



Meftun AKARSU

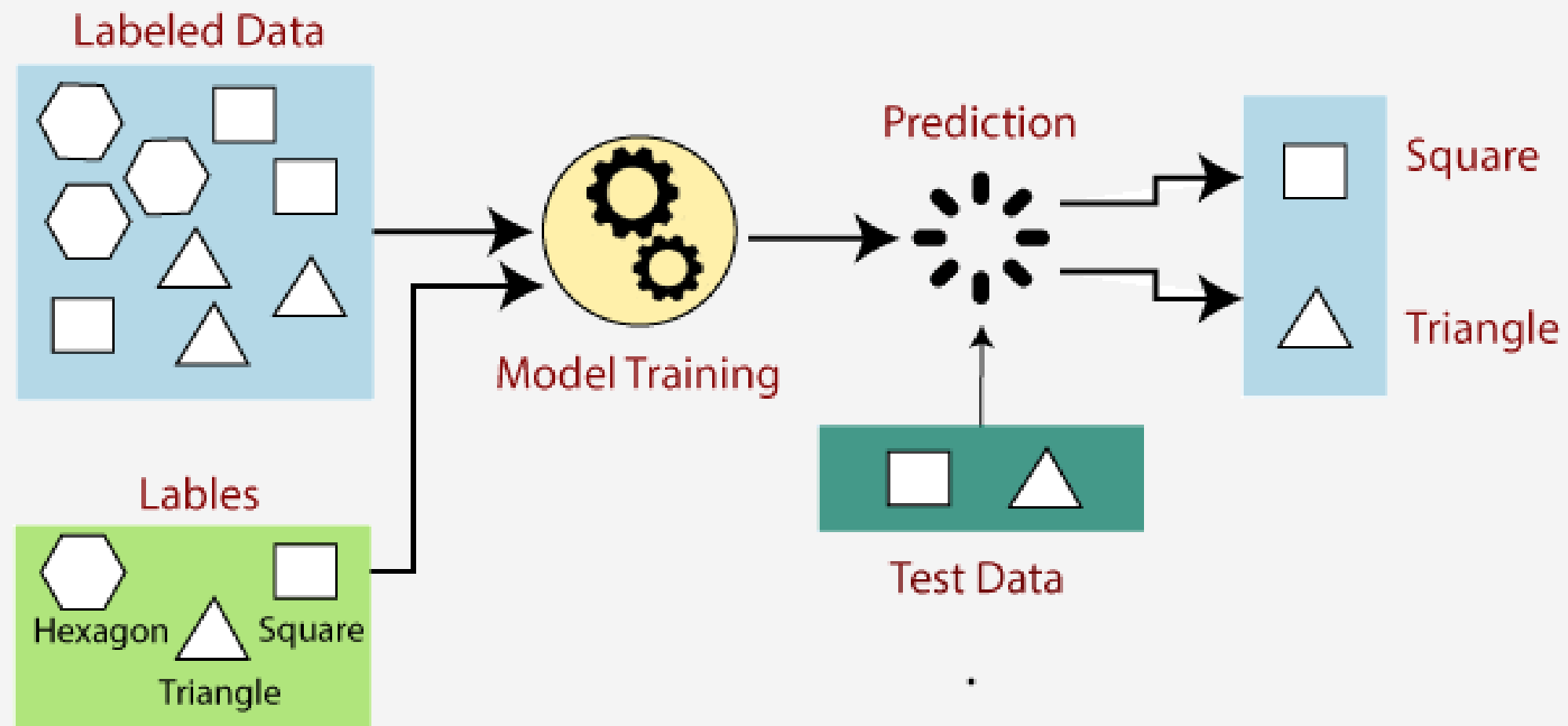
# Machine Learning

Part II

# Content

- Regression
- Classification
- Train Test Split Mantığı
- Supervised Learning Algoritmalar
- Unsupervised Learning Algoritmalar
- Reinforcement Learning Algoritmaları
- Training

# Intro



# Machine Learning

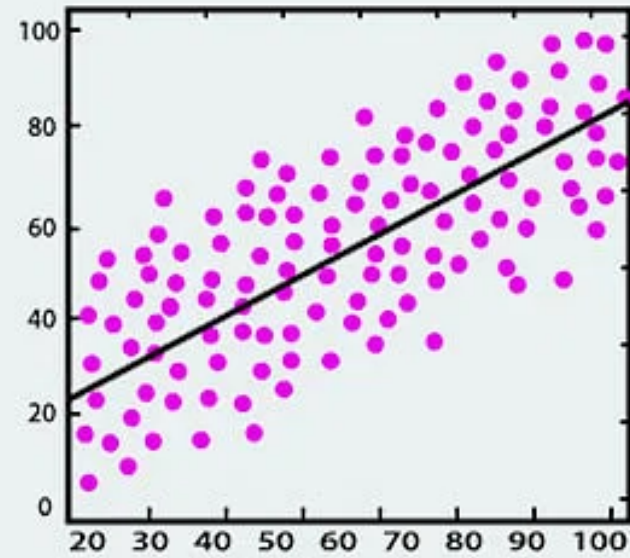
- Eğitim (Training) işlemi, makine öğrenimi modellerinin veri kümesini kullanarak öğrenmeyi ve desenleri tanımayı içeren süreçtir.
- Model, eğitim verilerine dayalı olarak girişlerle çıktılar arasındaki ilişkileri anlamaya çalışır. Bu süreçte model, verilerdeki desenleri algılar ve belirli bir görevi (örneğin, sınıflandırma veya regresyon) gerçekleştirebilmek için gereken parametreleri ve ağırlıkları ayarlar.

# Machine Learning

- <https://www.youtube.com/watch?v=Aut32pR5PQA>
- <https://www.youtube.com/watch?v=aircAruvnKk&t=561s>

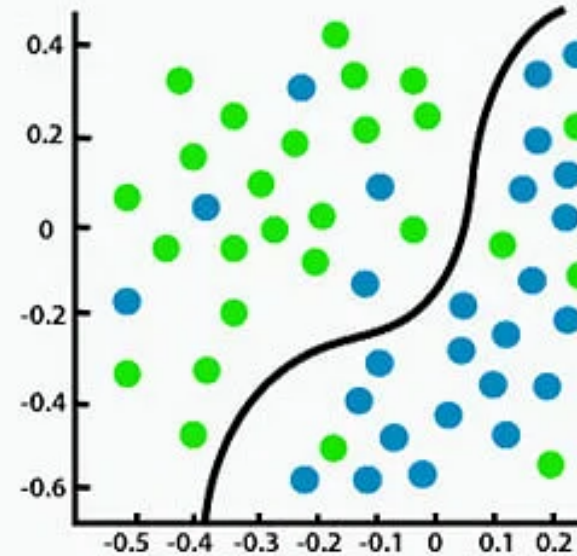
# Machine Learning

## Regression



Regression

versus



Classification

- Regresyon, belirli tarım ürünlerinin verimini tahmin etmek için kullanılabilir. Örneğin, bir çiftçi, buğday ekimi yapacağı bir tarla için geçmiş yıllarda elde ettiği verim verilerine dayanarak, bu yılki beklenen buğday verimini tahmin etmek isteyebilir.

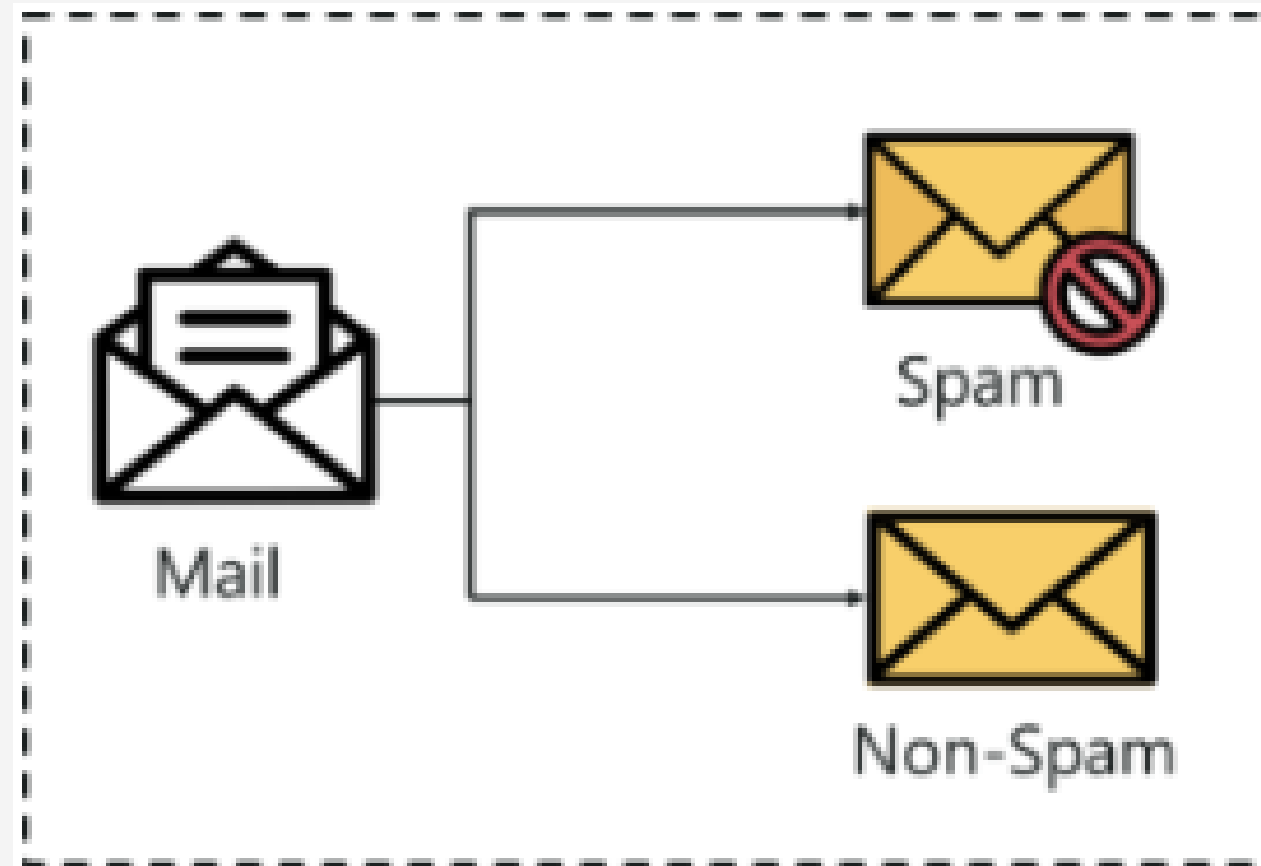
# Machine Learning

## Regression

- Bu senaryoda, bağımlı değişken "buğday verimi" olurken, bağımsız değişkenler "toprak özellikleri" (toprak pH, nem, mineral içeriği vb.), "hava koşulları" (yağış miktarı, sıcaklık vb.), "gübre kullanımı" gibi faktörler olabilir.
- Regresyon modeli, bu bağımsız değişkenlerin veri noktaları üzerinde analiz yaparak, buğday verimini tahmin etmeye yardımcı olur. (sayısal veri gibi düşünülebilir)

