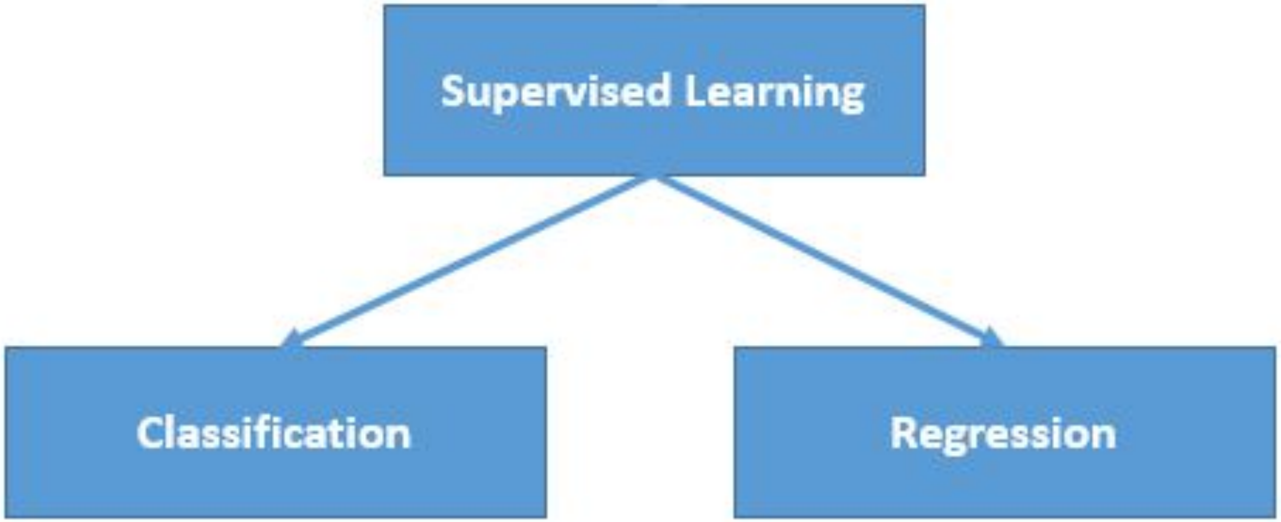
Models

Onur Sahil Cerit

Supervised Learning

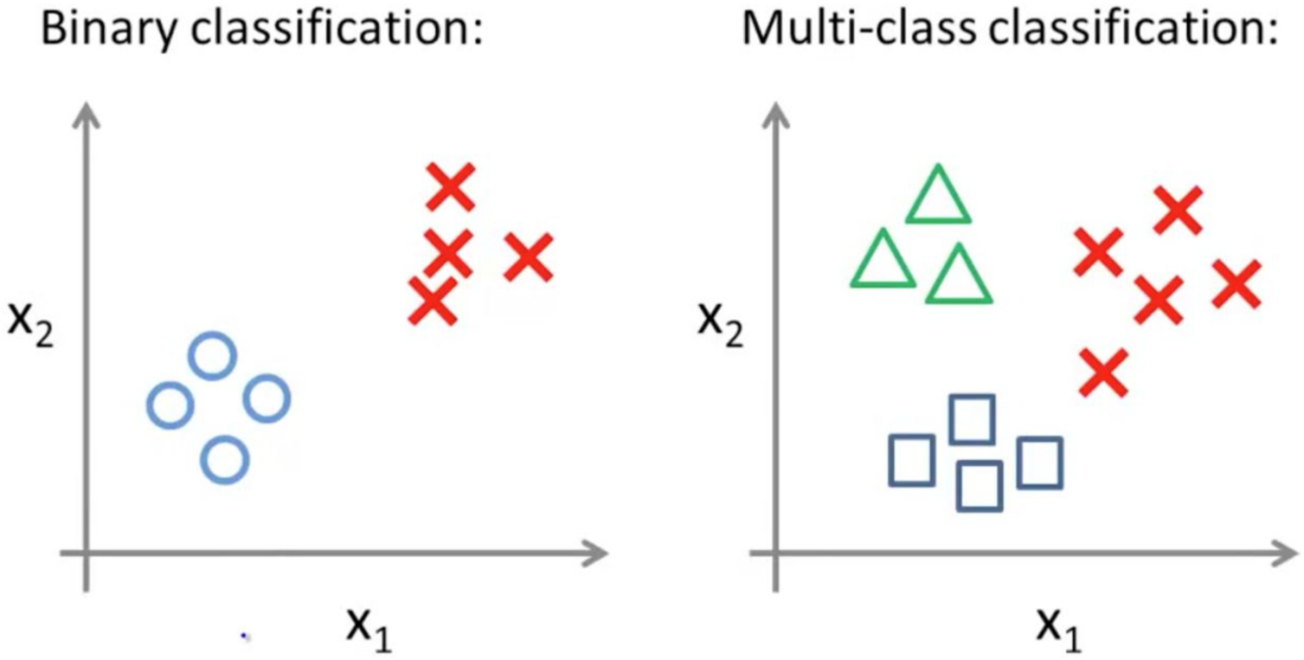


Regression Analysis

Bağımlı Değişken	Bağımsız Değişken
• Bu değişken tahmin etmeye çalıştığımız değişkendir	• Bu değişken tahmin etmek için kullandığımız giriş değişkenidir
• “y” olarak ifade edilir	• “X” olarak ifade edilir

** **<u>Bir regresyon probleminde sürekli aralıkta olan sonuçları tahmin etmeye çalışırız.</u>**

Classification Analysis



Linear Regression

Tek-değişkenli lineer regresyon modeli

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = h(x_i) + \varepsilon_i \Rightarrow \varepsilon_i = y_i - h(x_i)$$

y^i => ‘i’ numaralı gözlem için bağımlı değişken (ev fiyatı)

