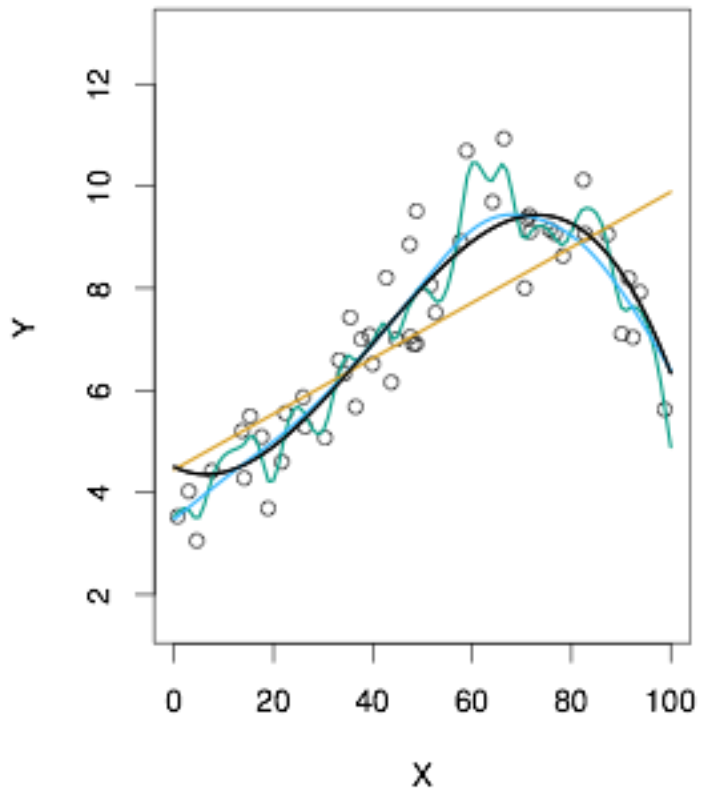
$$\{(x_i, y_i) : 1, \ldots, n\}$$

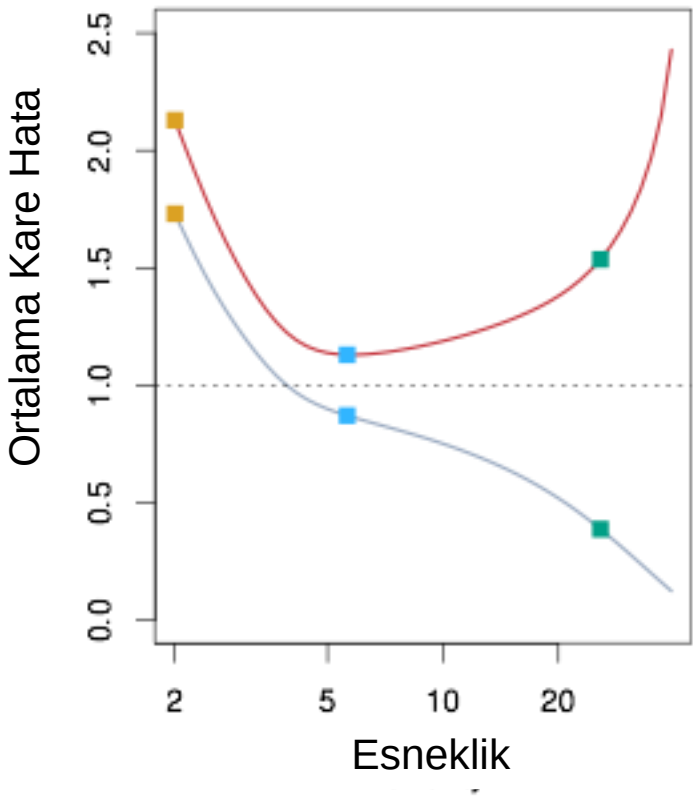
Eğitim Verisi Ortalama Kare Hata (Mean Square Error)