# Machine Learning

## Classification



- Tarım alanında sınıflandırma (classification), bitki hastalıklarının teşhisi veya zararlılarla mücadelede kullanılabilir. Örneğin, belirli bir bitki yaprağının üzerindeki lekelerin, belirli bir hastalığın semptomları mı yoksa normal bitki gelişimi sırasında ortaya çıkan olağan bir durum mu olduğunu belirlemek için sınıflandırma kullanılabilir.



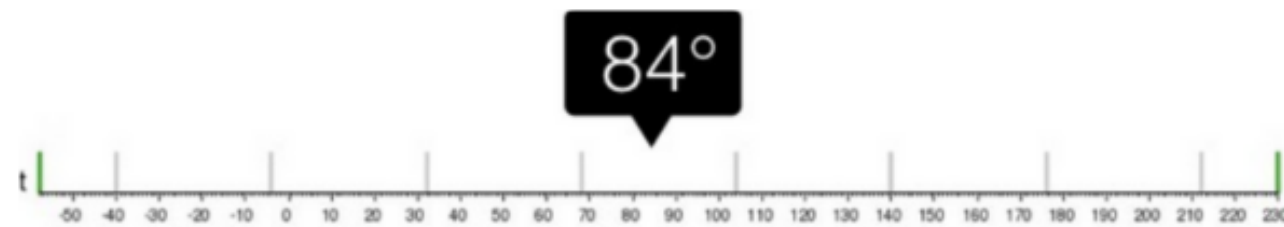
# Machine Learning

## Classification

### Regression



What will be the temperature tomorrow?



Fahrenheit

### Classification



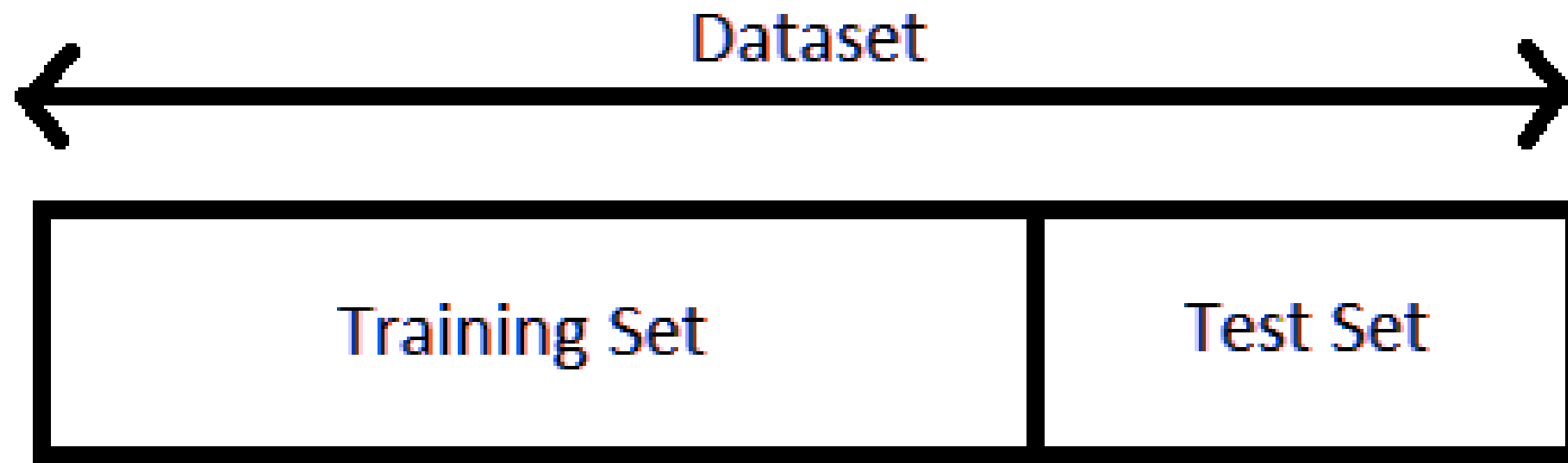
Will it be hot or cold tomorrow?