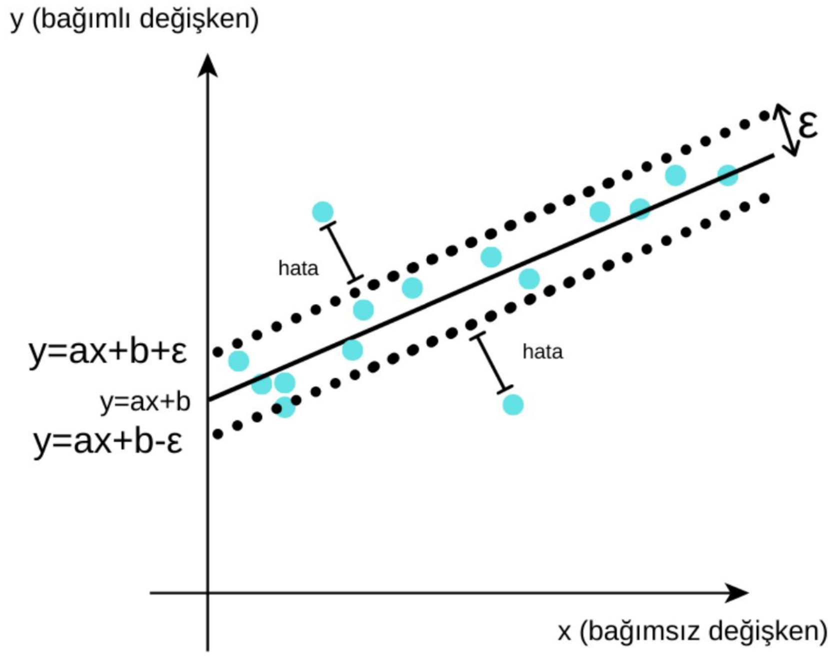
x^i => ‘i’ numaralı gözlem için bağımsız değişken (ev özellikleri)

ϵ^i => ‘i’ numaralı gözlem için hata değeri

β_0 => ‘i’ numaralı gözlem için sabit katsayı değeri

β_1 => ‘i’ numaralı gözlem için bağımsız değişkenin katsayı değeri

Linear Regression



Multiple Linear Regression

k-değişkenli çoklu lineer regresyon modeli

$$y^i = \beta_0 + \beta_1 x_1^i + \beta_2 x_2^i + \ldots + \beta_k x_k^i + \epsilon^i$$

y^i => ‘i’ numaralı gözlem için bağımlı değişken (ev fiyatı)

x_j^i => ‘i’ numaralı gözlem için j numaralı bağımsız değişken

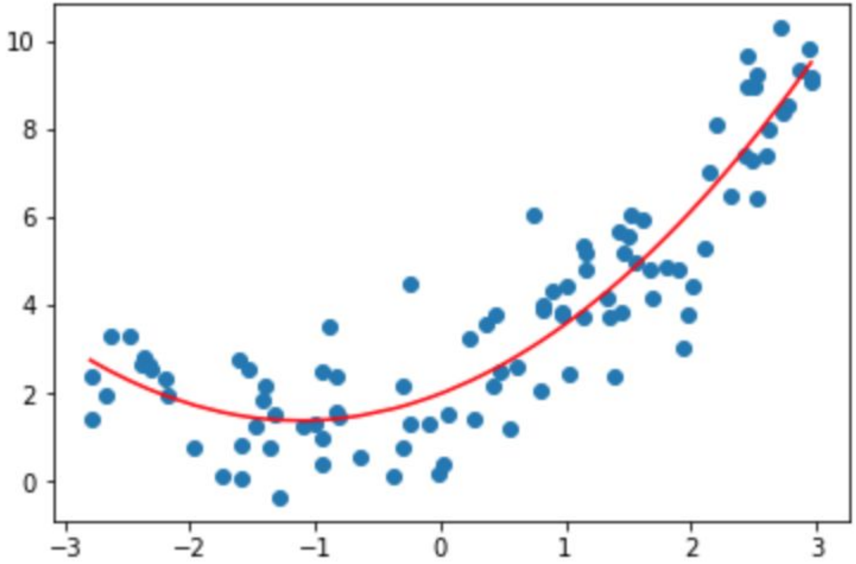
ϵ^i => ‘i’ numaralı gözlem için hata değeri

β_0 => ‘i’ numaralı gözlem için sabit katsayı değeri

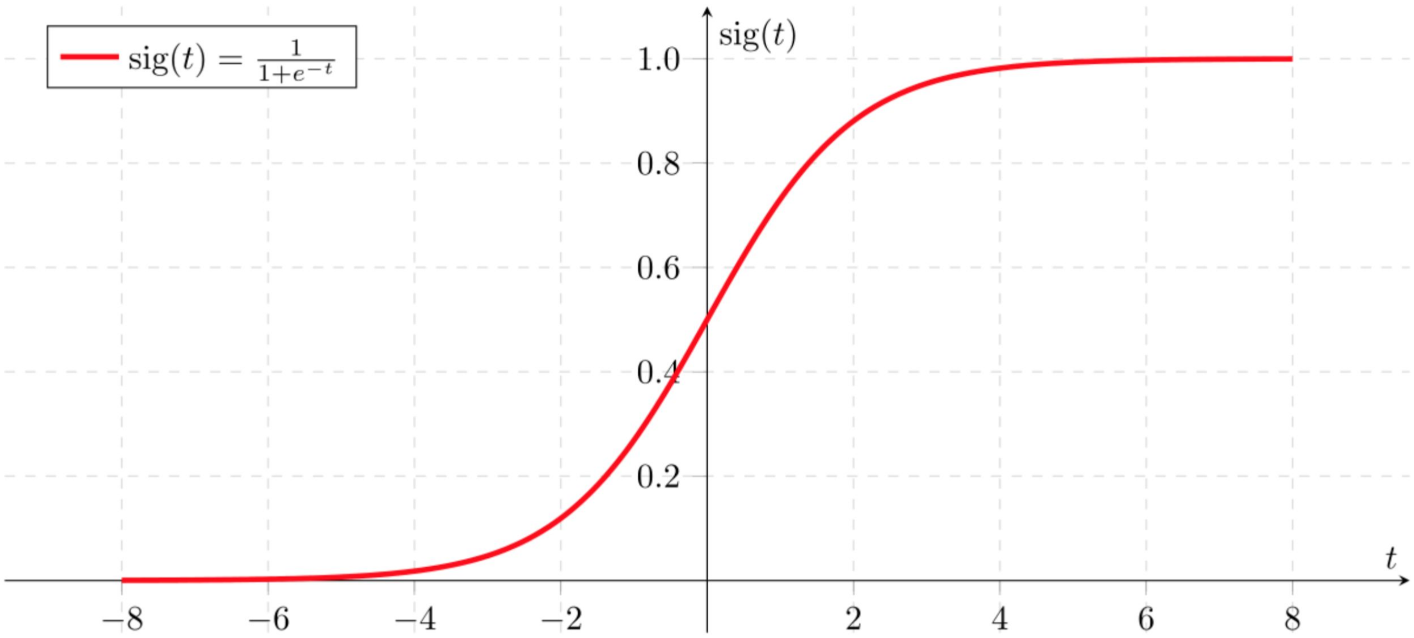
β_j => ‘j’ numaralı bağımsız değişken için regresyon katsayısı

