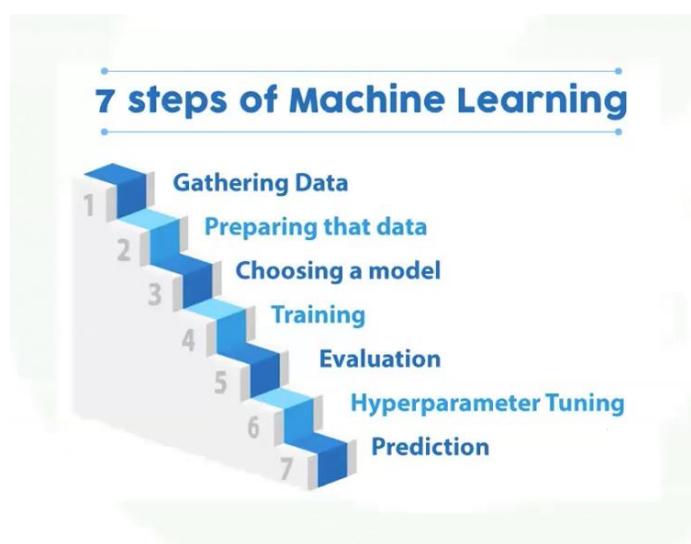
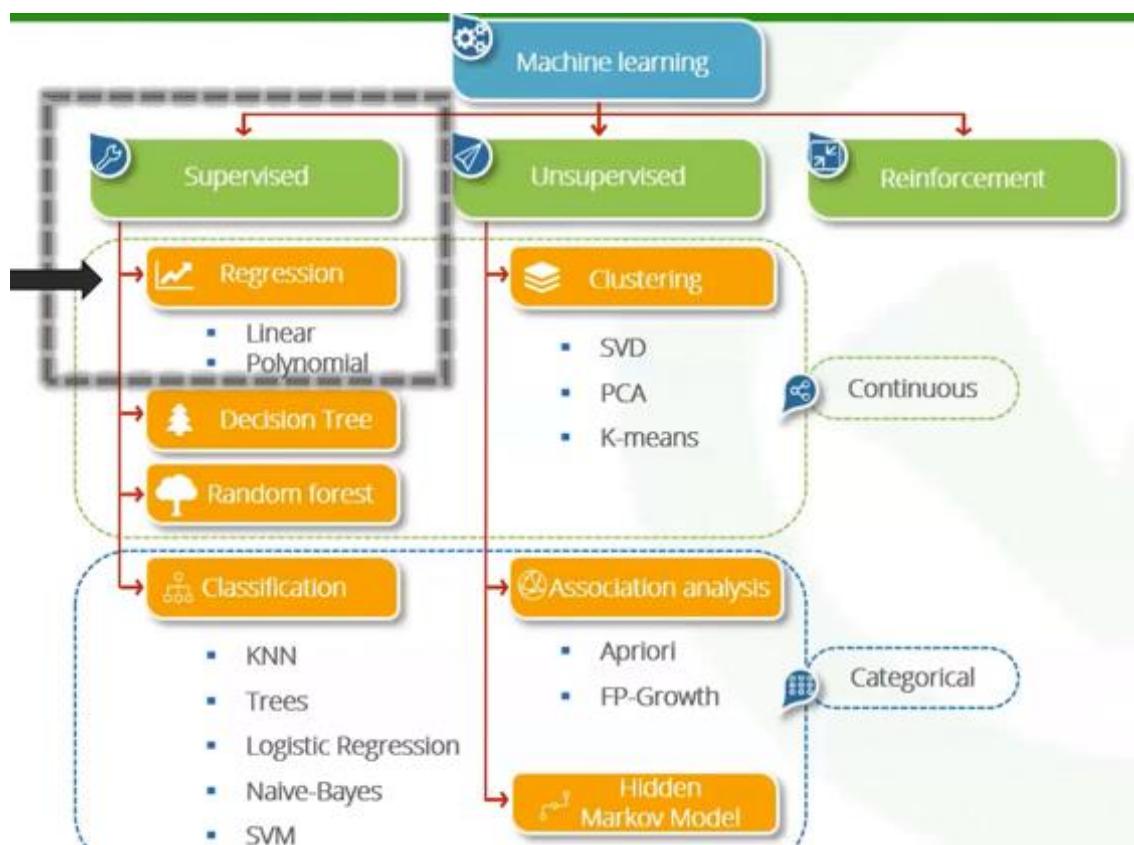