Fahrenheit

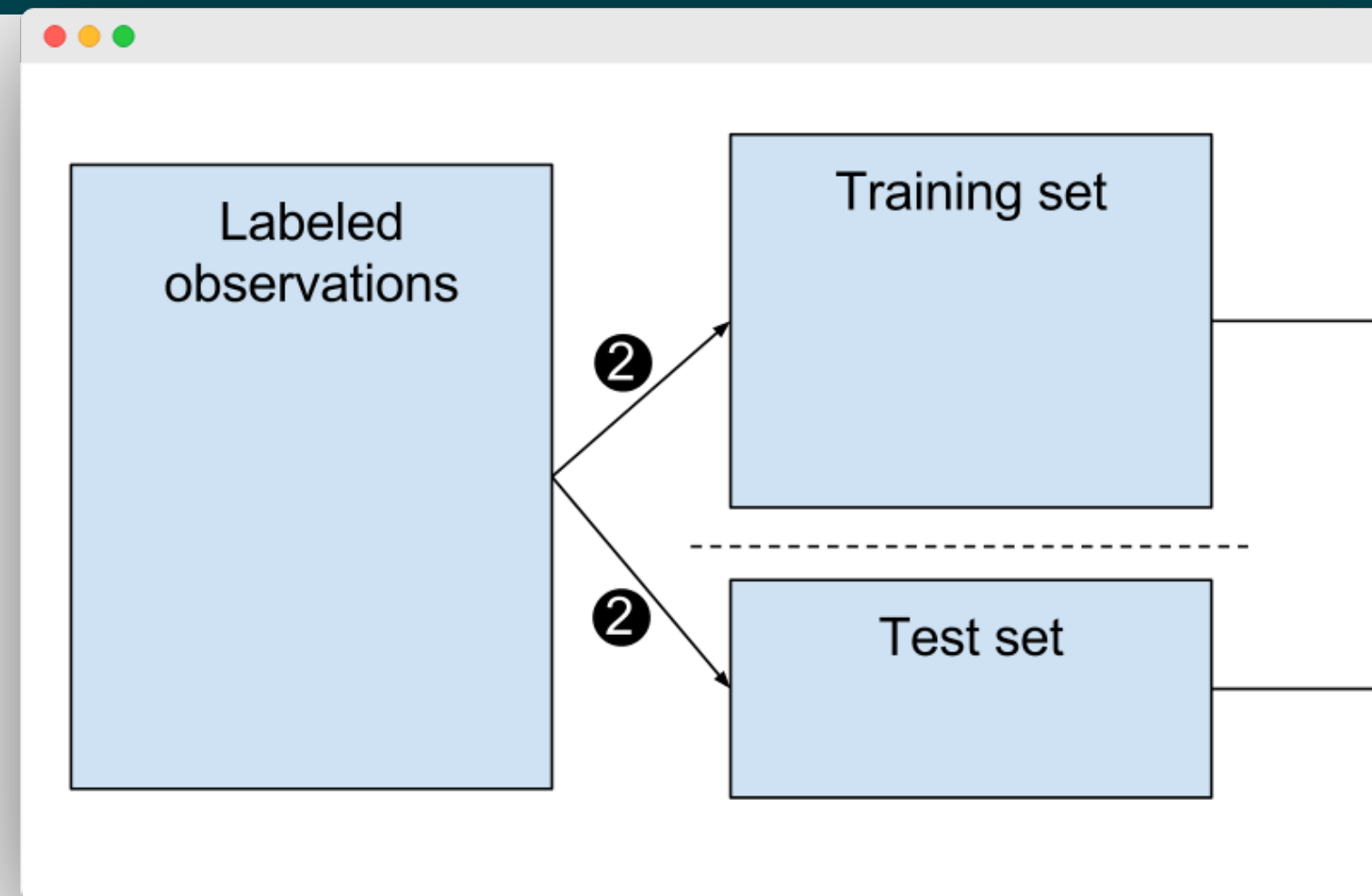
# Machine Learning

## Train Test Split



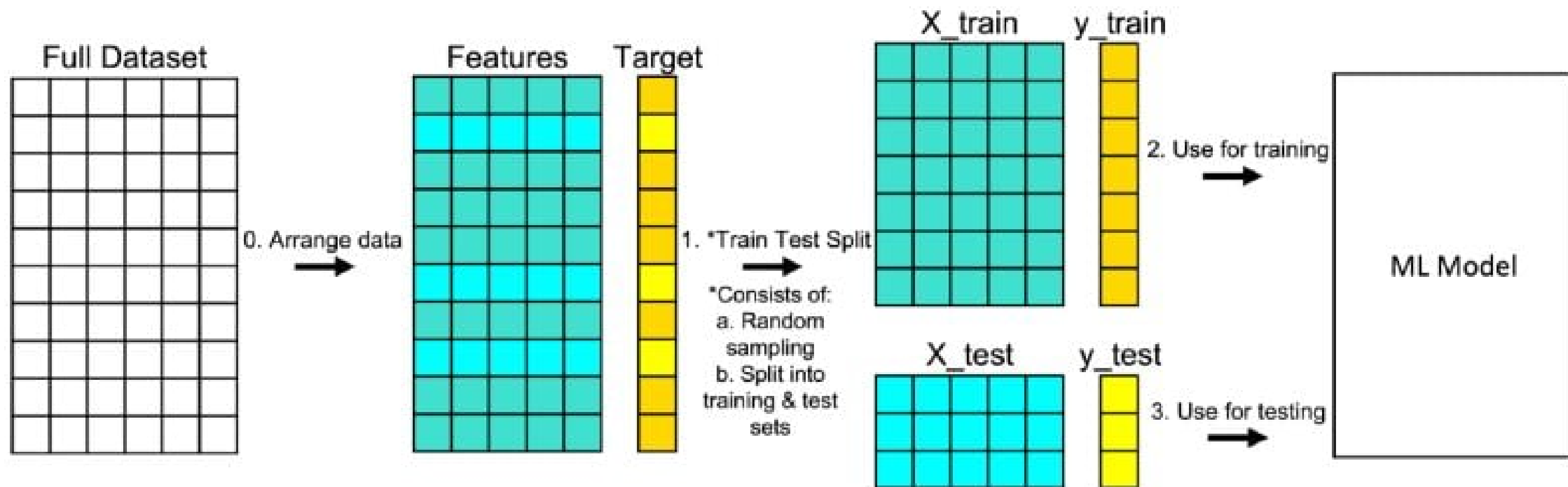
# Machine Learning

## Train Test Split



# Machine Learning

## Train Test Split



# Machine Learning

## Train Test Split

- *Train-test split*, makine öğreniminde modelinizi eğitmek ve test etmek için veri kümesini iki bölüme ayırmanın bir yoludur.
- *Amacımız*, eğittiğimiz modelin gerçek dünya verilerini gördüğünde ne kadar iyi performans göstereceğini tahmin etmektir.

# Machine Learning

## Train Test Split

- Makine öğrenimi modelimizi eğitmek ve test etmek için kullanacağımız veri kümesi elimizde hazır olmalıdır.
- Bu veri kümesi, giriş özelliklerini ve ilgili çıktı etiketlerini içerir.
- Örneğin, bitki yapraklarının hastalıklı mı yoksa sağlıklı mı olduğunu belirlemek istediğimizi düşünelim. Veri kümemiz, yaprakların görüntülerini ve hastalıklı veya sağlıklı olduğu bilgisini içermelidir.

# Machine Learning

## Train Test Split

- Veri kümesini, eğitim (train) ve test (test) verileri olarak iki parçaya ayırmak için train-test split yöntemini kullanacağız.
- Eğitim verileri, modelimizi eğitmek için kullanılacak ve test verileri ise modelin performansını ölçmek için kullanılacaktır.

# Machine Learning

## Train Test Split

- Train-test split işlemi, veri kümesini rastgele iki bölüme ayırarak yapılır.
- Örneğin, veri kümesinin %80'ini eğitim verisi olarak, %20'sini test verisi olarak ayırabiliriz.
- Eğitim verisi, modelin öğrenme sürecinde kullanılacak ve model, bu veri kümesini kullanarak çeşitli özellikleri ve etiketleri ilişkilendirmeyi öğrenecektir.



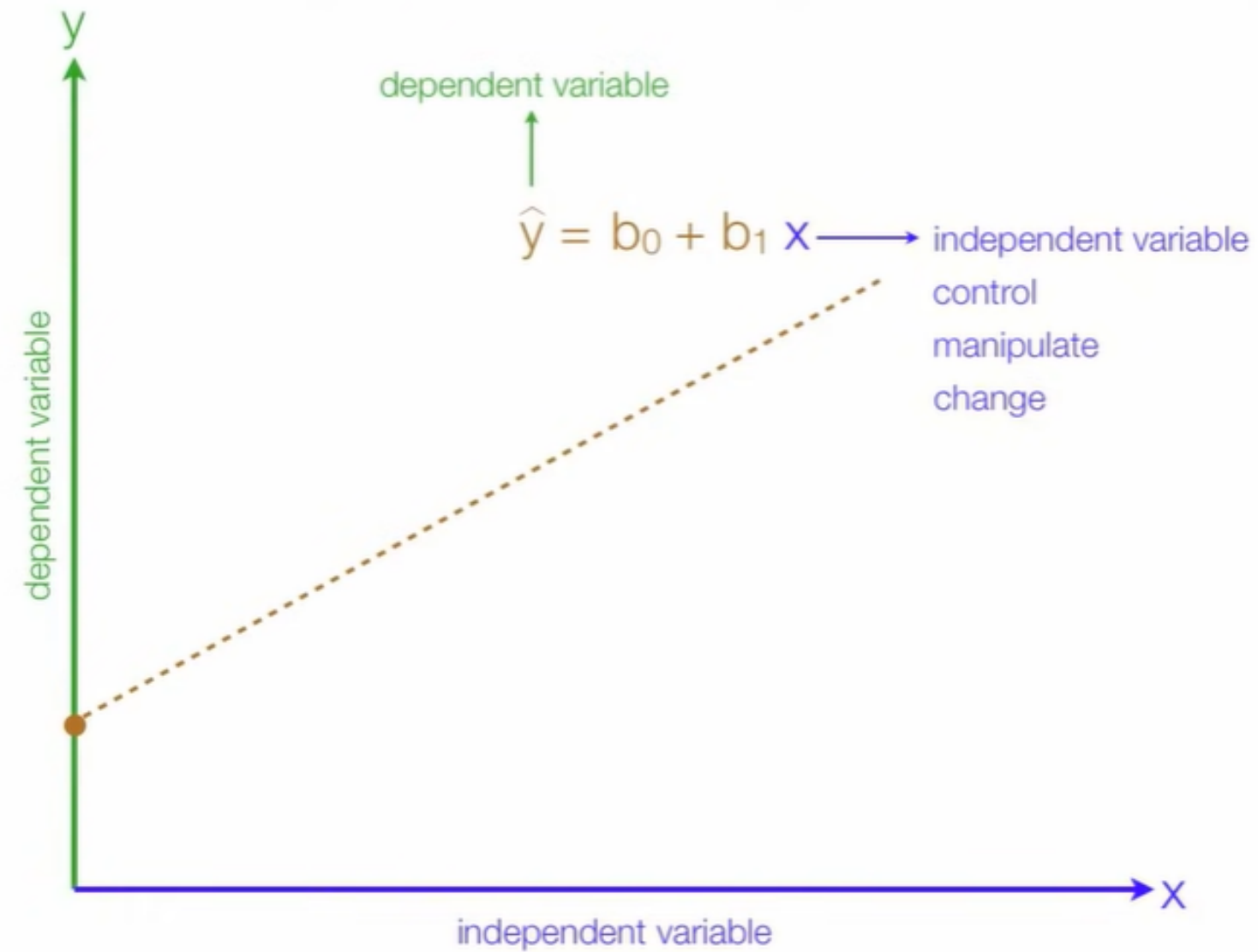
# Machine Learning

## Train Test Split

- Test verisi, modelin performansını değerlendirmek için kullanılacak ve modelin gerçek dünya verilerine ne kadar iyi genelleştirebildiğini göreceğiz.

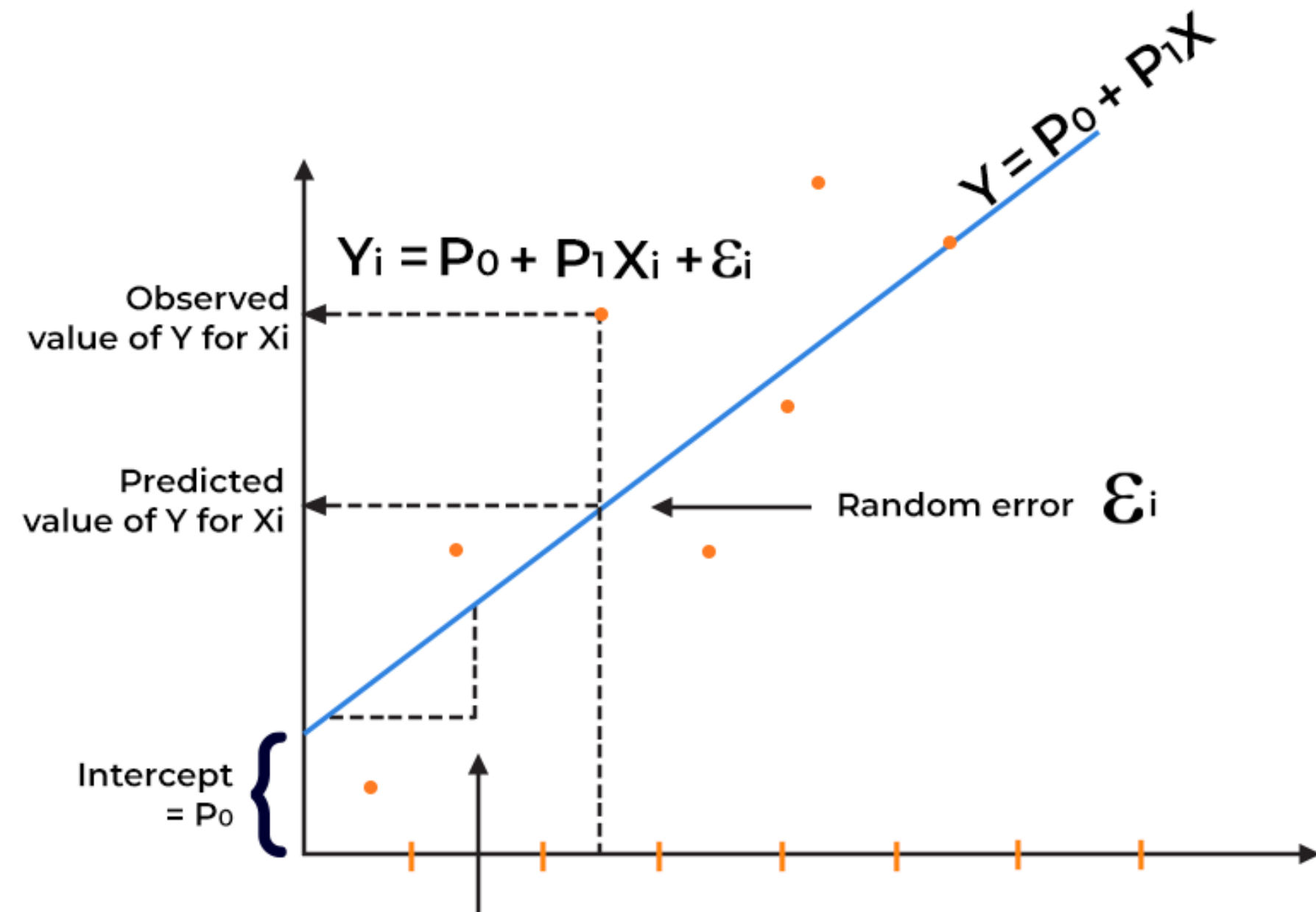
# Supervised Learning

## Linear Regression



# Supervised Learning

## Linear Regression



# Supervised Learning

## Linear Regression

Notasyon:

- $Y$ : Bağımlı değişken (örneğin, bitki verimi)
- $X$ : Bağımsız değişken (örneğin, gübre miktarı)
- $n$ : Veri noktalarının sayısı
- $\hat{Y}$ : Model tarafından yapılan tahmin ( $Y$ 'nin tahmin edilen değeri)
- $\beta_0$ : Kesme terimi (intercept), doğrunun  $Y$  eksenini kestiği nokta
- $w_1$ : Eğim katsayısı (slope), doğrunun eğimini temsil eder

# Supervised Learning

## Linear Regression

Lineer regresyon modeli,  $\hat{Y} = \beta_0 + w_1X$  şeklinde ifade edilir. Bu denklem, bağımlı değişkenin tahmin edilen değerini ( $\hat{Y}$ ) belirtir.  $\beta_0$  ve  $w_1$ , modelin eğitim sürecinde belirlenen parametrelerdir.

# Supervised Learning

## Linear Regression

Tarım alanında, gübre miktarı ve buğday verimi arasındaki ilişkiyi incelemek istediğinizi düşünelim. Elimizdeki verileri kullanarak bir lineer regresyon modeli oluşturmak istiyoruz.

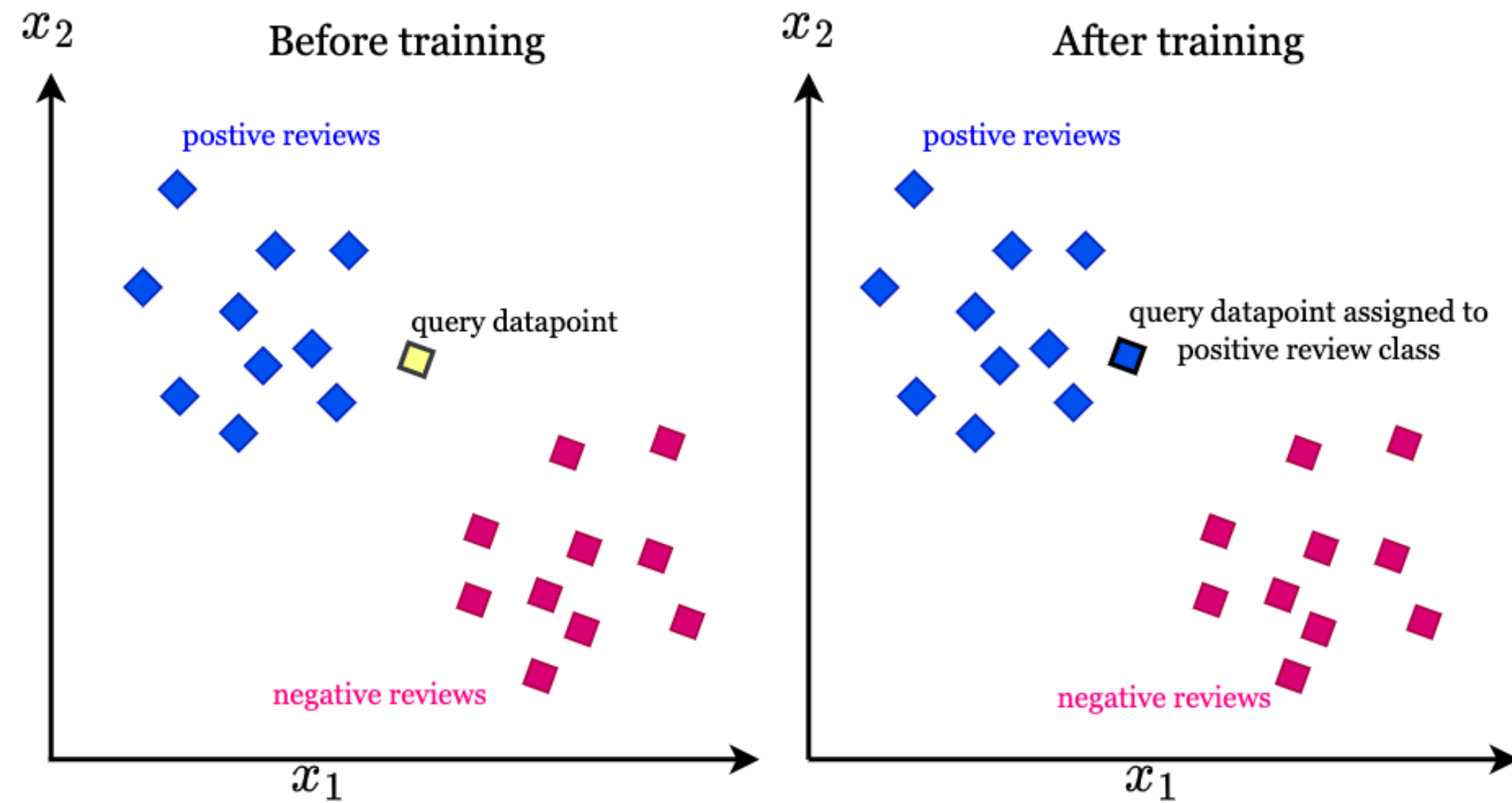
- X: Gübre miktarı (kg)
- Y: Buğday verimi (kg/alan birimi)
- $(X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n)$ : Veri noktaları

Lineer regresyon modeli, aşağıdaki denklemi takip eder:

$$\hat{Y} = \beta_0 + w_1 X$$

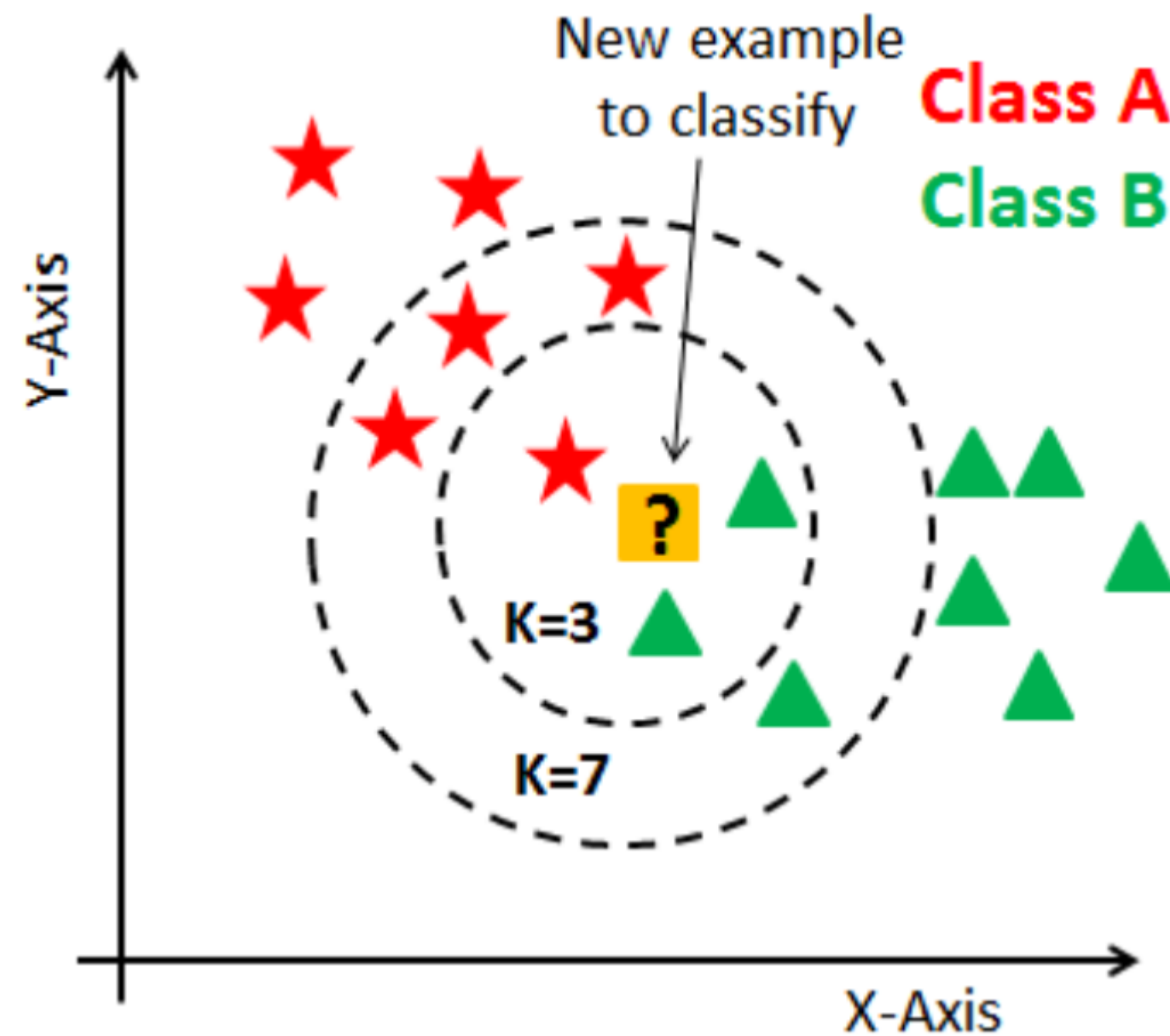
# Supervised Learning

## KNN (K-En Yakın Komşu)



# Supervised Learning

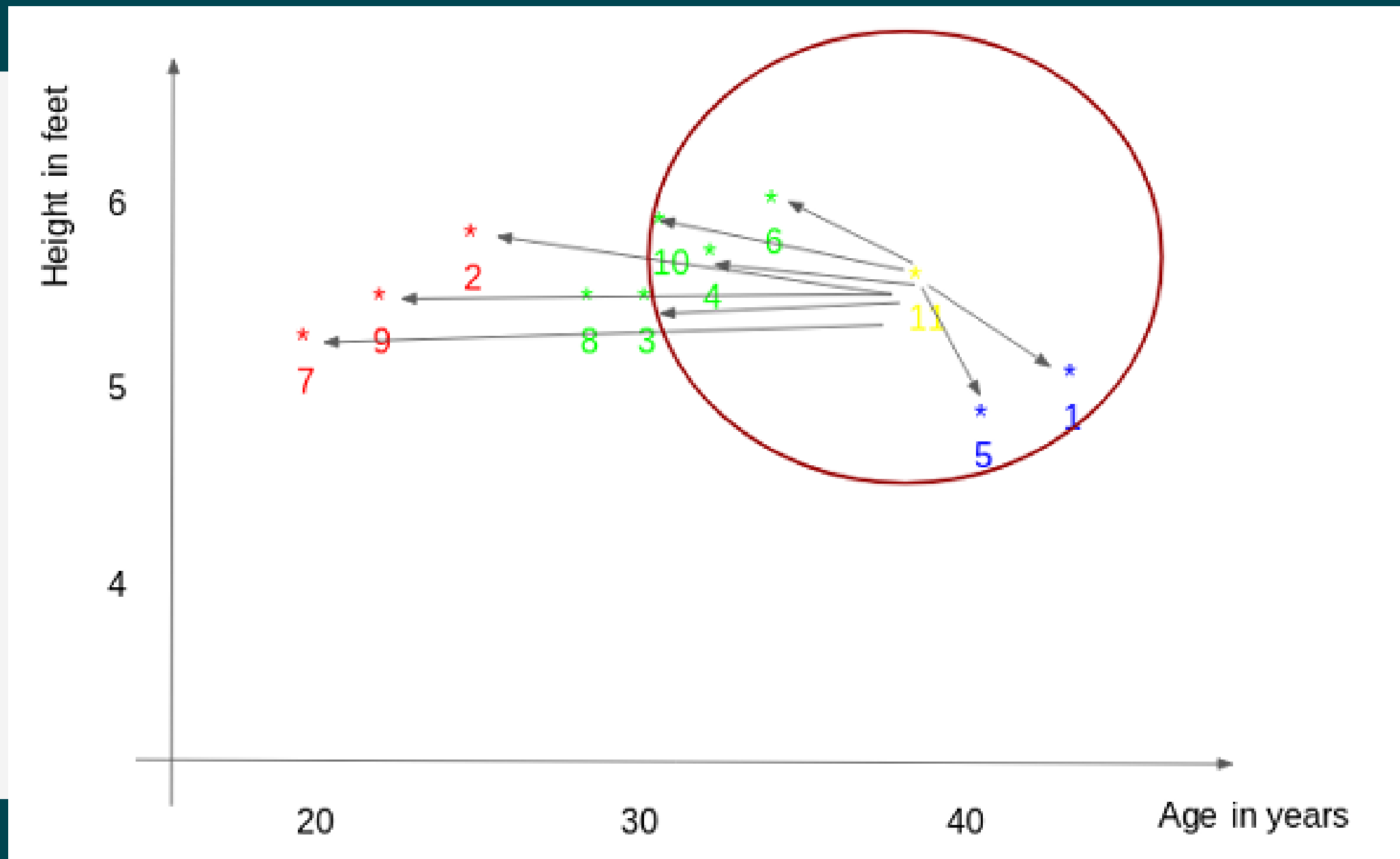
## KNN





# Supervised Learning

## KNN



# Supervised Learning

## KNN

Notasyon:

- $Y$ : Bağımlı değişken (örneğin, bitki verimi)
- $X$ : Bağımsız değişken (örneğin, gübre miktarı)
- $n$ : Veri noktalarının sayısı
- $\hat{Y}$ : Model tarafından yapılan tahmin ( $Y$ 'nin tahmin edilen değeri)
- $K$ : Komşu sayısı, tahmin yaparken kullanılacak yakındaki noktaların sayısı
- $D$ : Veri noktaları arasındaki uzaklığı ölçen bir metrik (genellikle Euclidean mesafesi kullanılır)

# Supervised Learning

## KNN

Notasyon:

- K-En Yakın Komşu regresyonu, veri noktalarının etiketleri (Y değerleri) arasındaki benzerlikleri değerlendirerek tahmin yapar.
- 
- Model, veri kümesindeki her bir veri noktasına dayalı olarak diğer veri noktalarına olan uzaklığını hesaplar ve belirtilen K sayısına göre en yakın K komşusunu seçer.

# Supervised Learning

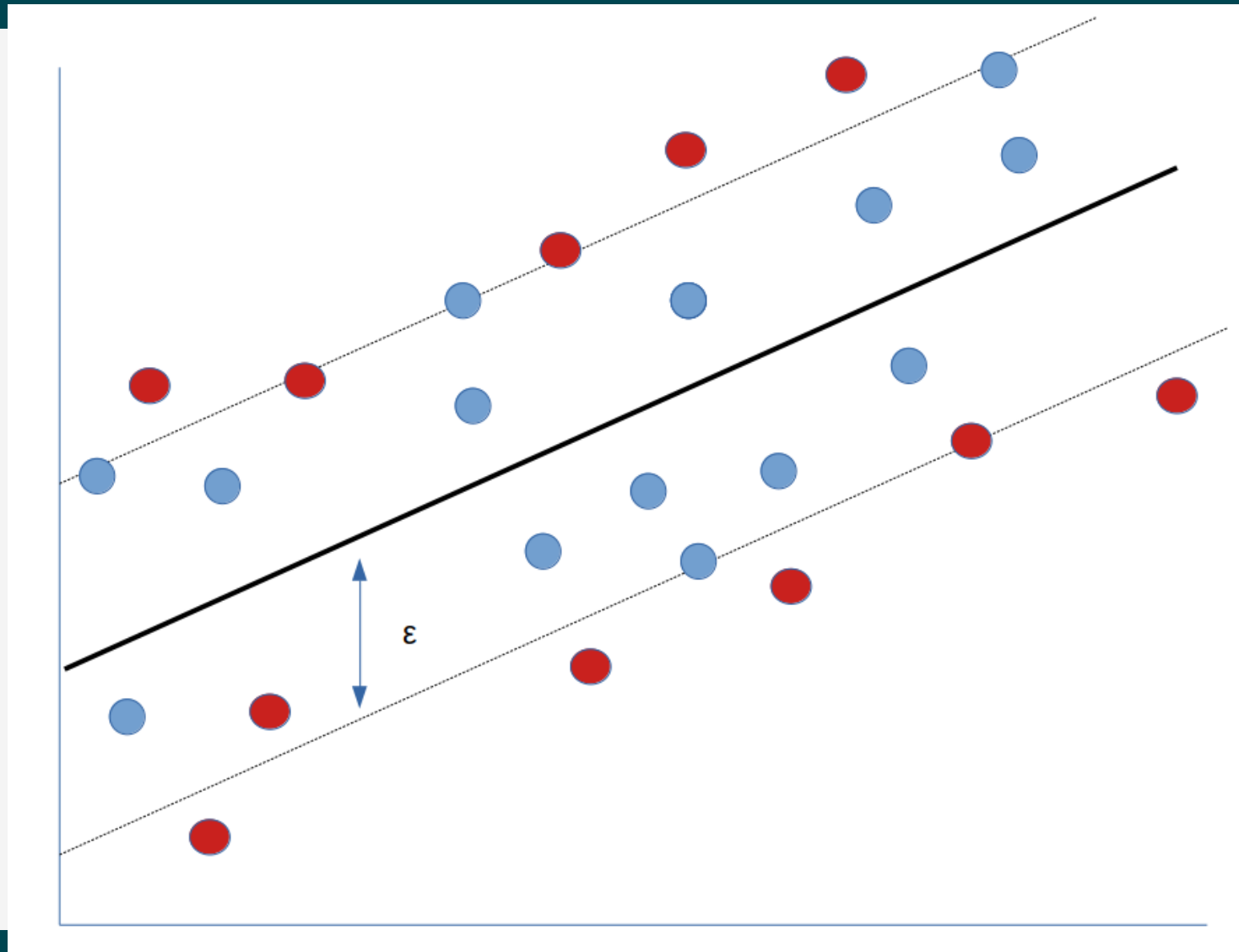
## KNN

Notasyon:

- KNN regresyonu, K komşusunun bağımlı değişken değerlerinin ortalamasını olarak tahmin yapar.
- Yani, K komşusunun Y değerlerini alır ve bu değerlerin ortalamasını olarak tahmin edilen Y değerini bulur.
- Örneğin, K=3 seçilirse, en yakın 3 komşu alınır ve bu komşuların Y değerlerinin ortalaması alınarak tahmin edilen Y değeri hesaplanır.