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

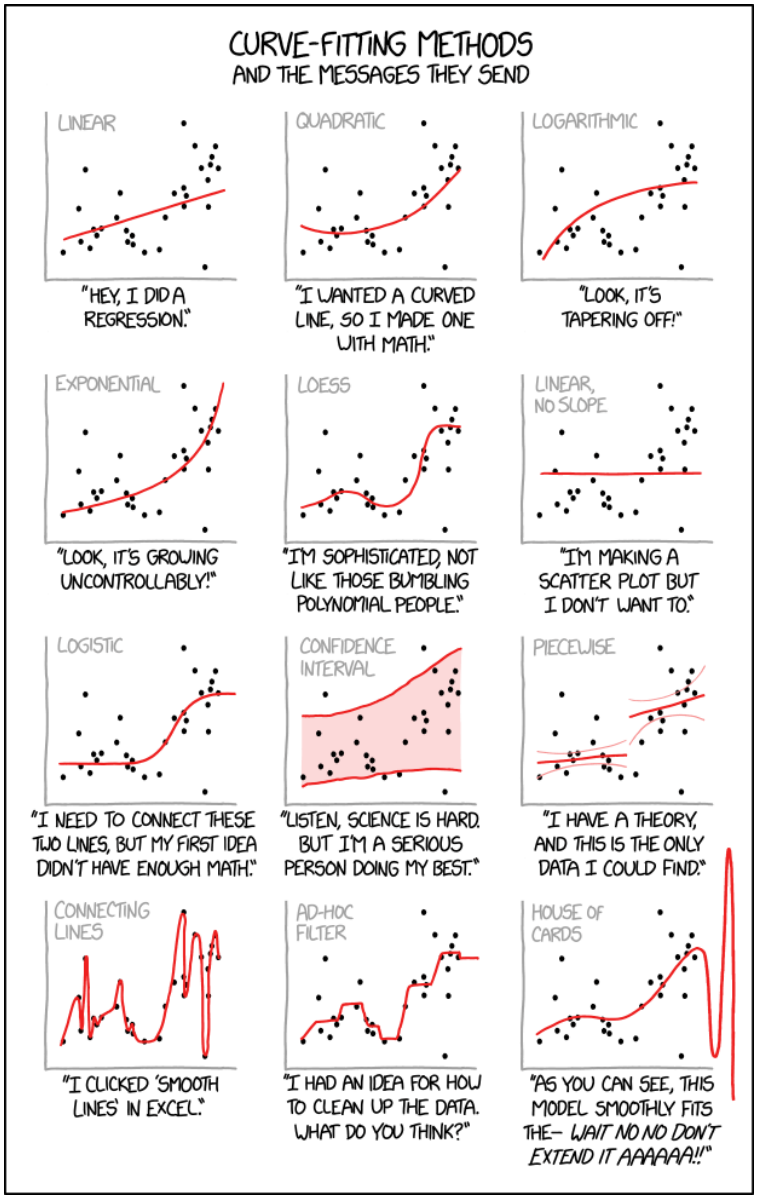
(x_0, y_0) : test verisi ???

$$\mathbb{E}(Y - \hat{Y})^2 = \left(f(X) - \hat{f}(X)\right)^2 + \mathrm{Var}(\epsilon)$$





aşırı öğrenme (overfitting)


CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND
LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."
LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE– WAIT NO NO DON'T EXTEND IT AAAAAA!!"

(kaynak)

Test verisinin beklenen ortalama kare hatası:

$$\mathbb{E}(y_0 - \hat{f}(x_0))^2 = \mathbb{E}\left(\left(f(x_0) + \epsilon - \hat{f}(x_0)\right)^2\right)$$

$$= \mathbb{E}\left((f(x_0))^2 + 2\epsilon f(x_0) + \epsilon^2 - 2(f(x_0) + \epsilon)\hat{f}(x_0) + (\hat{f}(x_0))^2\right)$$

$$= \mathbb{E}\big((f(x_0))^2\big) + 2\mathbb{E}\big(\epsilon f(x_0)\big) - 2\mathbb{E}\big(f(x_0)\hat{f}(x_0)\big) + \mathbb{E}\big((\hat{f}(x_0))^2\big) + \mathrm{Var}(\epsilon)$$

