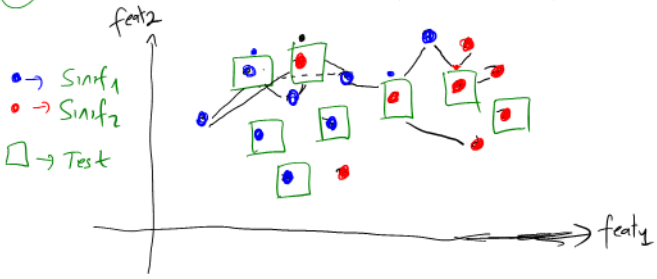
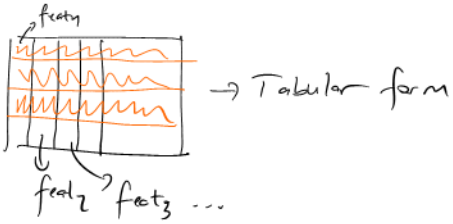
k-NN (Distance based) Classification Supervised (Sınıf / Etiket var)



k=3



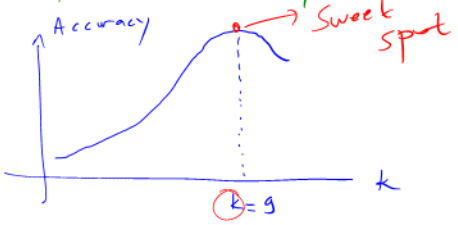
→ Tabular form

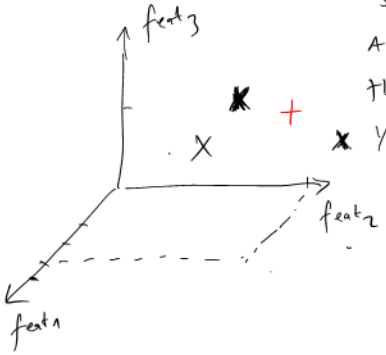
Ayrık (Class)

	Tansiyon	Kolestr.	Şeker	Kriz
Süha	130	220	110	Yok
Ahmet	140	230	95	Var
Hakan	120	250	98	Yok
Yasin	115	165	72	Yok
	↓ feat1	↓ feat2	↓ feat3	

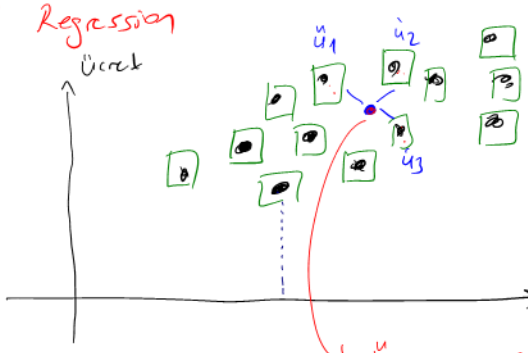
Özniteliklik

k: En yakın komşu sayısı





k-NN Regression



D: Train

k=3

Tecrübe	Ücret
25	---
12	---
42	---
30	---
3	---

$$\ddot{u}_{predicted} = \frac{\ddot{u}_1 + \ddot{u}_2 + \dots + \ddot{u}_k}{k}$$

→ Default

Euclid

(x_1, y_1, z_1)

(x_2, y_2, z_2)

$$d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$$

- Euclid
- Minkowski
- Cosine
- Jaccard
- Mahalanobis

Linear Regression → Cont. numerical values

$$y \approx b_0 + \underbrace{b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_4 \cdot x_4}_{\text{linear model}}$$

- b_1, b_2, b_3, b_4: coefficients
- b_0: intercept (bias)
- linear comb.

	x_1 feat$_1$	x_2 feat$_2$	x_3 feat$_3$	x_4 feat$_4$	y
Sample$_1$	3.7	4.2	1.3	0.8	0.75
Sample$_2$	6.2	6.1	4.4	1.2	0.56
⋮	⋮	⋮	⋮	⋮	⋮

x_1, x_2, x_3, x_4: independent variables

y: dependent variable

$y = f(x)$

Amaç: $b_0, b_1, \dots, b_4$ katsayılarını optimal bir şekilde bulmak!