# Supervised Learning

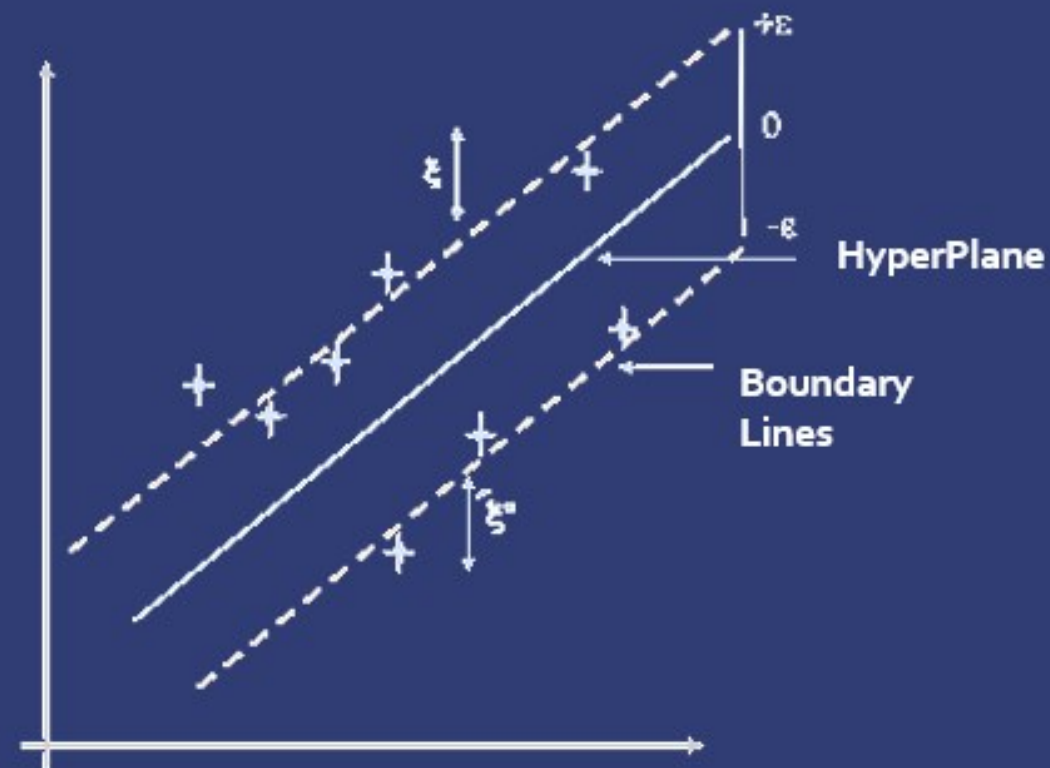
## SVM (Support Vector Machine,Regressor)



# Supervised Learning

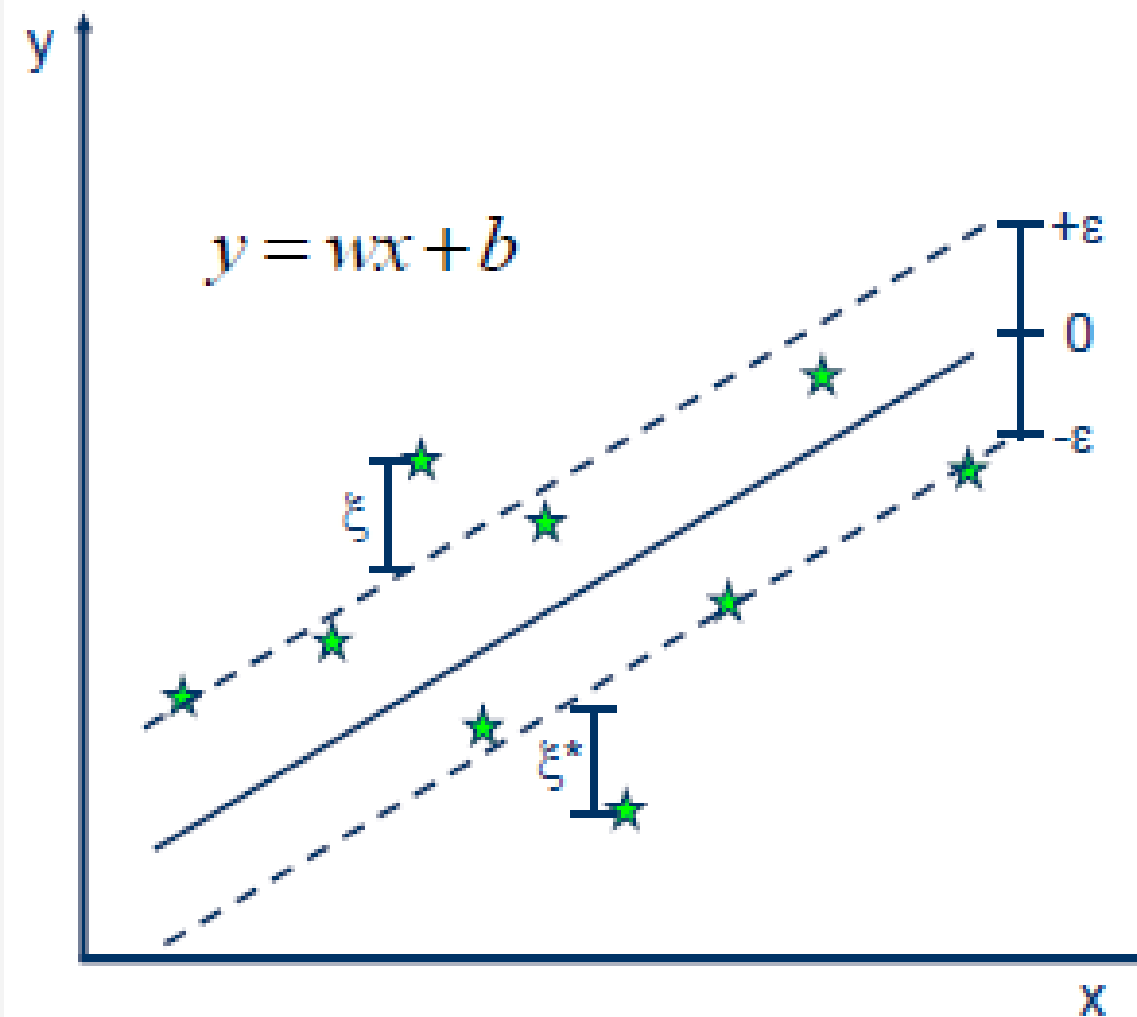
SVM (Support Vector Machine,Regressor)

## Support Vector Regression



# Supervised Learning

## SVM (Support Vector Machine,Regressor)



- Minimize:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

- Constraints:

$$y_i - wx_i - b \leq \epsilon + \xi_i$$

$$wx_i + b - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

# Supervised Learning

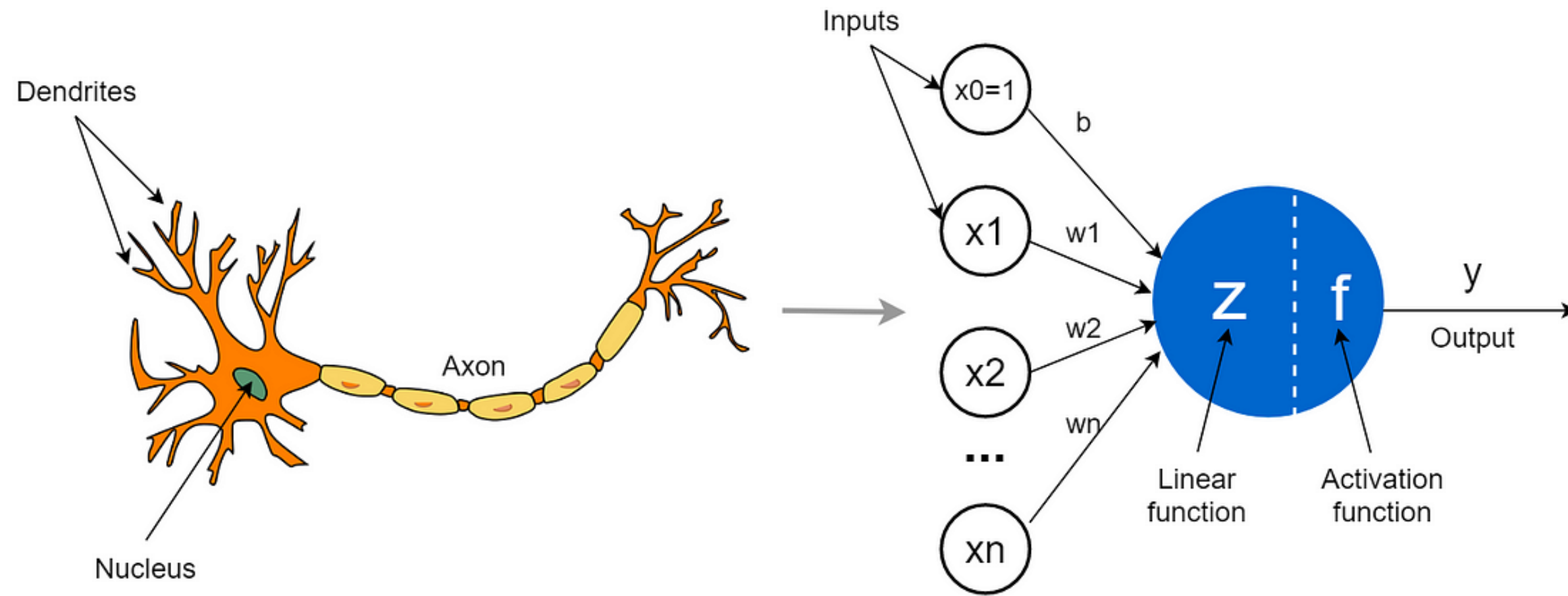
## SVM (Support Vector Machine,Regressor)

- Destek Vektör Regresyonu (Support Vector Regression - SVR), veriler arasındaki doğrusal olmayan bir ilişkiyi modellemek için kullanılan bir regresyon algoritmasıdır. SVR, makine öğrenimi alanında yaygın olarak kullanılan bir metottur ve lineer olmayan regresyon problemlerini çözmek için güçlü bir araçtır.