Polinomial Regression

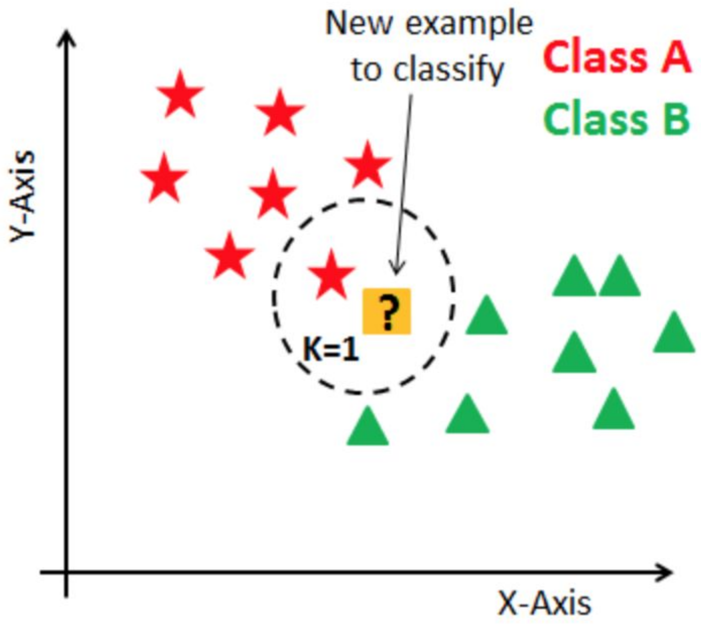
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i \ (i = 1, 2, \ldots, n)$$



Logistic Regression



k-Nearest Neighbor

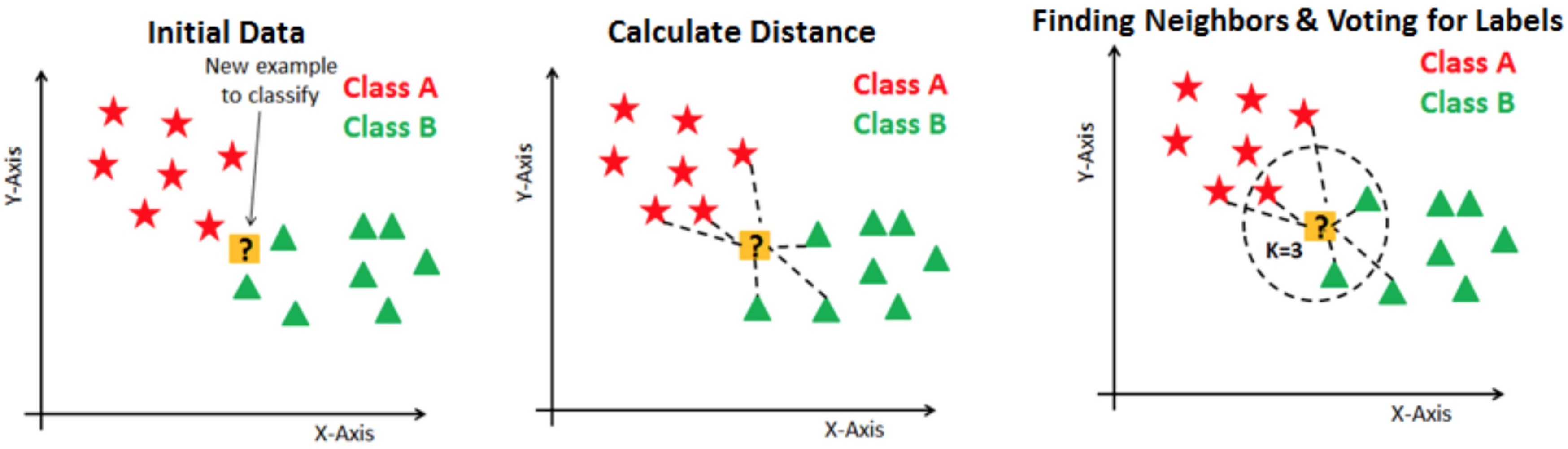


k-Nearest Neighbor

Steps:

1. Uzaklığı Hesapla
2. Yakın Komşuları Bul
3. Etiket/Sınıf için Oy Ver

k-Nearest Neighbor

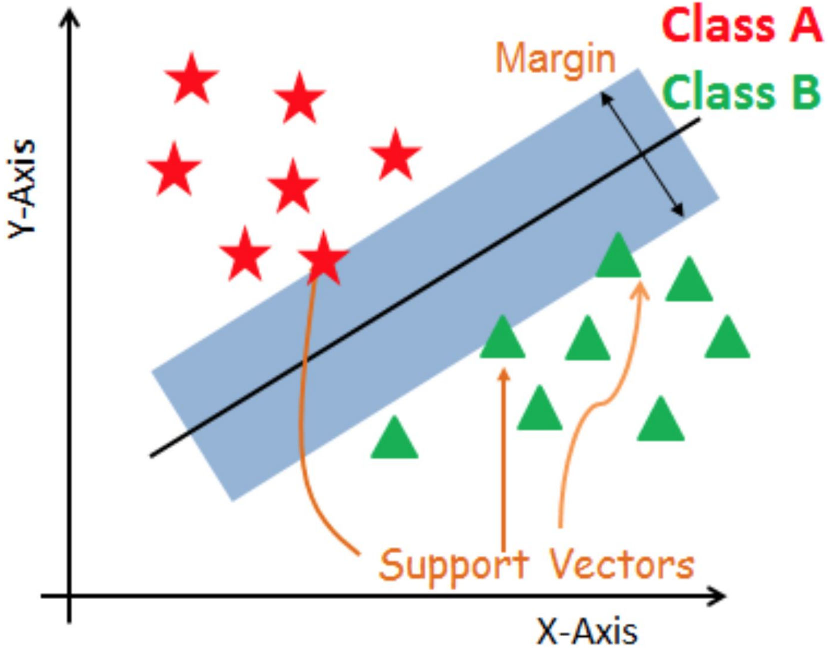


k-Nearest Neighbor - Euclidean Distance

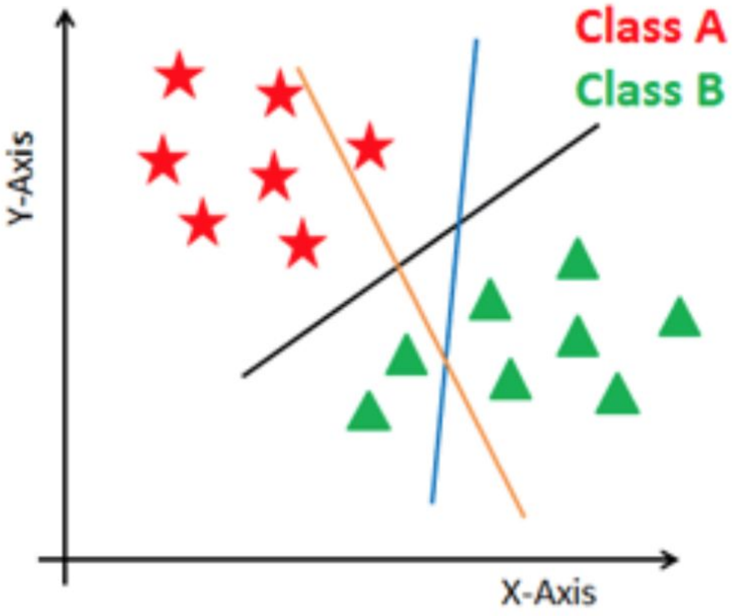
$$A = (x_1, x_2, \ldots, x_m) \qquad B = (y_1, y_2, \ldots, y_m)$$

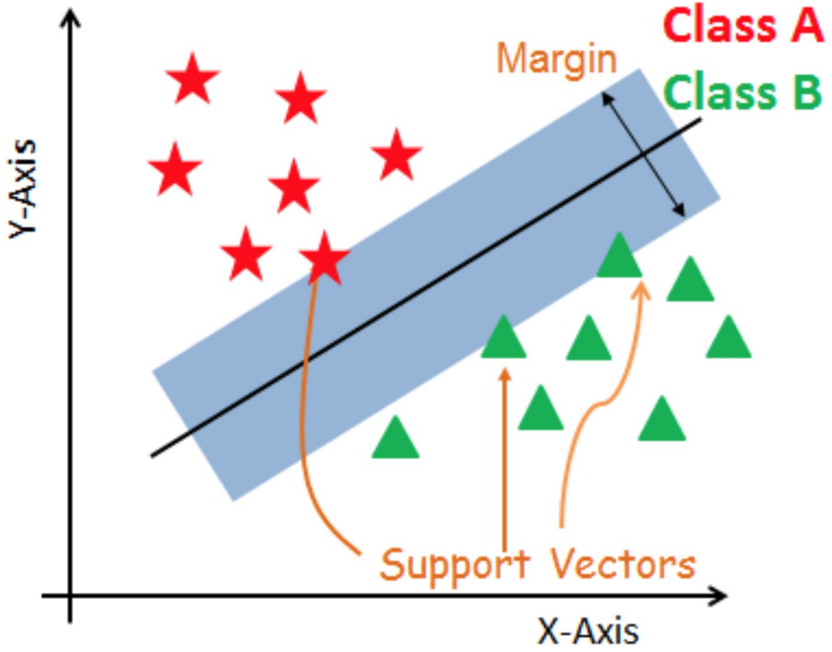
$$dist\,(A, B) = \sqrt{\frac{\sum_{i=1}^{m} (x_i - y_i)^2}{m}}$$