$$= (f(x_0))^2 - 2f(x_0)\mathbb{E}\big(\hat{f}(x_0)\big) + \mathbb{E}\big((\hat{f}(x_0))^2\big) + \mathrm{Var}(\epsilon) + \mathbb{E}\big(\hat{f}(x_0)\big)^2 - \mathbb{E}\big(\hat{f}(x_0)\big)^2$$

$$= \underbrace{\mathbb{E}\big((\hat{f}(x_0))^2\big) - \mathbb{E}\big(\hat{f}(x_0)\big)^2}_{\text{Varyans (variance)}} + \underbrace{\big(\mathbb{E}(\hat{f}(x_0)) - f(x_0)\big)^2}_{\text{Yanlılık (bias)}} + \mathrm{Var}(\epsilon)$$

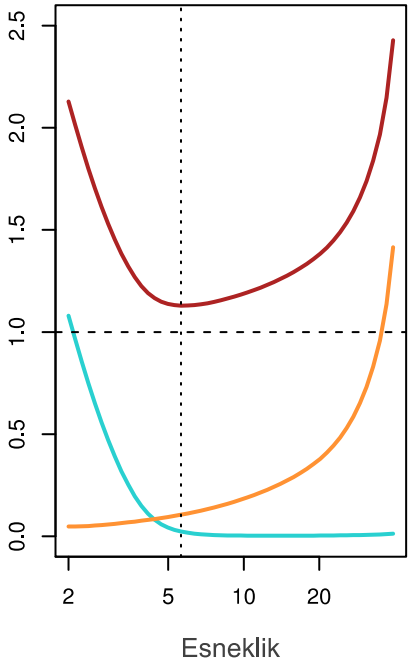
Test verisinin beklenen ortalama kare hatası:

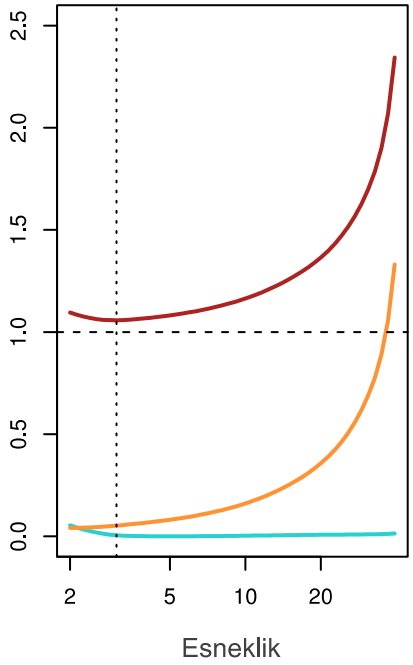
$$\begin{aligned}\mathbb{E}(y_0 - \hat{f}(x_0))^2 &= \dots \\ &= \mathbb{E}\big((\hat{f}(x_0))^2\big) - \mathbb{E}\big(\hat{f}(x_0)\big)^2 + \big(\mathbb{E}(\hat{f}(x_0)) - f(x_0)\big)^2 + \mathrm{Var}(\epsilon) \\ &= \mathrm{Var}(\hat{f}(x_0)) + \big(\mathrm{Bias}(\hat{f}(x_0))\big)^2 + \mathrm{Var}(\epsilon)\end{aligned}$$

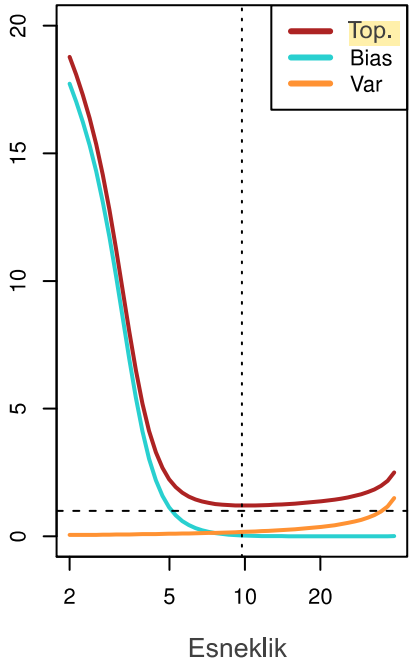
Kabaca:

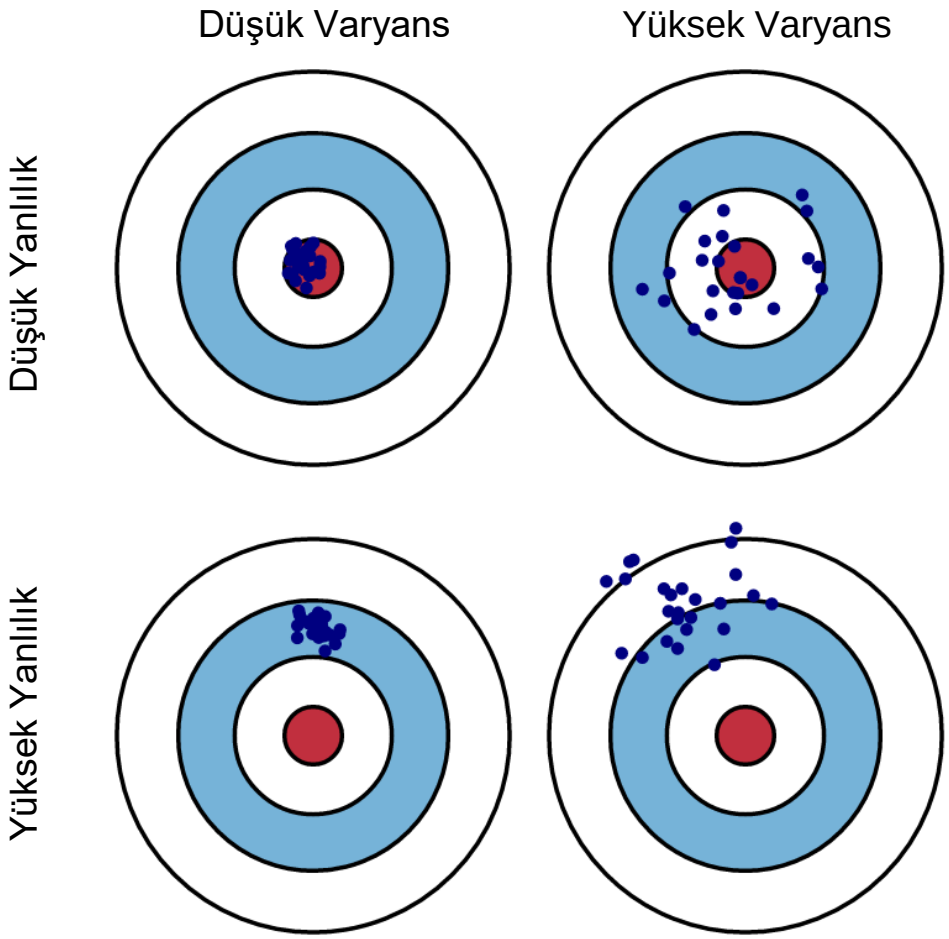
Model Karmaşıklığı ↑ Varyans ↑ Yanlılık ↓

$$\mathbb{E}(y_0 - \hat{f}(x_0))^2 = \mathrm{Var}(\hat{f}(x_0)) + \left(\mathrm{Bias}(\hat{f}(x_0))\right)^2 + \mathrm{Var}(\epsilon)$$







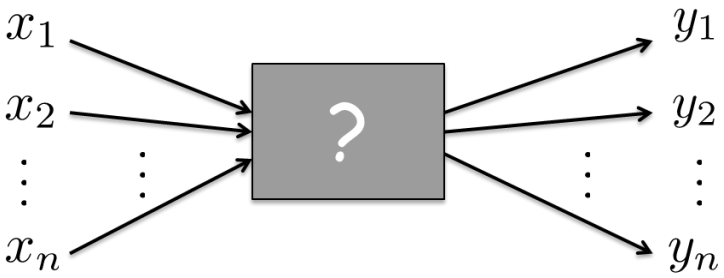

Düşük Varyans
Yüksek Varyans
Düşük Yanlılık
Yüksek Yanlılık

SINIFLANDIRMA

girdi değişkenleri

çıktı değişkenleri (kategorik)

