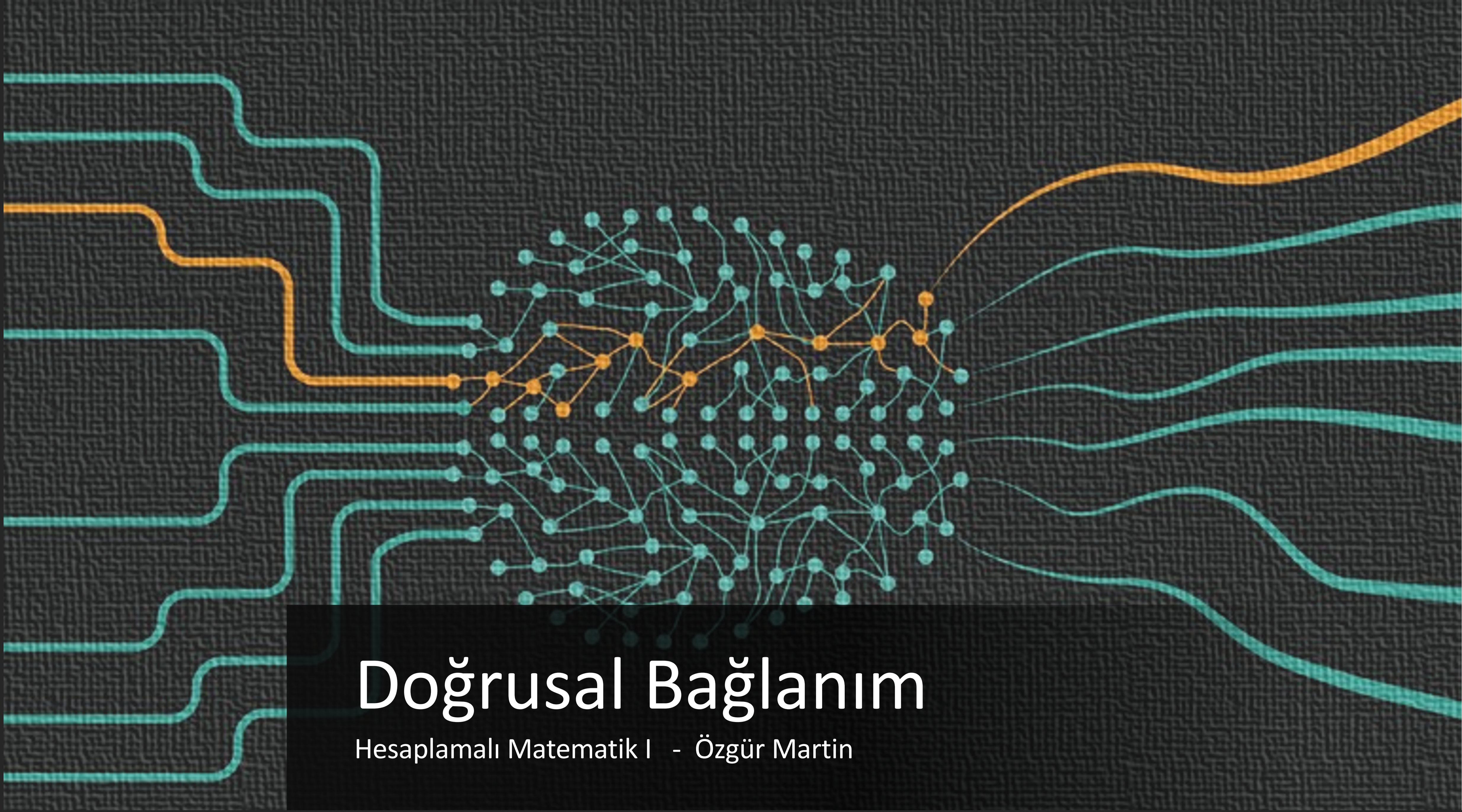


Doğrusal Bağlanım

Hesaplamalı Matematik I - Özgür Martin

Hatırlatma: Öğrenme Problemi

Hedef Fonksiyon

$$f : X \to Y \qquad y = f(x) + \epsilon$$
$$x \mapsto f(x)$$

Eğitim Verisi

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$$

Öğrenme Algoritması

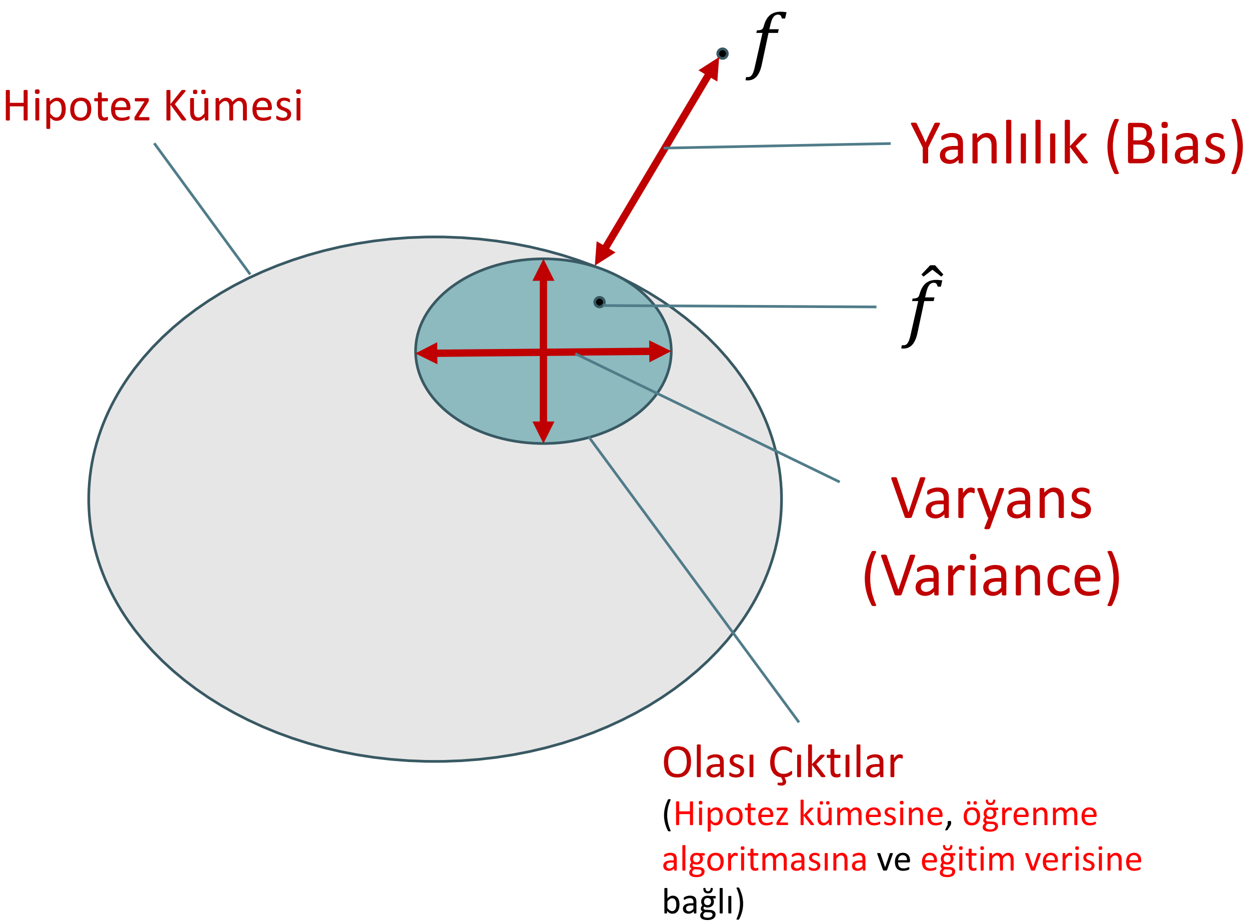