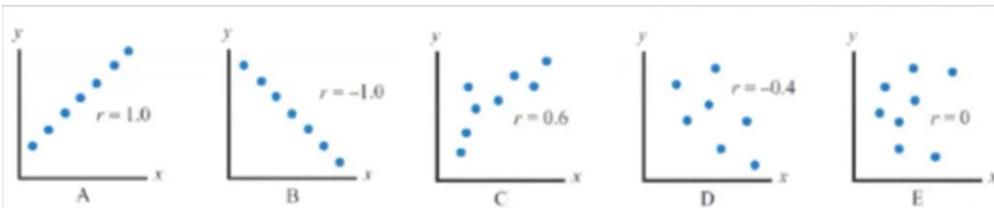
MACHINE LEARNING

Kütüphanesi scikit learn

Python en büyük kütüphanelerinden birisidir



Supervised Learning



Graph A ($r = 1.0$): perfect positive correlation between x and y

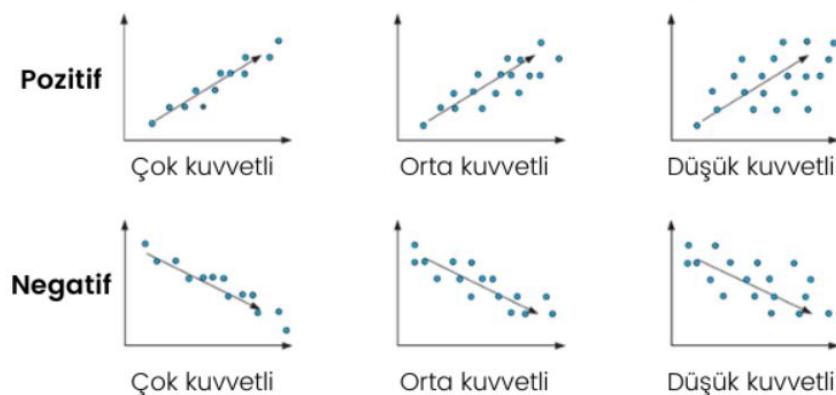
Graph B ($r = -1.0$): perfect negative correlation between x and y

Graph C ($r = 0.6$): a moderately positive relationship: y tends to increase as x increases, but not necessarily at the steady rate we observed in Graph A

Graph D ($r = -0.4$): a relatively weak negative relationship: the correlation coefficient is closer to zero, negative r value so y tends to decrease as x increases

Graph E ($r = 0$): no relationship between x and y

Doğrusal (Lineer) Korelasyon



Korelasyon katsayı değeri	Yorumu
0.9 ile 1 veya -0.9 ile -1 arası	Çok kuvvetli doğrusal ilişki
0.7 ile 0.9 veya -0.7 ile -0.9 arası	Kuvvetli doğrusal ilişki
0.5 ile 0.7 veya -0.5 ile -0.7 arası	Orta kuvvetli doğrusal ilişki
0.3 ile 0.5 veya -0.3 ile -0.5 arası	Düşük kuvvetli doğrusal ilişki
0.0 ile 0.3 veya 0.0 ile -0.3 arası	Hiç ilişki yok veya çok düşük kuvvetli doğrusal ilişki

Pipeline: iş hattı gibi düşün

1. Regresyon Nedir?

Regresyon, **sürekli (numerik) bir değeri tahmin etmek** için kullanılan bir makine öğrenmesi tekniğidir.

- Bağımsız değişkenler (özellikler) ile bağımlı değişken (hedef) arasındaki ilişkileri öğrenir.
- Çıkış değişkeni **sayı (kesintisiz değer)** olmalıdır.
- Kullanım alanları: Fiyat tahmini, hava sıcaklığı tahmini, satış miktarı tahmini vb.

Örnek Regresyon Problemleri:

- Bir evin **fiyatını** tahmin etmek (Metrekare, oda sayısı gibi değişkenlere bağlı olarak).
- Önümüzdeki ay bir markette **kaç ürün satılacağını** tahmin etmek.
- Bir çalışanın **maasını** deneyim yılına göre tahmin etmek.