$$X = (X_1, X_2, \ldots, X_p)$$



eğitim verisi

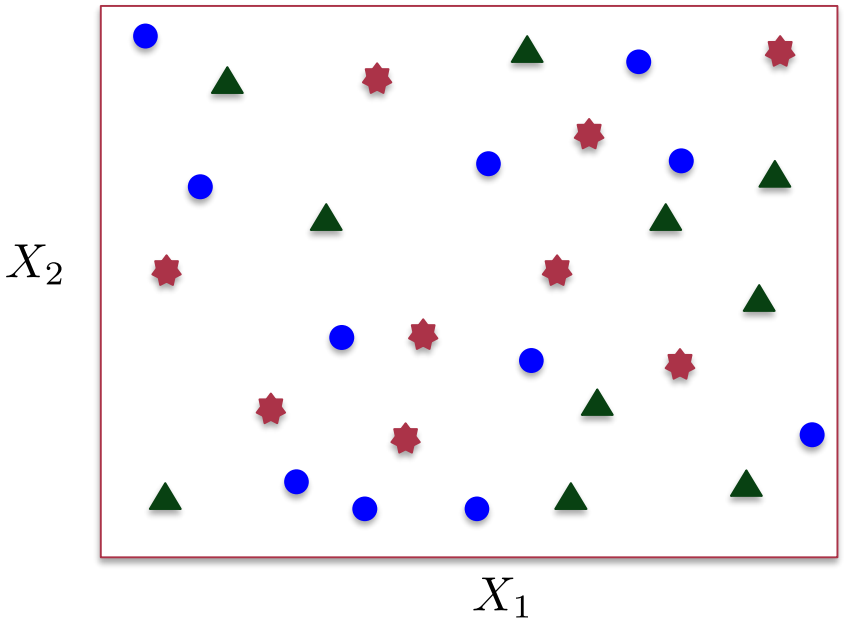
$$\{(x_i, y_i) : 1, \ldots, n\}$$

Eğitim Verisi Hata Oranı

$$\frac{1}{n} \sum_{i=1}^{n} I(y_i \neq \hat{y}_i)$$

(x_0, y_0) : test verisi ???

K - En Yakın Komşu (KEYK)