Sonuç

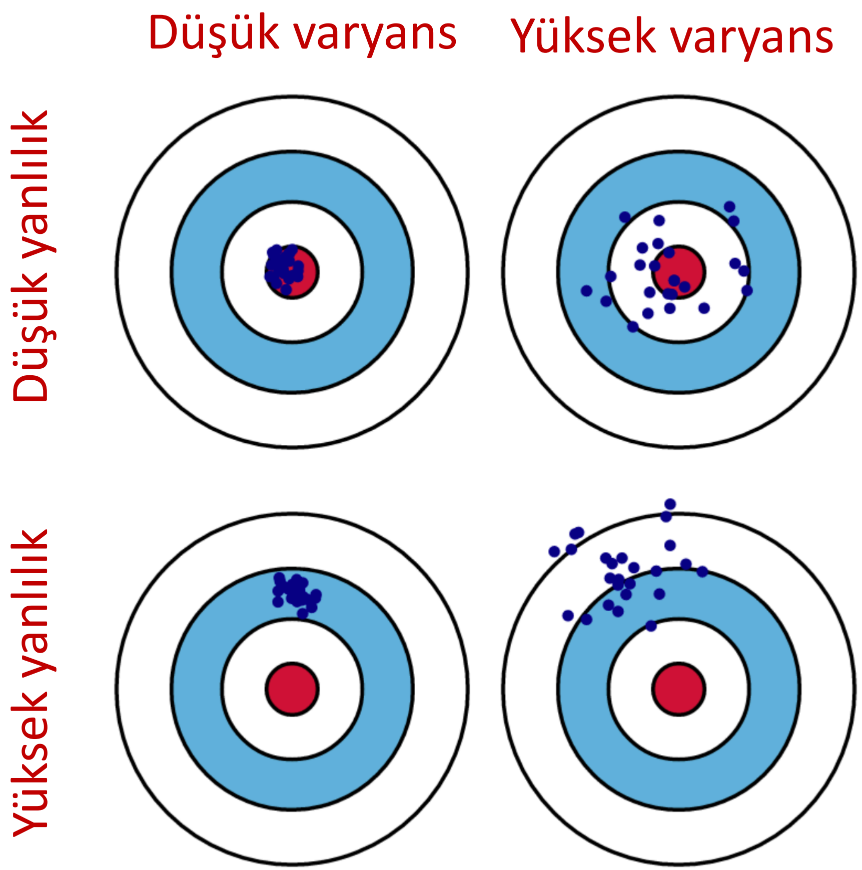
$$\hat{f} : X \to Y \qquad \hat{f}(x) \sim f(x)$$
$$x \mapsto \hat{f}(x)$$

Hipotez Seti:

- Bütün olası fonksiyonları değerlendirmek imkansız
- Bütün olası fonksiyonları değerlendirmek en iyi sonucu vermeyebilir
- Bu yüzden sınırlandırılmış bir fonksiyonlar kümesi ($\mathcal{H}$) seçiyoruz.