Support Vector Machines

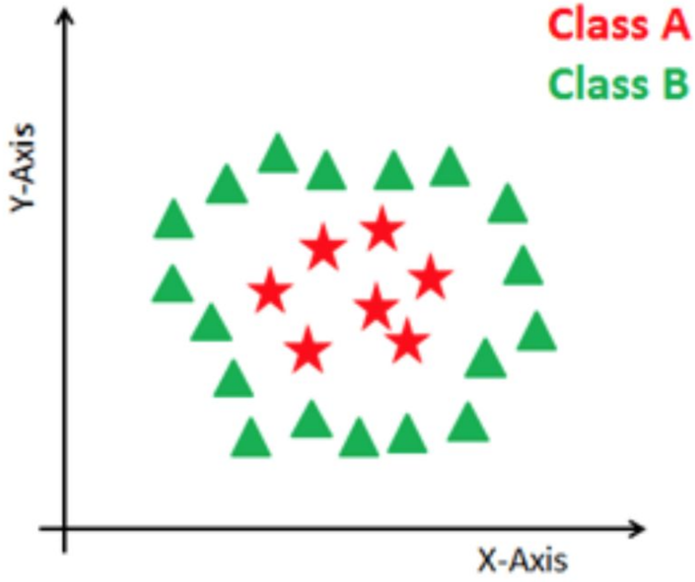


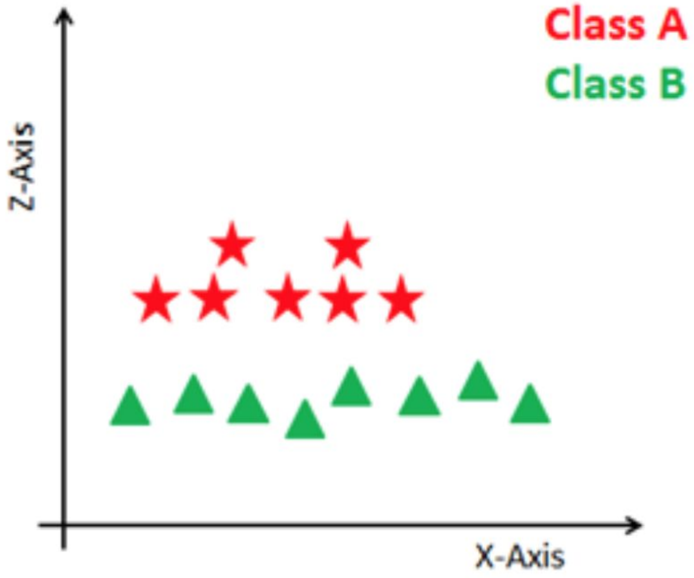
Support Vector Machines





Support Vector Machines - Kernels



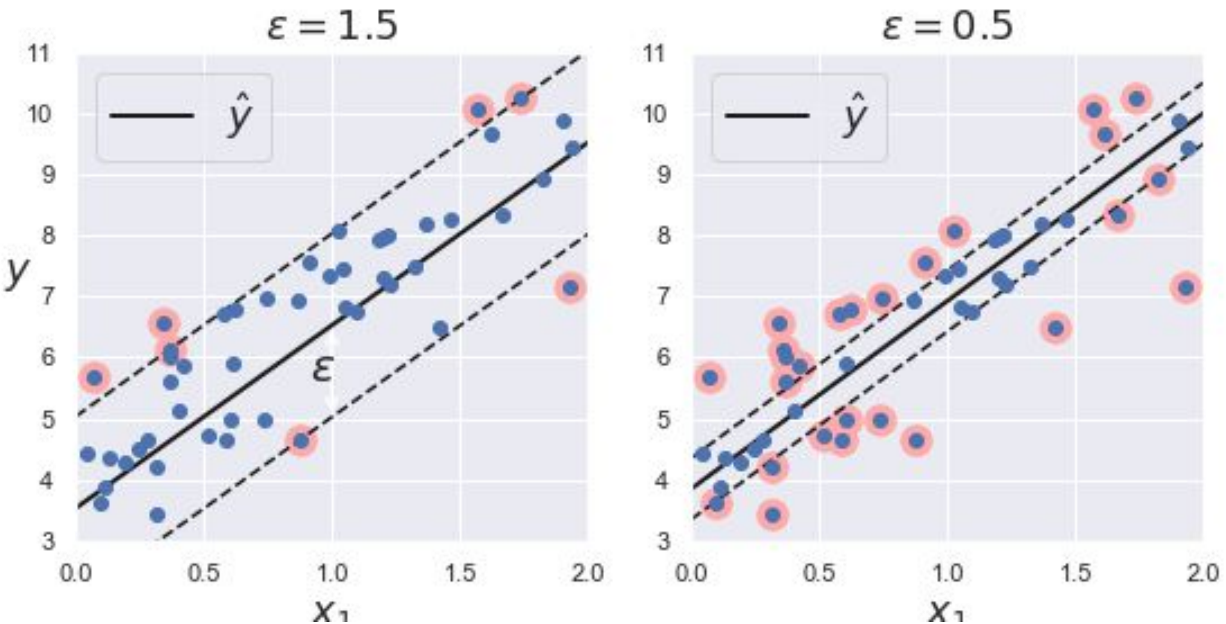


Support Vector Machines - Kernels

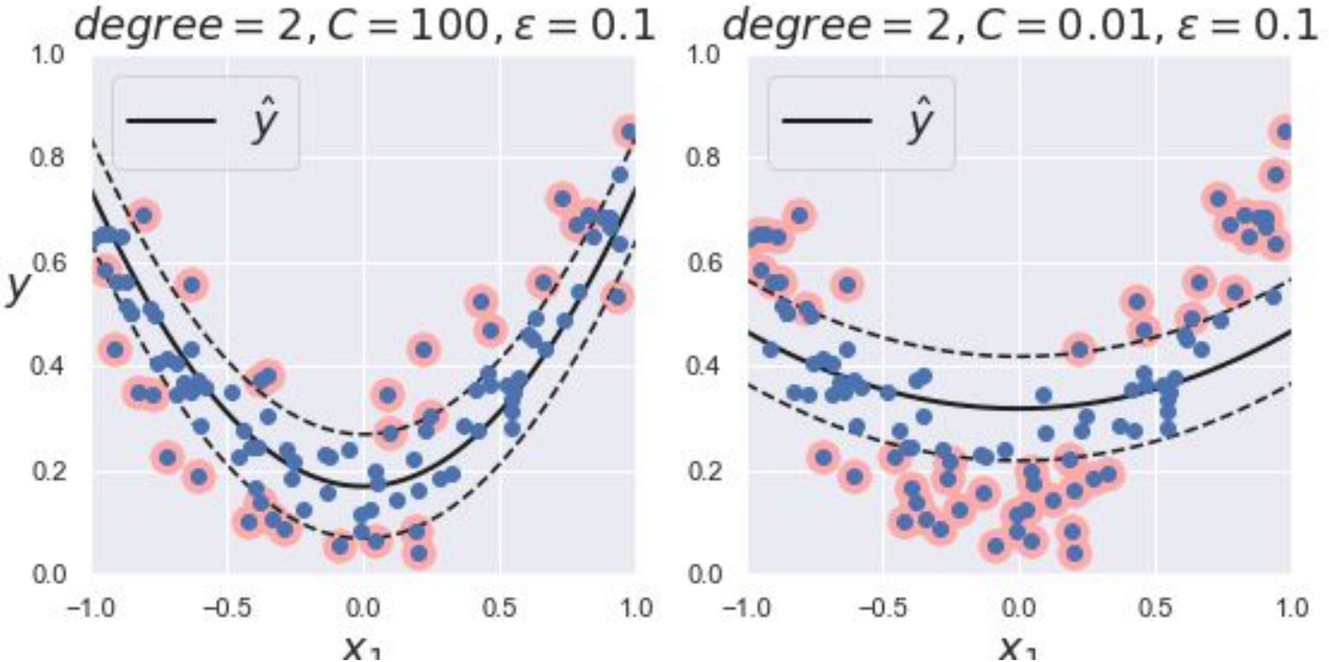
SAMPLES

Name of the Kernel	Mathematical Formula
Linear	$k(x,y) = x^T.y$
Polynomial	$k(x,y) = (x^T,y)^P$ or $k(x,y) = (x^T.y+1)^p$ where p is the polynomial degree
RBF(Gaussian)	$\phi(x) = \exp(-\frac{x^2}{2\sigma^2}), \sigma > 0$