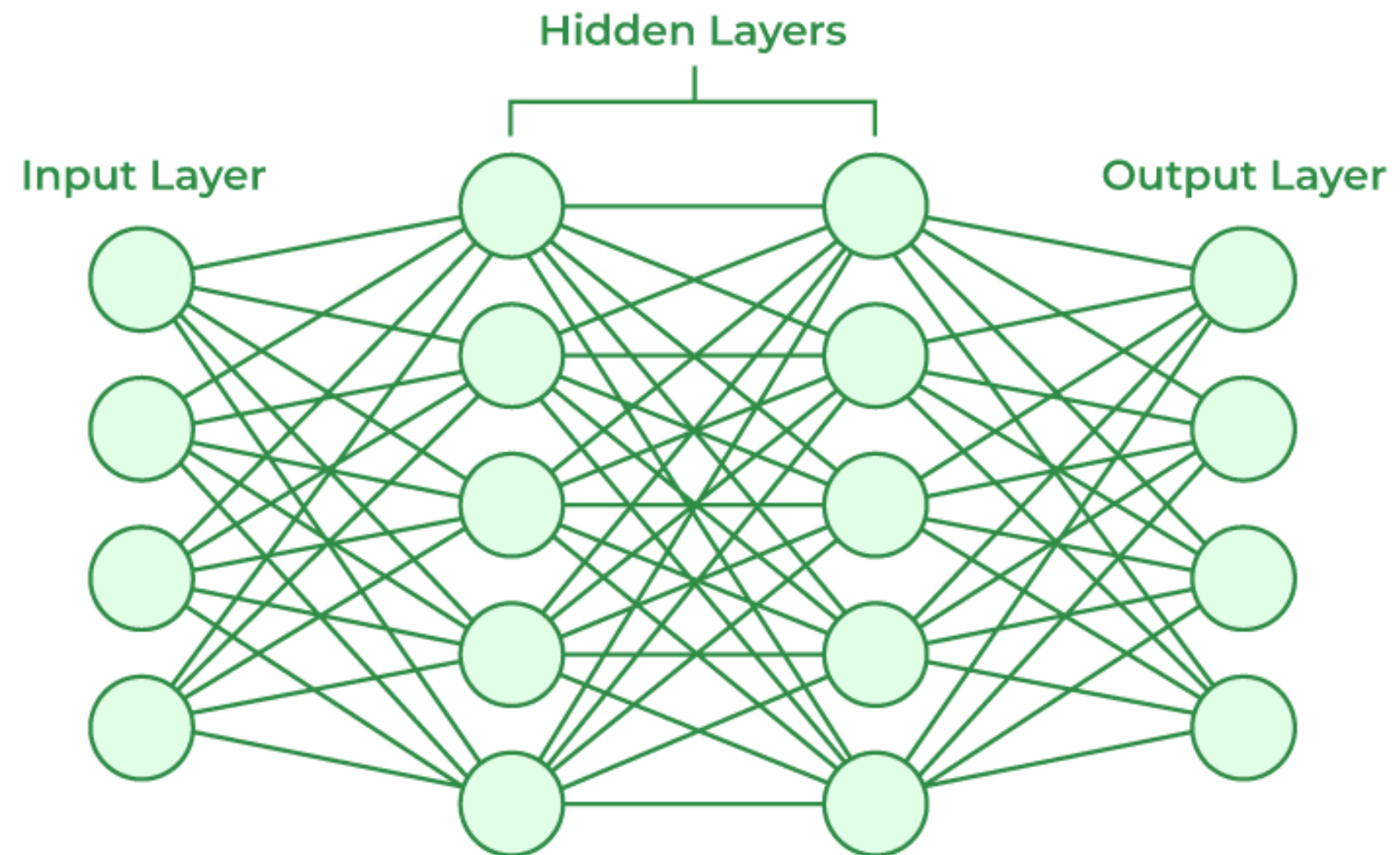
# Supervised Learning

## ANN (Artificial Neural Network)



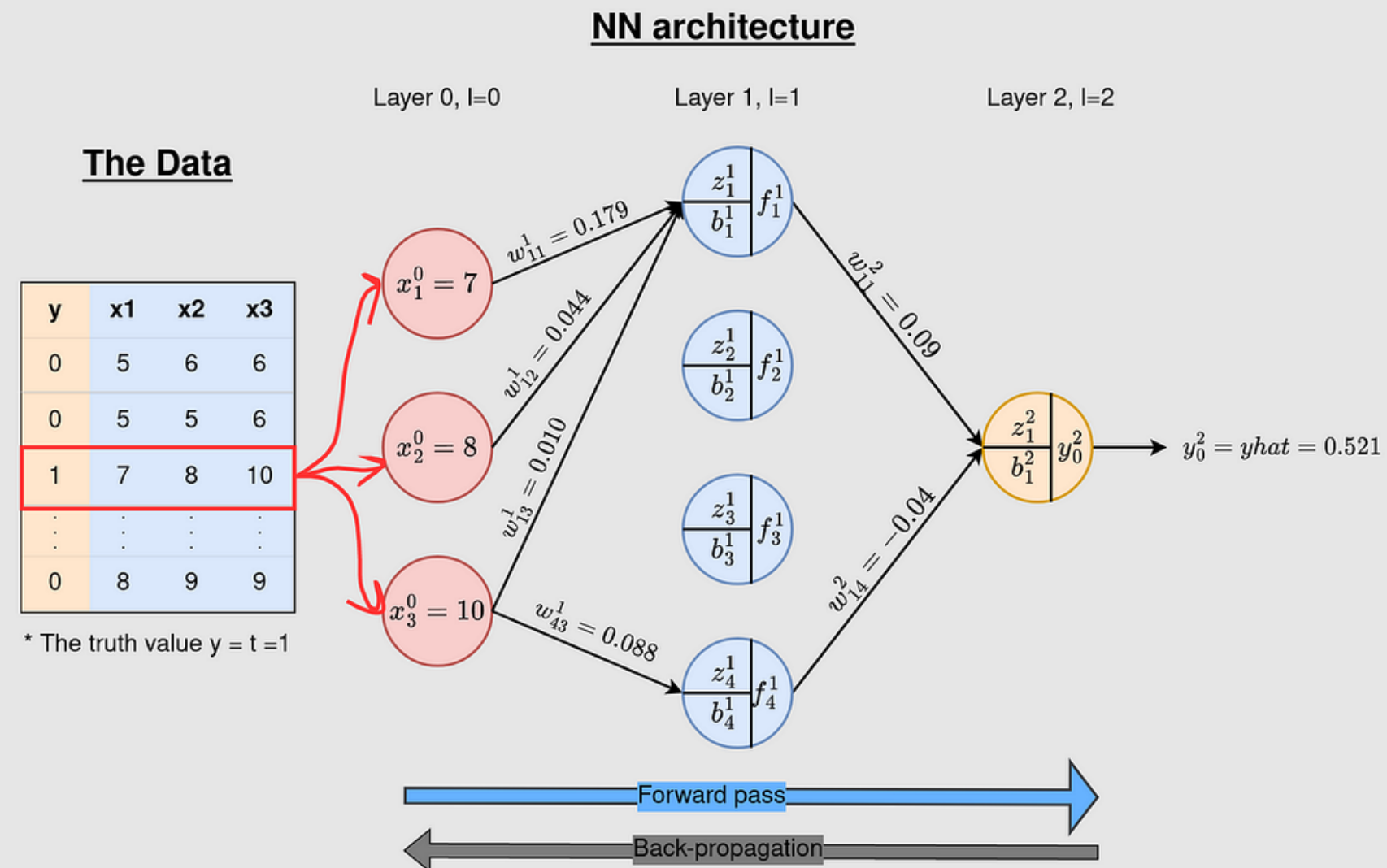
# Supervised Learning

ANN (Artificial Neural Network)



# Supervised Learning

## ANN (Artificial Neural Network)



# Supervised Learning

## ANN (Artificial Neural Network)

- Yapay Sinir Ağları (ANN), biyolojik sinir sistemlerinden ilham alınarak tasarlanmış matematiksel modelledir. Bu modelde, karmaşık bilgi işleme görevlerini gerçekleştirmek ve örüntüleri algılamak için büyük miktarda veri kullanılabilir. ANN, makine öğrenimi ve yapay zeka alanında başarıyla uygulanan en güçlü araçlardan biridir.

# Supervised Learning

## ANN (Artificial Neural Network)

- ANN, katmanlar halinde düzenlenmiştir. Genellikle üç temel katmanı vardır: Giriş katmanı, gizli katman(lar) ve çıkış katmanı.
- Giriş katmanı, verileri ANN'ye girdi olarak alır.
- Gizli katmanlar, girdileri işleyerek verileri daha yüksek seviyede temsil etmek için kullanılır.
- Çıkış katmanı, sonuçları verir ve modelin tahminlerini yapar.