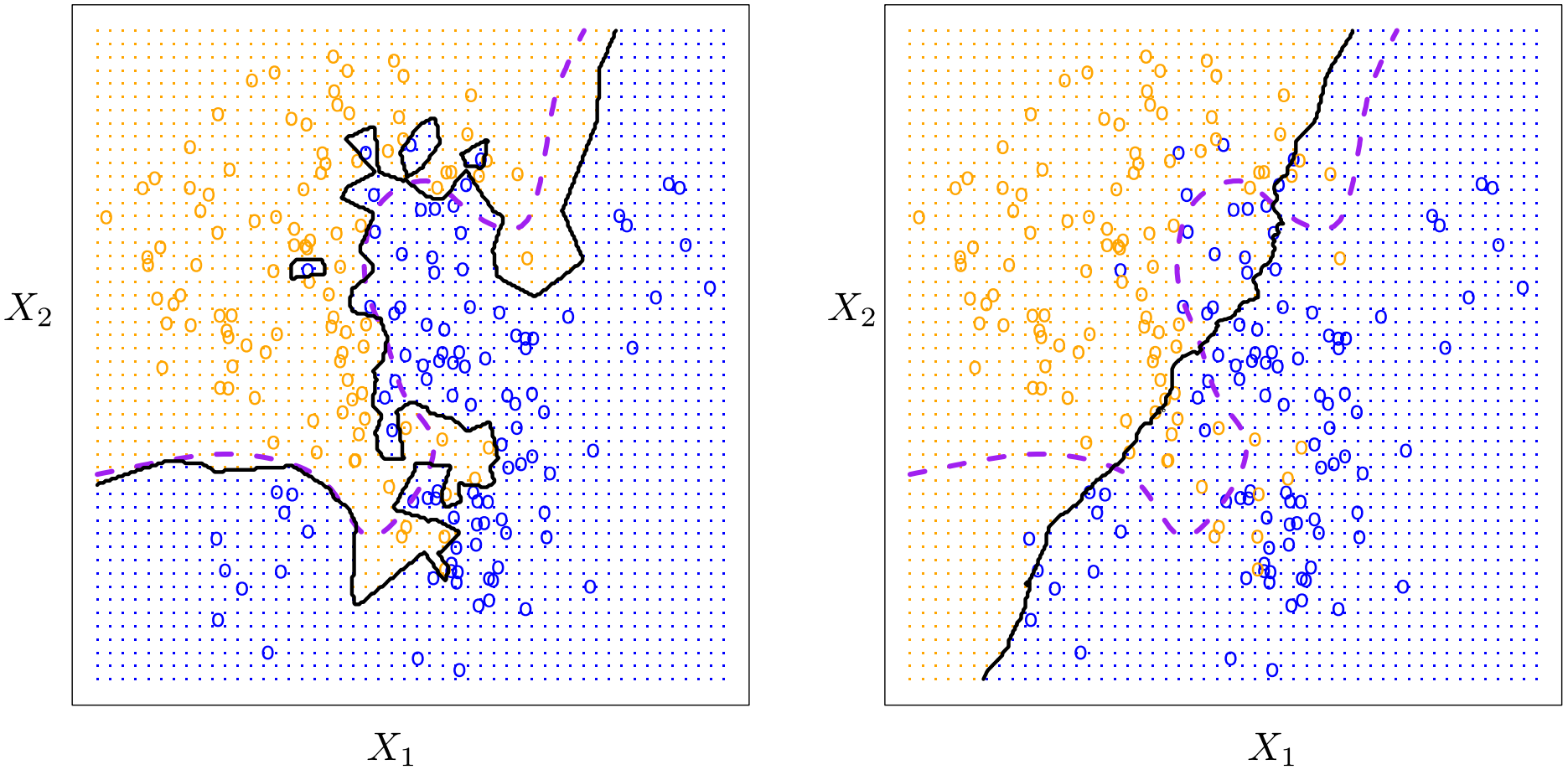
Yakınlık ???

K ???

oylama

$$\mathbb{P}(Y = j | X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j)$$

$$|\mathcal{N}_0| = K$$


$K = 1$
$K = 100$
X_2
X_1
X_2
X_1

Bağlanım

X, Y rassal: $f(X) = ?$

$$\mathbb{E}((Y - f(X))^2) = \mathbb{E}(\mathbb{E}((Y - f(X))^2 \mid X))$$

$$f(x) = \arg\min_u \underbrace{\mathbb{E}((Y - u)^2 \mid X = x)}_{h(u)}$$

$$\frac{\partial h(u)}{\partial u} = 2u - 2\mathbb{E}(Y|X = x) = 0 \implies f(x) = \mathbb{E}(Y|X = x)$$

$$\boxed{f(x_0) = \mathbb{E}(Y|X = x_0)}$$

$$\hat{f}(x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} y_i$$

Bağlanıma *KEYK* yaklaşıklığı

★

Sınıflandırma

X, Y rassal: $\hat{Y}(X) = ?$

$$\mathbb{E}(I(Y \neq \hat{Y}(X))) = \mathbb{E}(\mathbb{E}(I(Y \neq \hat{Y}(X)) \mid X))$$

$$\begin{aligned}
\hat{Y}(x) &= \arg\min_j \ \mathbb{E}(I(Y \neq j) \mid X = x) \\
&= \arg\min_j \ \sum_k I(k \neq j)\mathbb{P}(Y = k|X = x) \\
&= \arg\min_j \ (1 - \mathbb{P}(Y = j|X = x)) \\
&= \arg\max_j \ \mathbb{P}(Y = j|X = x)
\end{aligned}$$

Bayes Sınıflandırıcısı (classifier)