Support Vector Machines - Regression



Support Vector Machines - Regression

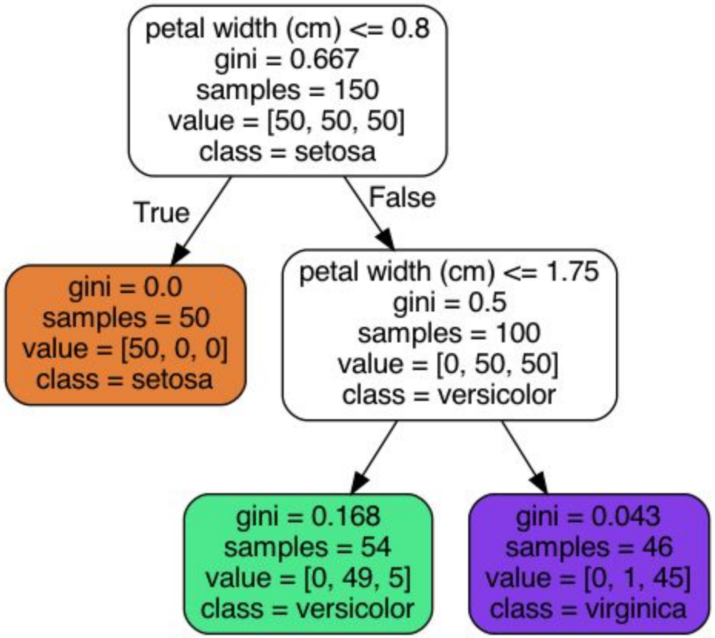


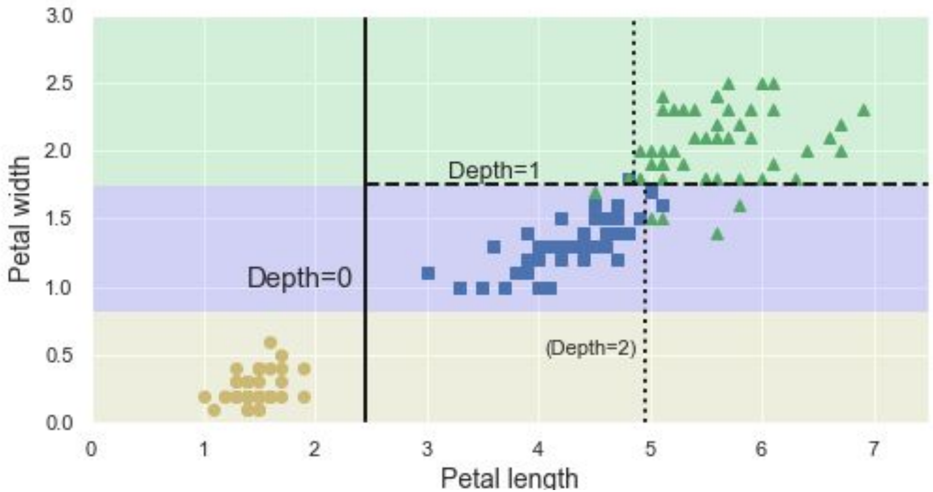
Decision Trees - Classification

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0.0
1	4.9	3.0	1.4	0.2	0.0
2	4.7	3.2	1.3	0.2	0.0
3	4.6	3.1	1.5	0.2	0.0
4	5.0	3.6	1.4	0.2	0.0

```
array(['setosa', 'versicolor', 'virginica']
```

Decision Trees - Classification





Decision Trees - Classification

Gini Impurity

***Impurity: Bir training set verisine etiket verirken o etiketin yanlış olma olma şansı (Hiç hata => Impurity = 0)**

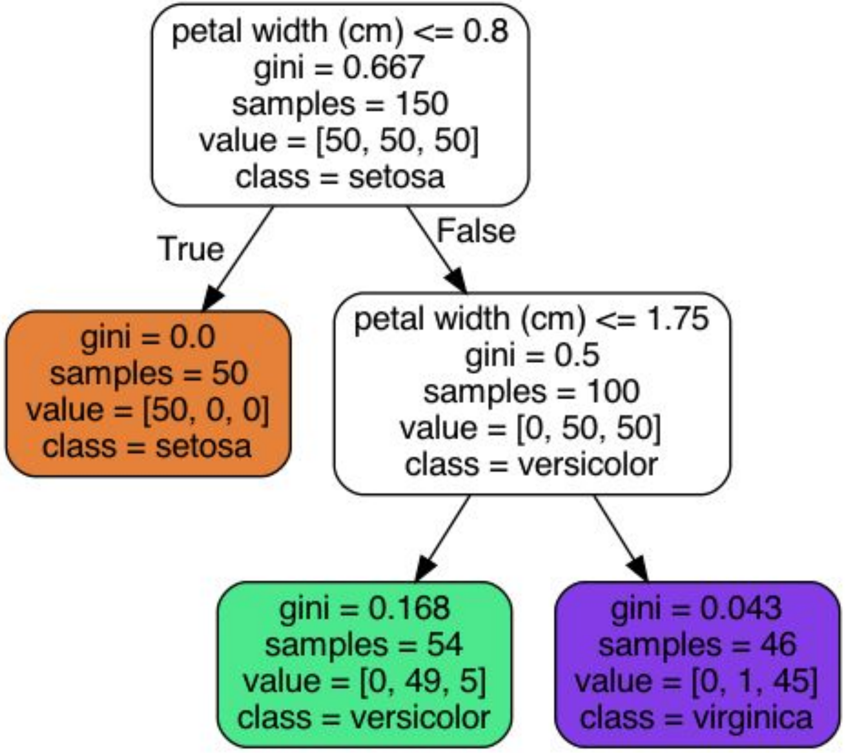
Gini Impurity, verilen bir değerin impurity(yanlış olma oranı) değerini bulur.