Öğrenme Problemi





Doğrusal Bağlanım (Linear Regression)

Güdümlü öğrenmeye basit bir yaklaşım

reklam.xlsx

TV	radyo	gazete	satış
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2

- Reklam bütçesi ile satış değerleri arasında ilişki var mı?
- Reklam bütçesi ile satış değerleri arasında ilişki ne kadar güçlü?
- Satışa hangi medyalar etki ediyor?
- Her bir medyanın satışa etkisini ne kadar hassas tahmin edebiliriz?
- Gelecekteki satışları ne kadar hassas kestirebiliriz?
- Değerler arasındaki ilişki doğrusal mı?

Basit Doğrusal Bağlanım

Hatırlatma:

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

$$Y \approx \hat{f}(X)$$

Basit Bağlanım

$$Y \approx \beta_0 + \beta_1 X$$

$$\text{satış} \approx \beta_0 + \beta_1 \text{TV}$$

$$Y \approx \beta_0 + \beta_1 X \qquad (\text{ satış} \approx \beta_0 + \beta_1 \text{TV})$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

?

eğitim verisi

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

$$y_i \approx \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i}, \quad i = 1, \ldots, n$$

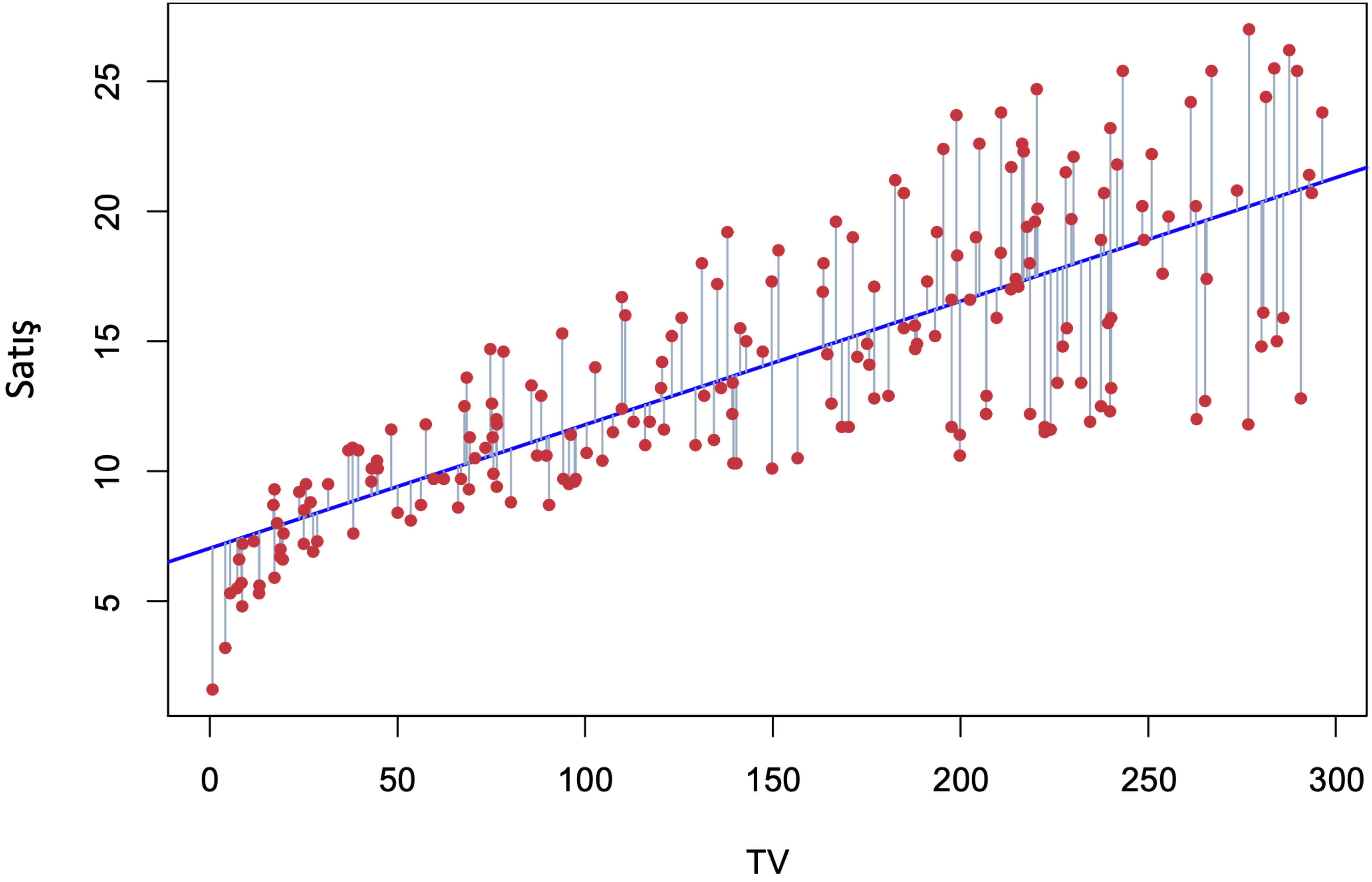
$$y_i \approx \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i}, \quad i = 1, \ldots, n$$