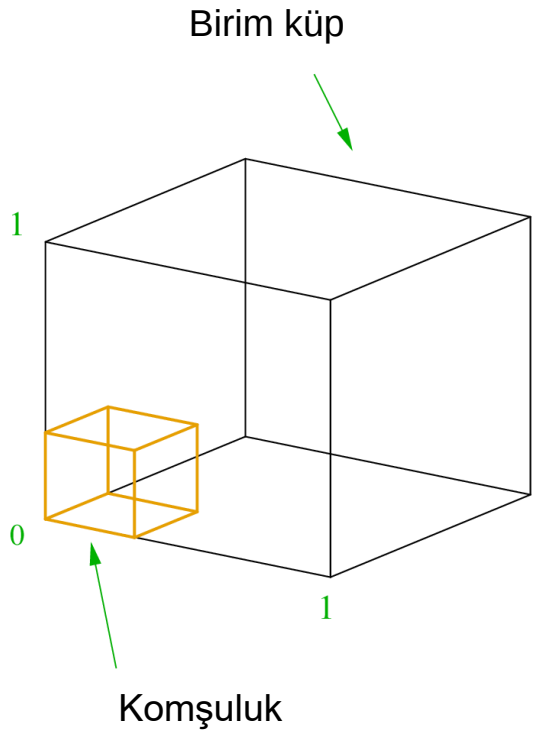
(kaynak)

$$\hat{y}_0 = \max_j \ \mathbb{P}(Y = j | X = x_0)$$

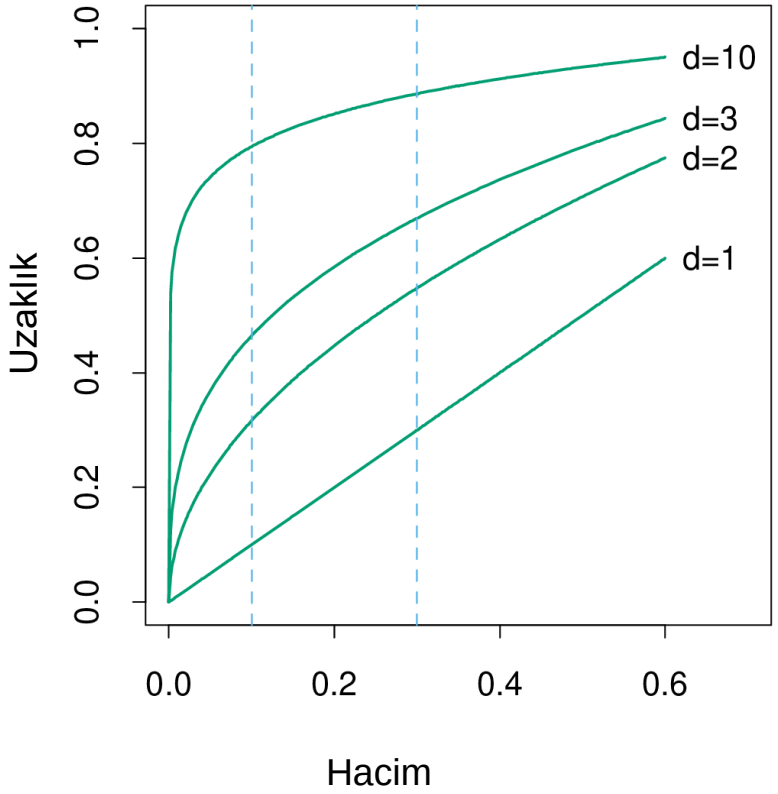
Bayes Hata Oranı

$$\mathbb{E}\big(1 - \max_j \ \mathbb{P}(Y = j|X)\big) = 1 - \mathbb{E}\big(\max_j \ \mathbb{P}(Y = j|X)\big)$$

Boyutluluk Belası (Curse of Dimensionality)



*



* *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, T. Hastie, R. Tibshirani ve J. Friedman, İkinci Basım, Springer, 2009, sf 23.

Özet

- Bağlanım ve sınıflandırma
- Güdümlü-güdümsüz Öğrenim
- Fonsiyon yaklaşıklama, parametrik olan ve olmayan yöntemler
- Eğitim ve test hataları
- Varyans – Yanlılık
- Bayes Sınıflandırıcısı ve K – En Yakın Komşu yöntemi
- Boyutluluk belası