$$I_G(n) = 1 - \sum_{i=1}^{J} (p_i)^2$$

Decision Trees - Classification

Gini Impurity

$$I_G(n) = 1 - \sum_{i=1}^{J} (p_i)^2$$



<u>Derinlik-2 Sol Nod:</u>

1 - (0/54)^2 - (49/54)^2 - (5/54)^2 ~= 0.168

Decision Trees - Classification

Entropy (Information Gain: sorulacak en iyi soruyu bulmakta fayda sağlar)

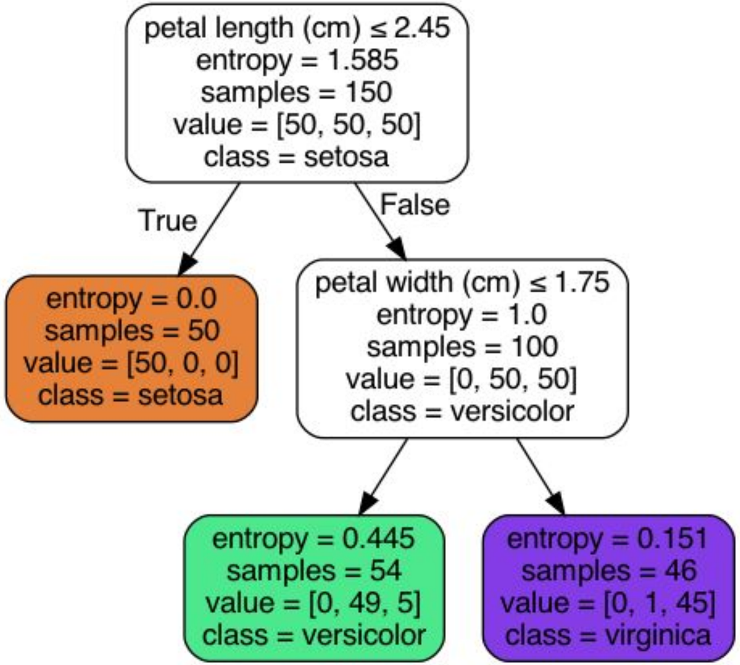
***Entropy: Makine öğrenmesinde entropy bir verinin sadece tek class değeri almasıyla 0 değerini alır.**

$$I_H = -\sum_{j=1}^{c} p_j log_2(p_j)$$

Decision Trees - Classification

Entropy

$$I_H = -\sum_{j=1}^{c} p_j log_2(p_j)$$

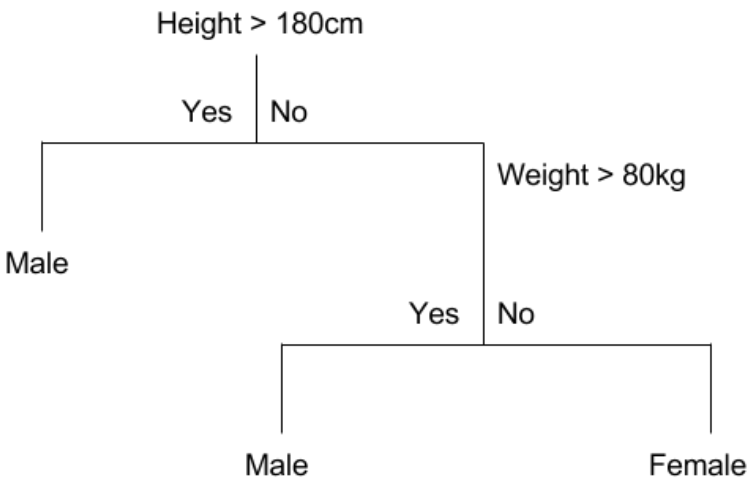


<u>Derinlik-2 Sol Nod:</u>

-(49/54)log(49/54) - 5/54log(5/54) ~= 0.44

Decision Trees - CART

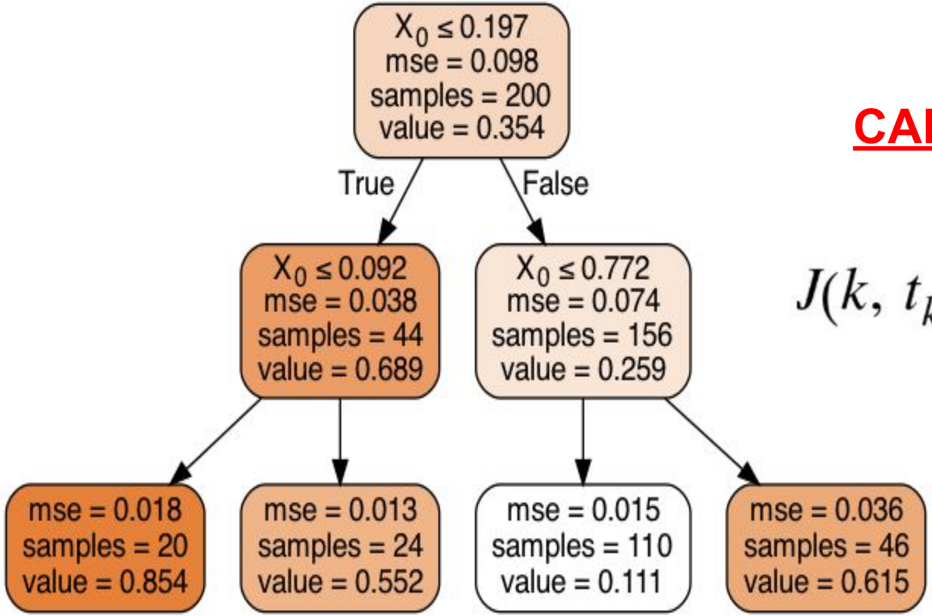
Classification and Regression Tree - CART