$$e_i = y_i - \hat{y}_i$$ *i*. kalıntı (*i*th residual)

Kalıntı Kareler Toplamı (Residual Sum of Squares)

$$\mathrm{KKT} = e_1^2 + e_2^2 + \cdots + e_n^2$$

(Amaç: Bu toplamı minimize eden $\hat{\beta}_0$ ve $\hat{\beta}_1$ değerlerini bulalım)

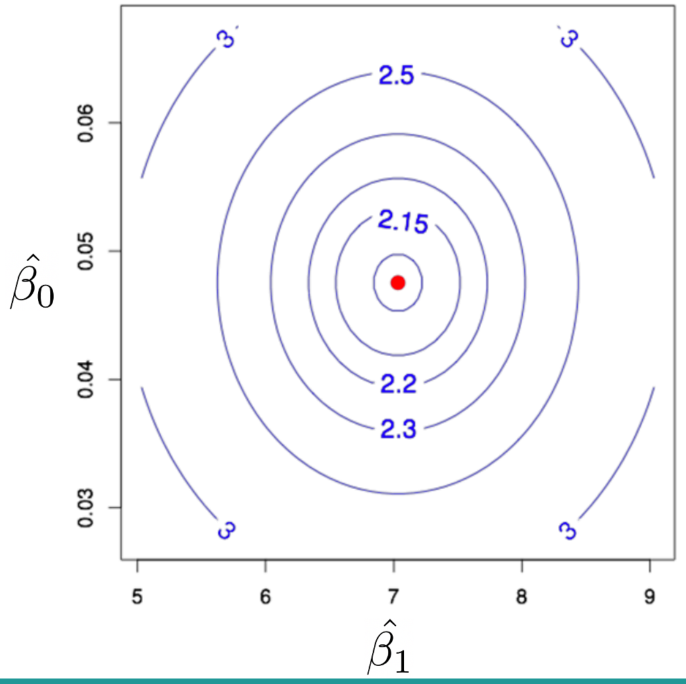

Satış
TV
0
50
100
150
200
250
300
5
10
15
20
25

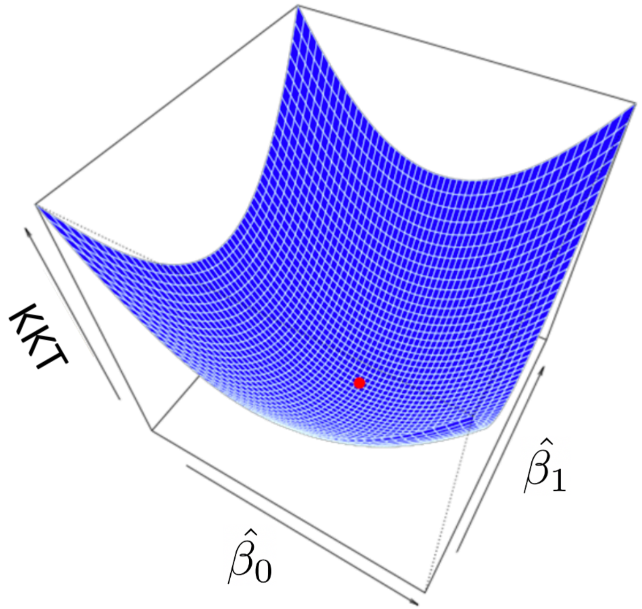
En Küçük Kareler (EKK) Yöntemi (Least Squares Method)

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \text{KKT} = (y_1-\hat{\beta}_0-\hat{\beta}_1 x_1)^2+(y_2-\hat{\beta}_0-\hat{\beta}_1 x_2)^2+\ldots+(y_n-\hat{\beta}_0-\hat{\beta}_1 x_n)^2$$

Çözüm:

Dışbükey Fonksiyon





Çözüm:

$$\frac{\partial(\mathrm{KKT})}{\partial\hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i), \qquad \frac{\partial(\mathrm{KKT})}{\partial\hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i$$

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \quad \hat{\beta}_0\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Burada $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ve $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\text{SST} = \text{SSR} + \text{SSE}$$