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(n)} & x_2^{(n)} & x_3^{(n)} & x_4^{(n)} \end{bmatrix} \quad \begin{bmatrix} y \\ y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

$$y^{(1)} \approx b_0 + b_1 x_1^{(1)} + \dots + b_4 x_4^{(1)}$$
$$y^{(2)} \approx b_0 + b_1 x_1^{(2)} + \dots + b_4 x_4^{(2)}$$
$$\vdots$$
$$y^{(n)} \approx b_0 + b_1 x_1^{(n)} + \dots + b_4 x_4^{(n)}$$

b_0, b_1, b_2, b_3, b_4 = ?

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ & & \vdots & & \\ 1 & x_1^{(n)} & x_2^{(n)} & x_3^{(n)} & x_4^{(n)} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \approx \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

5 bilinmeyen

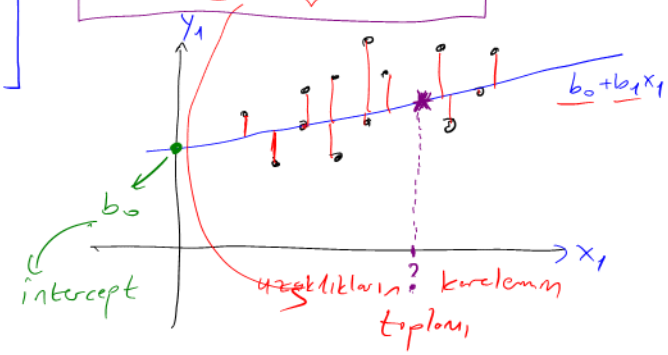
X → Dikdörtgen $n \times 5$

$\vec{b}$

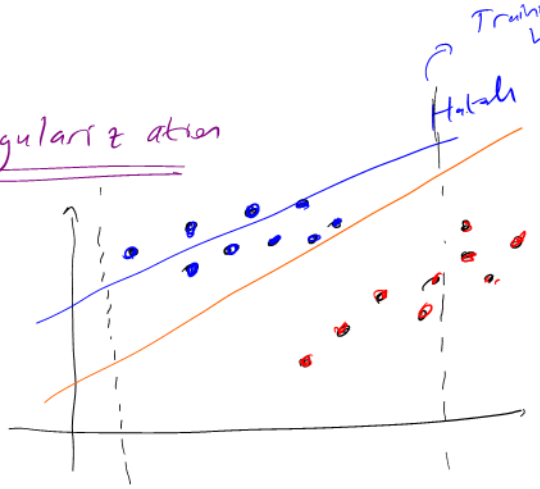
y

$$X\vec{b} \approx \vec{y} \Rightarrow \vec{b} = ?$$

$$\min \| X\vec{b} - \vec{y} \|_2^2$$

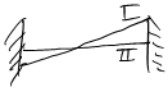


Regularization



Overfitting

• → Train
• → Test



RIDGE

$$\min_{\vec{b}} \|X\vec{b} - \vec{y}\|_2^2 + \alpha \|\vec{b}\|_2^2$$

Penalty term

$$b_0^2 + b_1^2 + \dots + b_4^2$$

LASSO

$$\min_{\vec{b}} \|X\vec{b} - \vec{y}\|_2^2 + \beta \|\vec{b}\|_1$$

Penalty norm

$$|b_0| + |b_1| + \dots + |b_4|$$

$$\min_{\vec{b}} \|X\vec{b} - \vec{y}\|_2^2 + \alpha \|\vec{b}\|_2 + (1-\alpha) \|\vec{b}\|_1$$

Elastic-Net problem

$\alpha = [0.1, 0.01, \boxed{0.001}, 0.0001] \Rightarrow$ grid-search (R^2 for each)