Yaygın Regresyon Algoritmaları:

- **Linear Regression (Doğrusal Regresyon)**
- **Polynomial Regression (Polinom Regresyon)**
- **Ridge Regression, Lasso Regression**
- **Decision Tree Regressor, Random Forest Regressor**
- **XGBoost, LightGBM (Gelişmiş ağaç tabanlı algoritmalar)**

Linear Regression (Doğrusal Regresyon)

Tanım:

Linear Regression, **bağımsız değişkenler (X) ile bağımlı değişken (Y) arasında doğrusal bir ilişki olduğunu varsayan ve sürekli (numerik) bir değer tahmin eden bir makine öğrenmesi modelidir.**

Simple lineer: $y = ax + b$ dayanır

Gerçek Dünya Örnekleri:

❖ Örnek 1: Ev Fiyatı Tahmini

- **Problem:** Bir evin fiyatını, metrekare, oda sayısı ve bulunduğu semte göre tahmin etmek istiyoruz.
- **Bağımsız Değişkenler (X):**
 - **Metrekare büyüklüğü** (100 m^2 , 150 m^2 , 200 m^2 ...)
 - **Oda sayısı** (2 oda, 3 oda, 4 oda...)
 - **Semt puanı** (0-10 arasında bir değer)
- **Bağımlı Değişken (Y):**
 - **Ev fiyatı** ($\text{₺}500,000$, $\text{₺}750,000$, $\text{₺}1,000,000$...)

Modelin tahmini şöyle olabilir:

$$\begin{aligned} \text{EvFiyatı} &= 50,000 + 3,000 \times (\text{Metrekare}) + 10,000 \times (\text{OdaSayısı}) + 20,000 \times (\text{SemtPuanı}) \\ |, \text{Fiyatı} &= 50,000 + 3,000 \times (\text{Metrekare}) + 10,000 \times (\text{Oda Sayısı}) + 20,000 \times (\text{Semt Puanı}) \\ \text{EvFiyatı} &= 50,000 + 3,000 \times (\text{Metrekare}) + 10,000 \times (\text{OdaSayısı}) + 20,000 \times (\text{SemtPuanı}) \end{aligned}$$

- **100 m², 3 odalı, semt puanı 8 olan bir ev için:**

$$\begin{aligned} \text{EvFiyatı} &= 50,000 + 3,000(100) + 10,000(3) + 20,000(8) = 840,000 \text{ TL} \\ |, \text{Fiyatı} &= 50,000 + 3,000(100) + 10,000(3) + 20,000(8) = 840,000 \text{ TL} \\ \text{EvFiyatı} &= 50,000 + 3,000(100) + 10,000(3) + 20,000(8) = 840,000 \text{ TL} \end{aligned}$$

❖ Örnek 2: Maaş Tahmini

- **Problem:** Çalışanların maaşlarını, deneyim yılına göre tahmin etmek istiyoruz.
- **Bağımsız Değişkenler (X):**
 - **Deneyim yılı** (1, 2, 3, 5, 10 yıl...)
- **Bağımlı Değişken (Y):**
 - **Maaş (TL)** ($\text{₺}10,000$, $\text{₺}15,000$, $\text{₺}25,000$...)

Model şöyle olabilir:

$$\begin{aligned} \text{Maas} | c &= 5,000 + 2,500 \times (\text{Deneyim Yılı}) \\ \text{Maas} &= 5,000 + 2,500 \times (\text{Deneyim Yılı}) \\ |, \text{Maas} &= 5,000 + 2,500 \times (\text{DeneyimYılı}) \end{aligned}$$

- **5 yıl deneyimi olan biri için:** $\text{Maas} | c = 5,000 + 2,500(5) = 17,500 \text{ TL}$ $\text{Maas} = 5,000 + 2,500(5) = 17,500 \text{ TL}$



Simple Linear Regression



Scott'ın almak istediği 9000 km'lik aracın fiyat tahminini de yine bu modelden yapabilir:

$$\text{Fiyat} = 93.048 - 1,385 (9000) = 80.583$$

Yine bu modele göre yeni ilanda yer alan 12.000 km'lik aracın fiyat tahminini yapabiliriz:

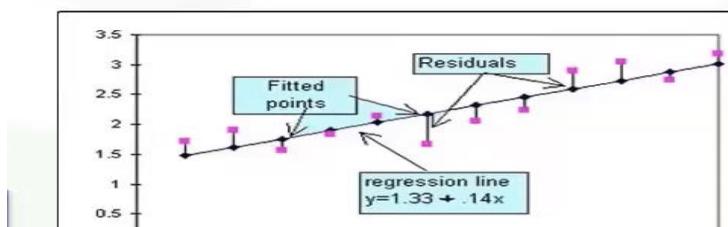
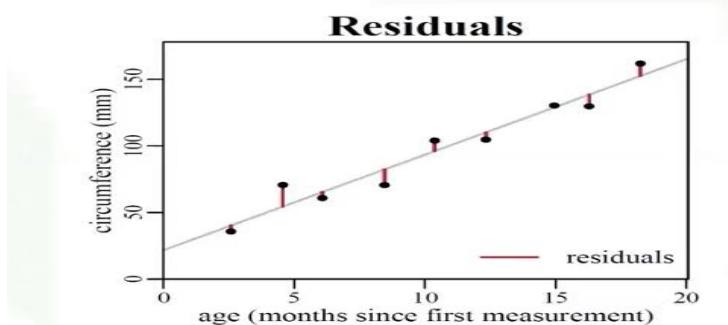
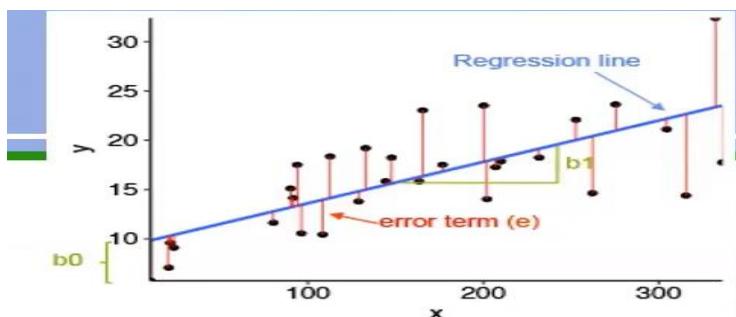
$$\text{Fiyat} = 93.048 - 1,385 (12000) = 76.478$$

Aynı şekilde Scott'in 10.000 km'deki aracı için bir satış fiyatı belirleyebiliriz:

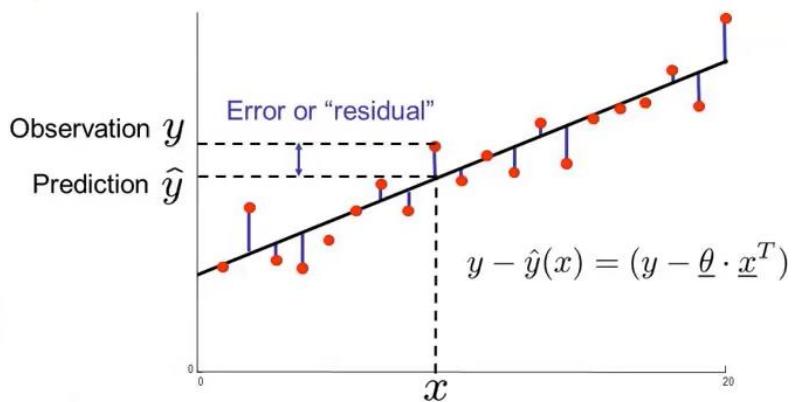
$$\text{Fiyat} = 93.048 - 1,385 (10000) = 79.198$$

- Peki, bu algoritma makine Öğrenmesi ile alternatifler arasından en doğru olanı nasıl belirledi?

Sıra	Fiyat (₺)	KM	Tahmin Edilen Fiyat (₺)
1	86000	6000	84737
2	82000	8200	81690
3	78000	9000	80582
4	75000	14200	73379
5	70000	16200	70609