Total Sum of Squares)	Regression Sum of Squares	Error Sum of Squares

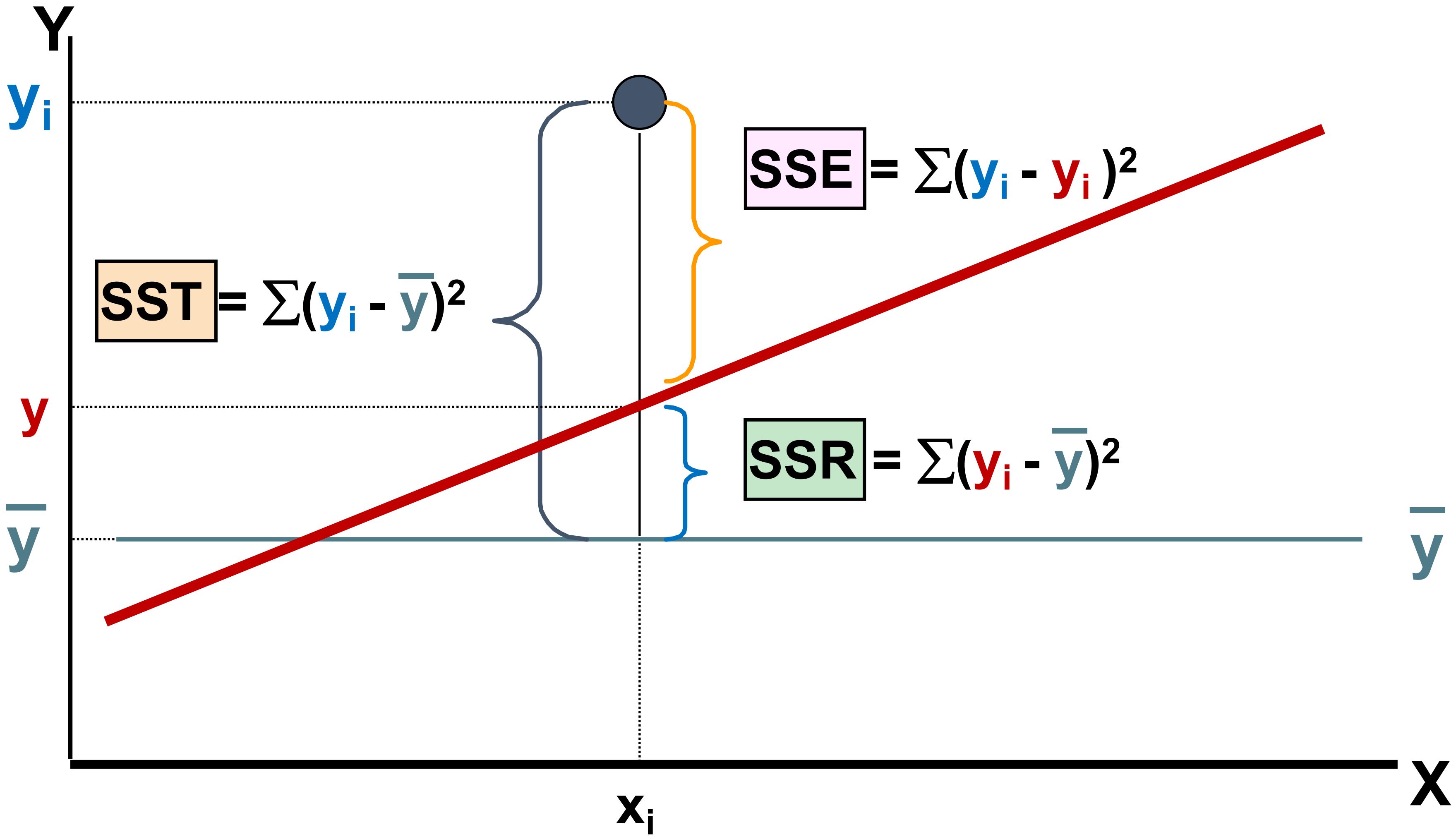
$$\boxed{\text{SST} = \sum (y_i - \bar{y})^2} \qquad \boxed{\text{SSR} = \sum (\hat{y}_i - \bar{y})^2} \qquad \boxed{\text{SSE} = \sum (y_i - \hat{y}_i)^2}$$

where:

$\bar{y}$ = y değerlerinin ortalaması

y_i = y'lerin gerçek değerleri

$\hat{y}_i$ = verilmiş x_i için y'lerin model tarafından tahmin edilmiş değeri


Y
yi
SSE = ∑(yi - yi)2
SST = ∑(yi - y)2
y
SSR = ∑(yi - y)2
y
y
X
xi

R-kare (R-squared) veya R^2

y_i değerlerindeki değişimin ne kadarının modelimiz tarafından açıklandığını ölçer!

$$\boxed{R^2 = \frac{SSR}{SST} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}}$$