CART Cost Function for Classification

$$J(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}$$

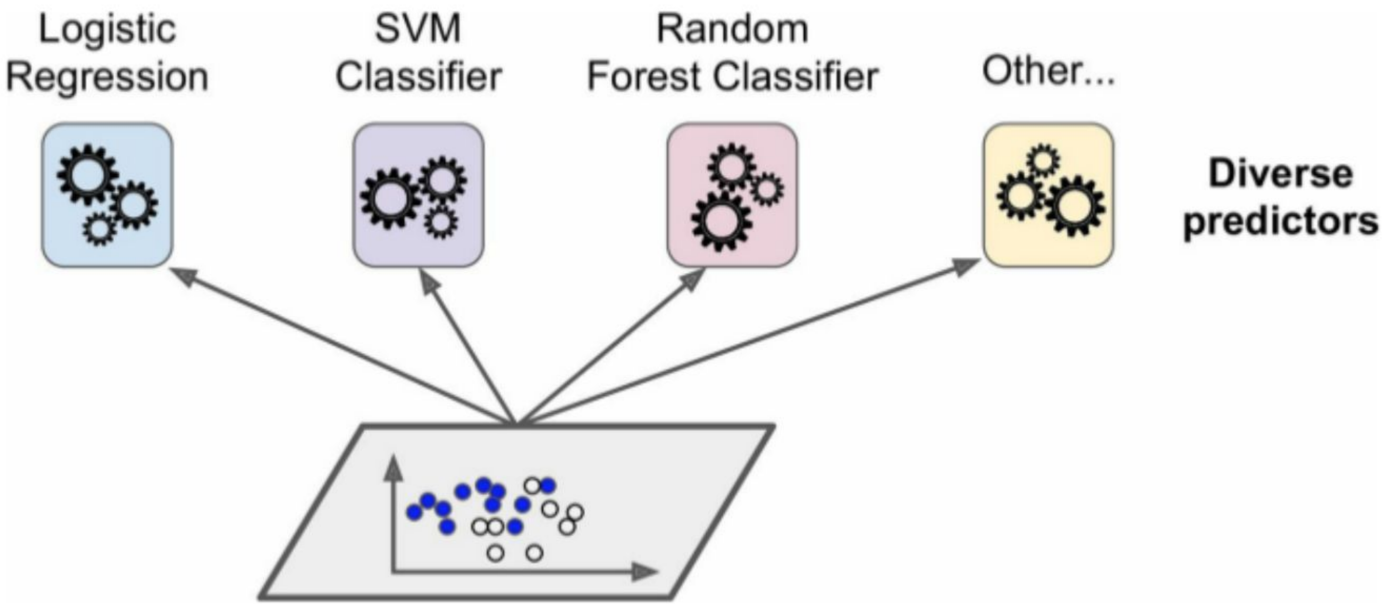
Decision Trees - Regression



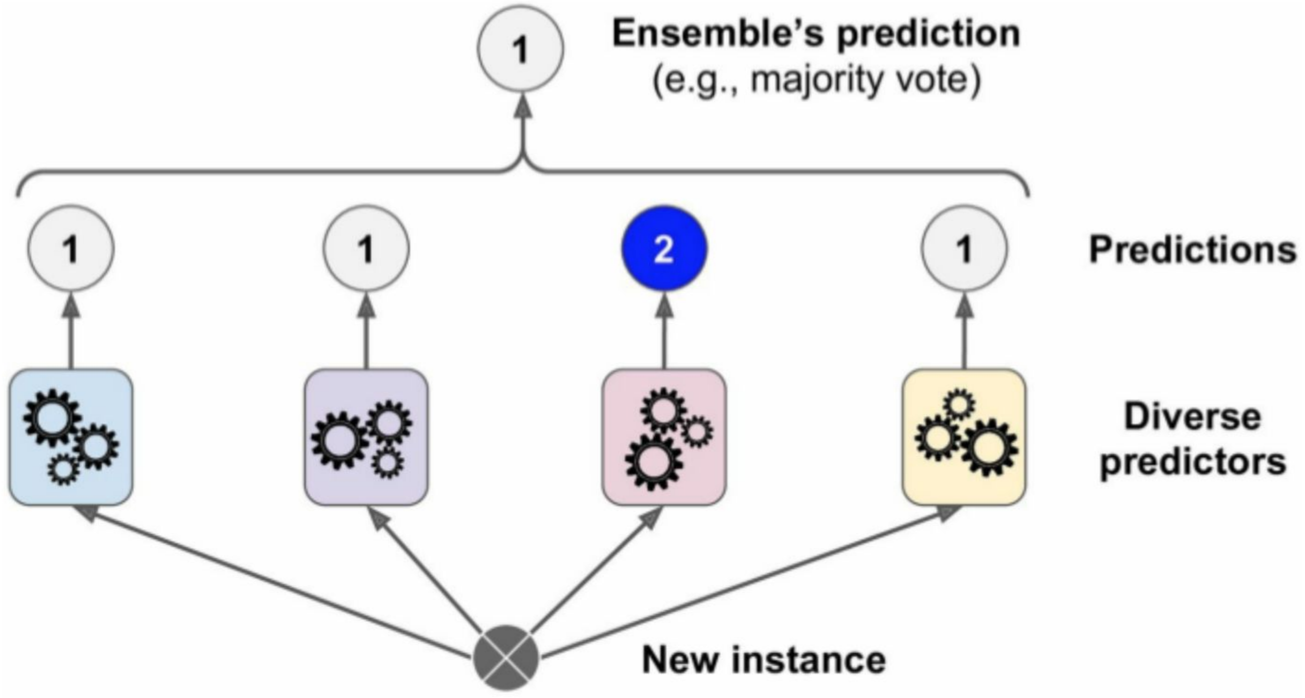
CART Cost Function for Regression

$$J(k, t_k) = \frac{m_{\text{left}}}{m}\text{MSE}_{\text{left}} + \frac{m_{\text{right}}}{m}\text{MSE}_{\text{right}}$$

Random Forest



Random Forest



Random Forest

Bagging and Pasting