Measuring error



İçin Dilevarılar İhbar

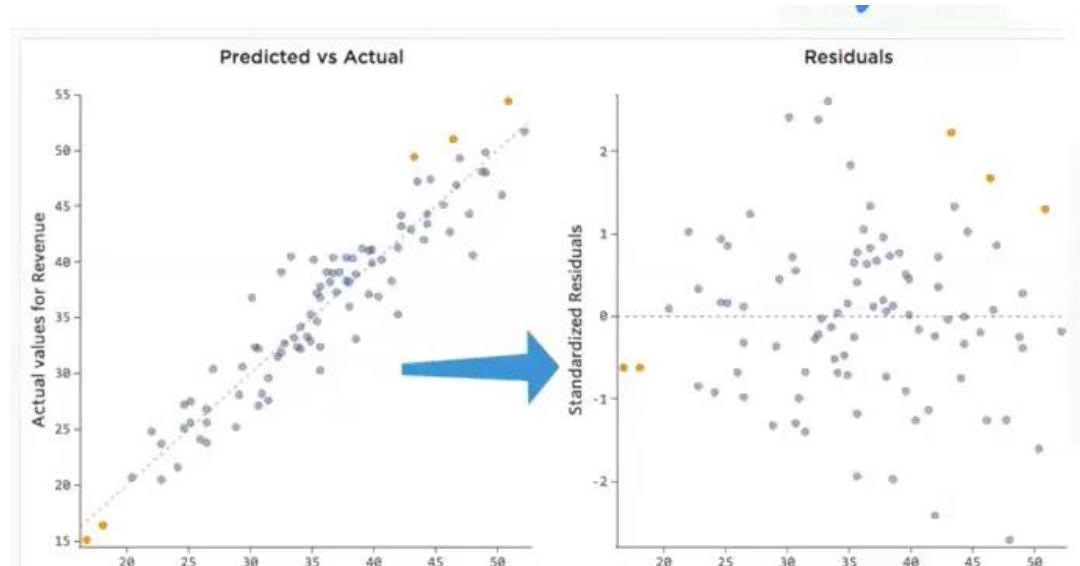
Residuals: gerçek noktaya doğru üzerinde denk geldiği nokta arasındaki fark

Error de denilebilir

Amacımız bu residuals'ları minimize etmek

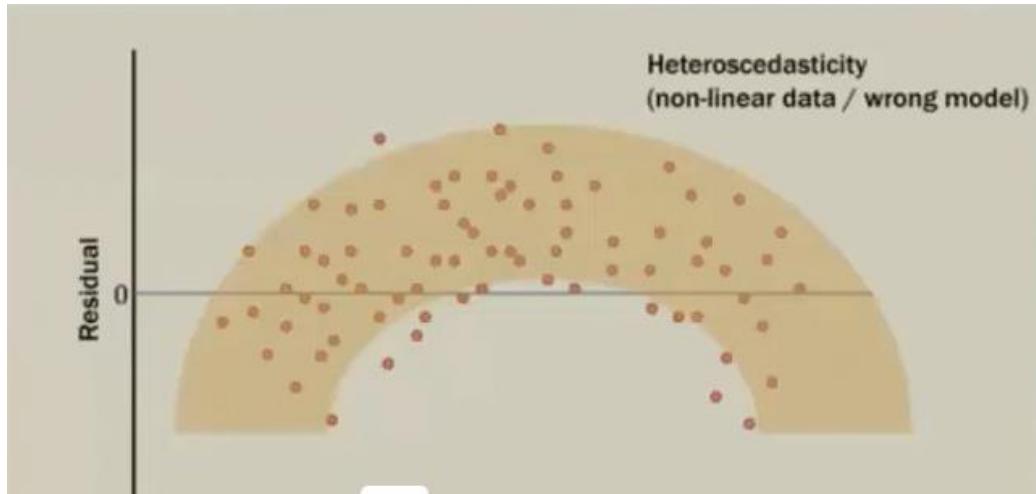
Bilgi kaybına sebep olurlar

Standardize edersek dağılım bu şekilde olur



pattern sergilememeli

Bu şekilde olursa nonlinear data olduğunu gösterir



Regresyon modelleri için performans değerlendirme ölçütleri

ERROR METRICS

MAE MSE RMSE R2

Regresyon Modelleri İçin Performans Değerlendirme Ölçütleri

Mean Absolute Error (MAE) Ortalama Mutlak Hata

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

n: Gözlem Sayısı
Her bir gerçek değer
Her bir tahmin edilmiş değer

Divide by the total number of data points
Actual output value
Predicted instant value
The absolute value of the residual

Mean Squared Error (MSE) Ortalama Kare Hata

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

n: Gözlem Sayısı
Her bir gerçek değer
Her bir tahmin edilmiş değer

Mean Squared Error
• Average (squared) deviation from truth $\sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$
• Very sensitive to outliers
• 99 exact, 1 off by \$10^6\$ → large effect on models
• Sensitive to mean / scale
• $\mu_y = \bar{y}, \Sigma y_i \dots$ good baseline
• Relative squared error (Weka)
$$\frac{\sum (U(x_i) - \bar{y})^2}{\sum (\bar{y} - \bar{y})^2}$$

R2'ye baksan bile RMSE'ye de mutlaka bak