Not: $\boxed{0 \leq R^2 \leq 1}$

Veriyi şu fonksiyondan üretelim: $Y = 2 + 3X + \epsilon$

Rastgele hata

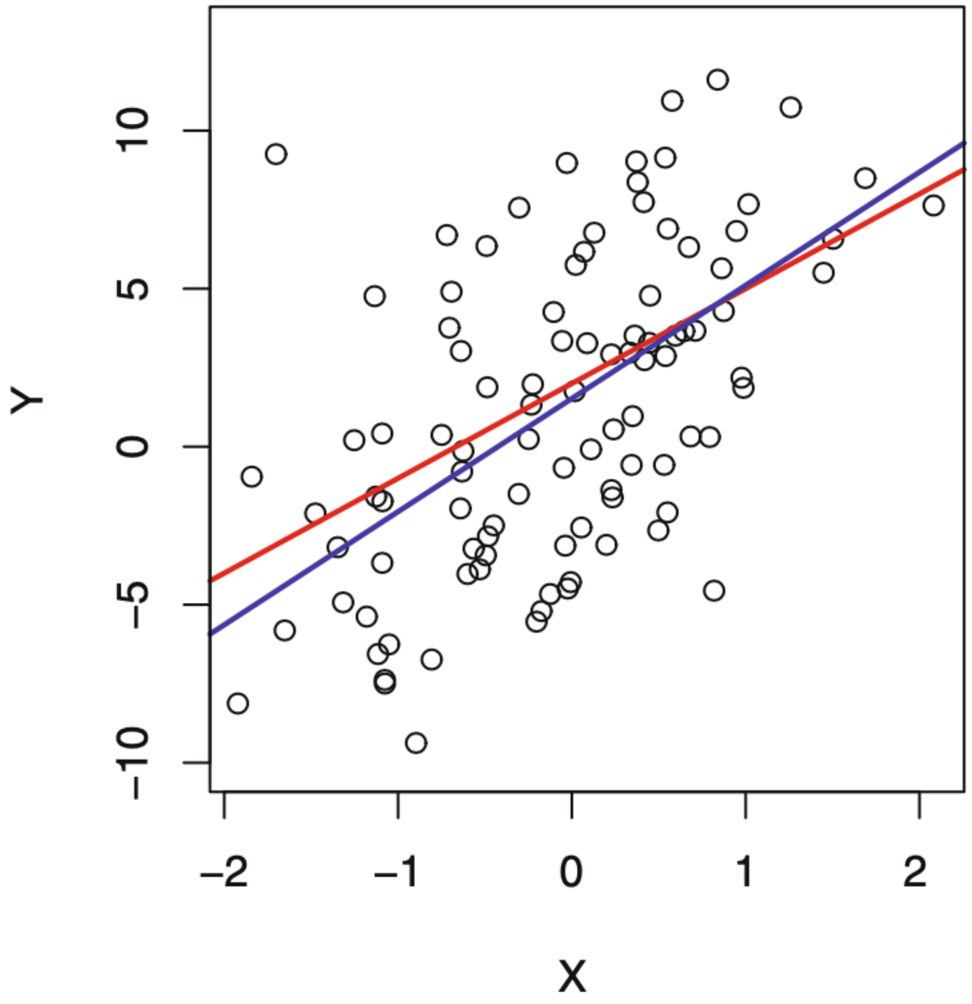
$\{(x_i, y_i): i = 1,\ldots, n\}$

Üretilmiş veri seti

Kırmızı doğru: Gerçek ilişki

$$f(X) = 2 + 3X$$

Mavi doğru: $f(X)$ fonsiyonuna EKK yöntemi ile yaklaşım



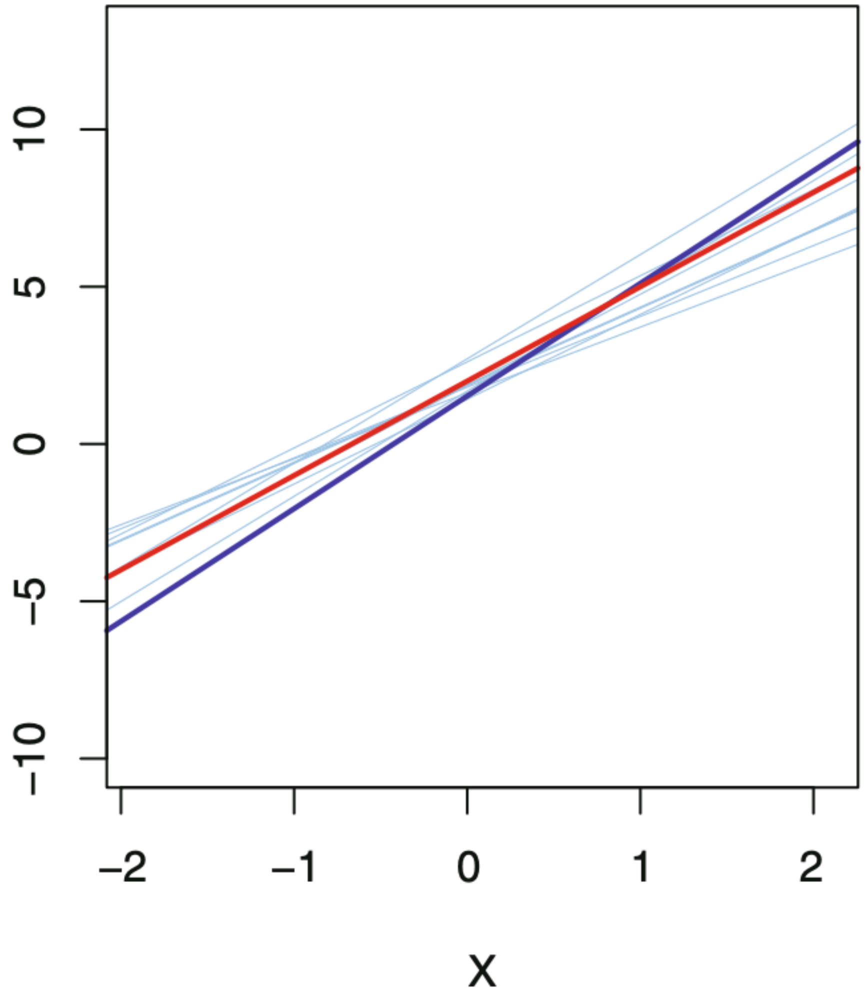
Yeni veri üretip ek küçük kareler (EKK) yöntemi ile yeni veriye karşılık gelen bağlanım doğrularını bulalım:

$$\{(x_i^{(1)}, y_i^{(1)}) \;:\; i = 1, \ldots, n\} \quad \overset{\text{EKK}}{\Longrightarrow} \quad \hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)}$$

$$\{(x_i^{(2)}, y_i^{(2)}) \;:\; i = 1, \ldots, n\} \quad \overset{\text{EKK}}{\Longrightarrow} \quad \hat{\beta}_0^{(2)}, \hat{\beta}_1^{(2)}$$

$$\vdots$$

$$\{(x_i^{(N)}, y_i^{(N)}) \;:\; i = 1, \ldots, n\} \quad \overset{\text{EKK}}{\Longrightarrow} \quad \hat{\beta}_0^{(N)}, \hat{\beta}_1^{(N)}$$



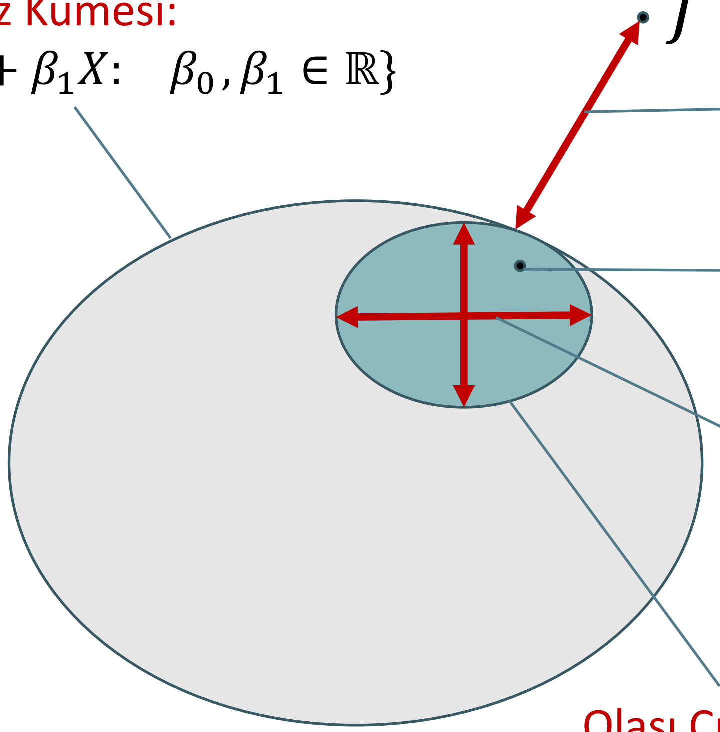
Veri kümesi değiştikçe bağlanım doğrusu da değişiyor!

Hatırlatma:

Hipotez Kümesi:
$\{\beta_0 + \beta_1 X : \quad \beta_0, \beta_1 \in \mathbb{R}\}$

f

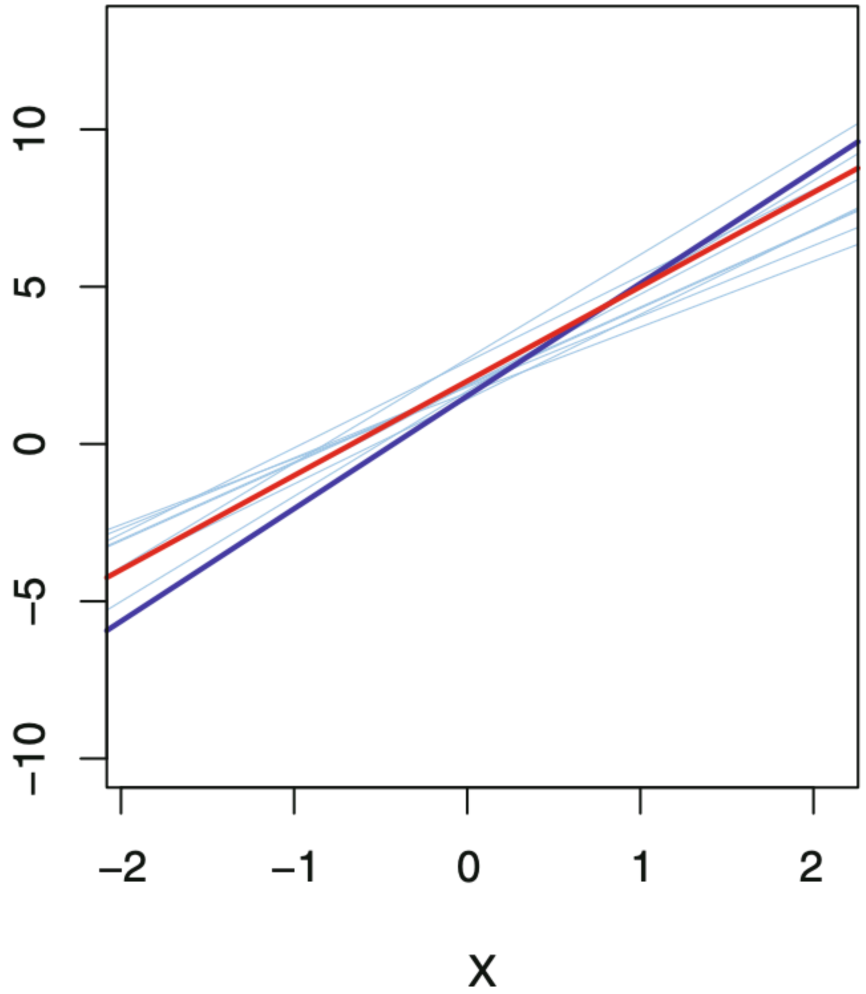
Yanlılık (Bias)



$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$

Varyans
(Variance)

Olası Çıktılar
(Hipotez kümesine, öğrenme algoritmasına ve eğitim verisine bağlı)



Basit Doğrusal Bağlanım Örneği

Reklam verisi: *reklam.xlsx*

TV	radyo	gazete	satış
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2

$$\text{satış} \approx \beta_0 + \beta_1 \text{TV}$$

Basit Doğrusal Bağlanım Örneği

[1]

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
```

```
reklam_verisi = pd.read_excel('/content/drive/My Drive/Colab Data/reklam.xlsx')

reklam_verisi.head()
```

	TV	radyo	gazete	satış
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

Basit Doğrusal Bağlanım Örneği

```
x = np.array(reklam_verisi['TV']).reshape(-1, 1)
y = np.array(reklam_verisi['satış'])

regr = linear_model.LinearRegression(fit_intercept =True)

regr.fit(x, y)

y_tahmin = regr.predict(x)
```

```
print('y =', regr.intercept_, ' + ', regr.coef_[0], '* x')
print('Ortalama kare hata: %.2f' % mean_squared_error(y, y_tahmin))
print('R-kare: %.2f' % r2_score(y, y_tahmin))
```

```
y = 7.032593549127695  +  0.04753664043301975 * x
Ortalama kare hata: 10.51
R-kare: 0.61
```

$$= \frac{KKT}{n}$$

Basit Doğrusal Bağlanım Örneği

$$\text{satış} \approx \beta_0 + \beta_1 \text{TV}$$

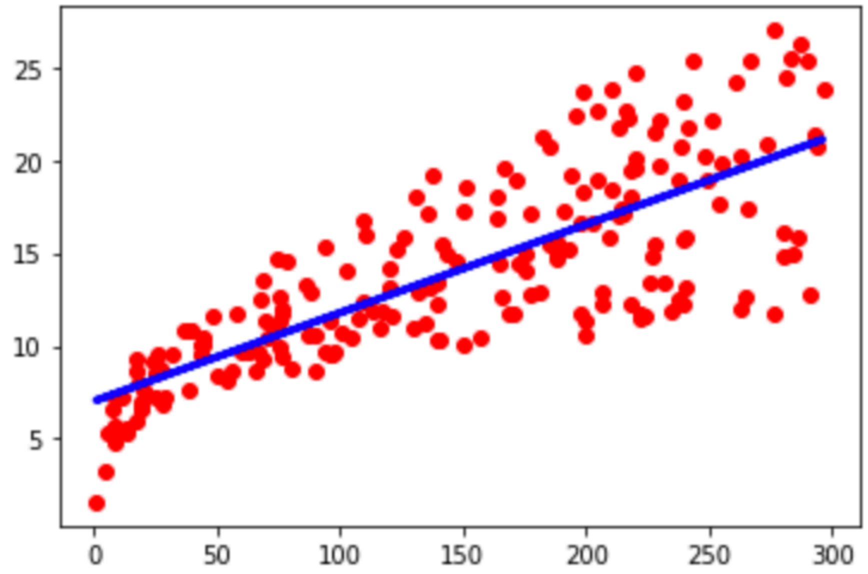
$$\text{satış} \approx 7.0325 + 0.0475\,\text{TV}$$

Televizyona verilecek her **1000 $**'lık fazladan reklam için yaklaşık **47** tane fazladan ürün satılmasını bekliyoruz.

Basit Doğrusal Bağlanım Örneği

```python
plt.scatter(x, y,  color='red')
plt.plot(x, y_tahmin, color='blue', linewidth=3)

plt.show()
```



Doğrusal Bağlanımdan Beklentilerimiz

Kestirim

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

Ne?

EVET

Çıkarım

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

Nasıl?

EVET