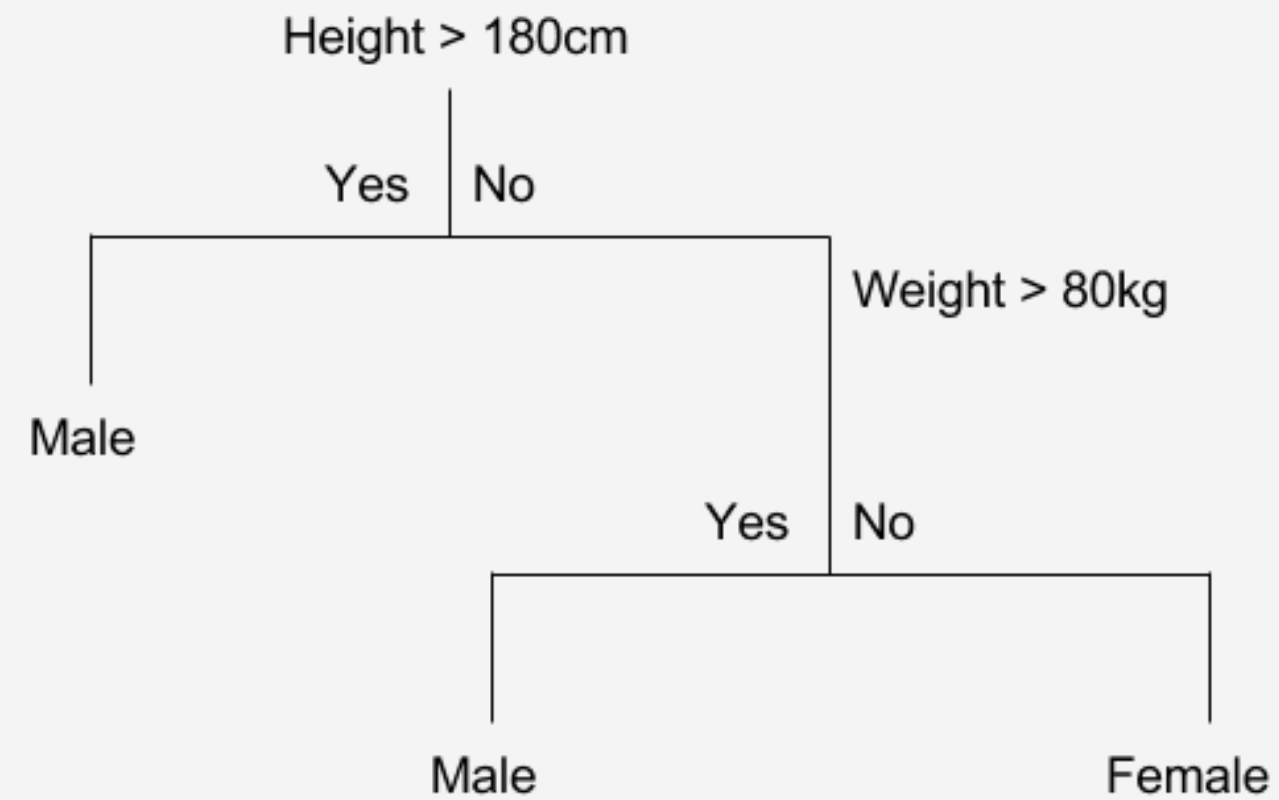
# Supervised Learning

## ANN (Artificial Neural Network)

- Başlangıçta, ağırlıklar ve bias (eşik) değerleri rastgele belirlenir. Bu değerler, nöronların girdi verilerini nasıl işleyeceğini ve aktive edileceğini belirler.
- Veri, giriş katmanına girer ve ardından gizli katmanlardan geçer.
- Her gizli katmandaki nöronlar, girdileri belirli bir işlemle işleyerek bir aktivasyon fonksiyonuna gönderir.
- Aktivasyon fonksiyonu, nöronun çıkışını belirler ve çıkış katmanına iletilir.
- Çıkış katmanı, sonuçları sağlar ve modelin tahminlerini verir.

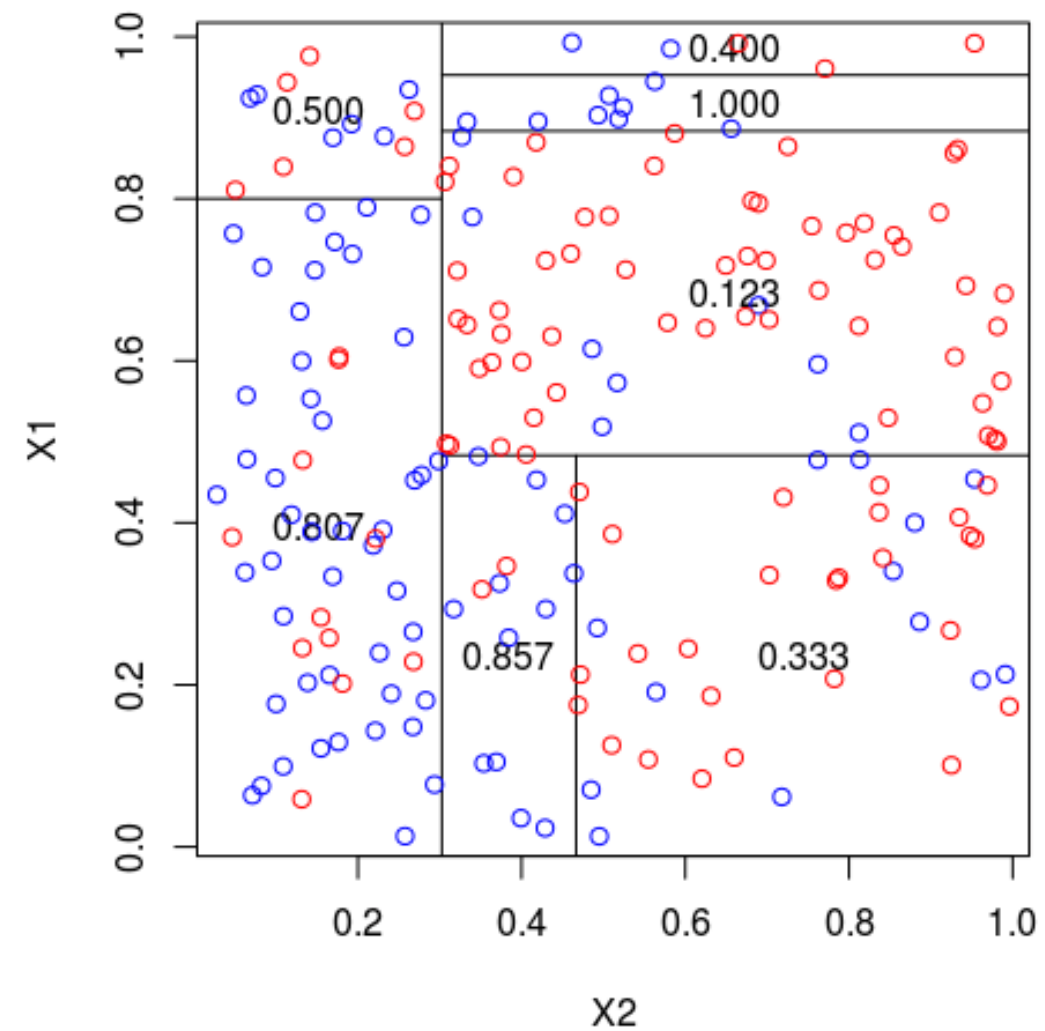
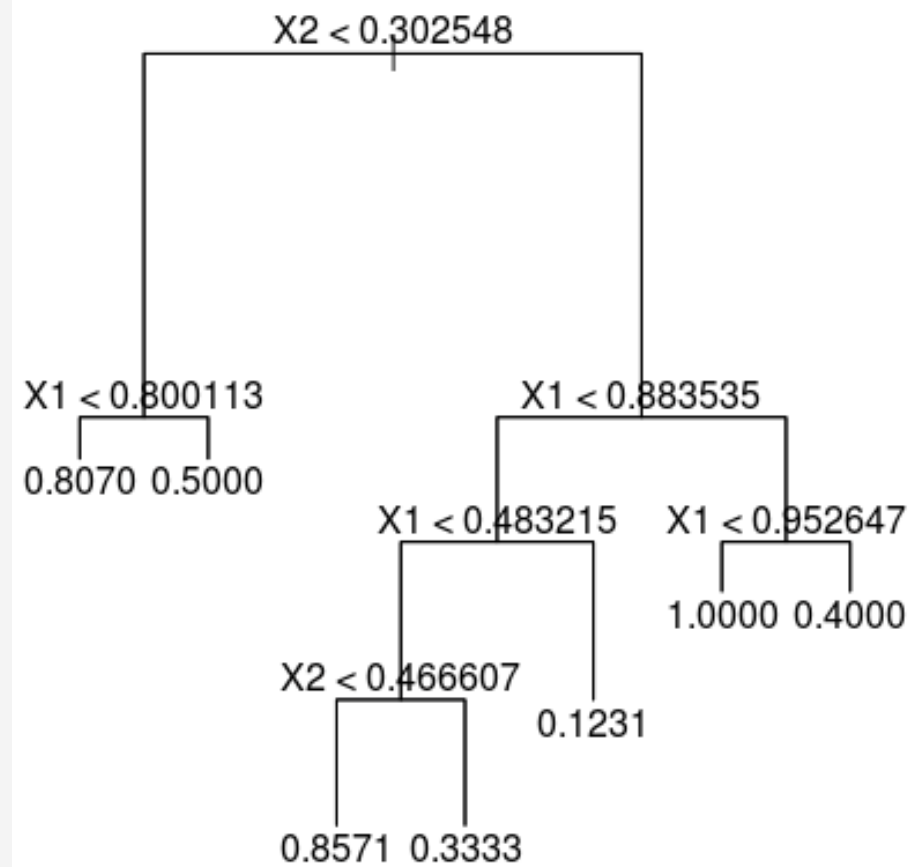
# Supervised Learning

## CART (Classification and Regression Tree)



# Supervised Learning

## CART (Classification and Regression Tree)





# Supervised Learning

## CART (Classification and Regression Tree)

- CART, hem sınıflandırma hem de regresyon problemlerini çözmek için kullanılan bir karar ağacı algoritmasıdır. Bu algoritma, veri kümesini adım adım böler ve her bölme adımında en iyi ayrımı yapacak özellikleri ve değerleri seçerek bir ağaç yapısı oluşturur.

# Supervised Learning

## CART (Classification and Regression Tree)

- Başlangıçta, veri kümesi kök düğüm olarak kabul edilir.
- Veri kümesi, belirli bir özellik ve eşiği kullanarak iki veya daha fazla alt küme (bölme) oluşturacak şekilde bölünür. Bu bölme, verileri daha homojen alt gruplara ayırmayı hedefler.
- Bu işlem, her alt kümedeki veriler neredeyse aynı sınıfa (sınıflandırma) veya aynı hedef değişkenine (regresyon) sahip olacak şekilde belirli bir duruma ulaşana kadar tekrarlanır.

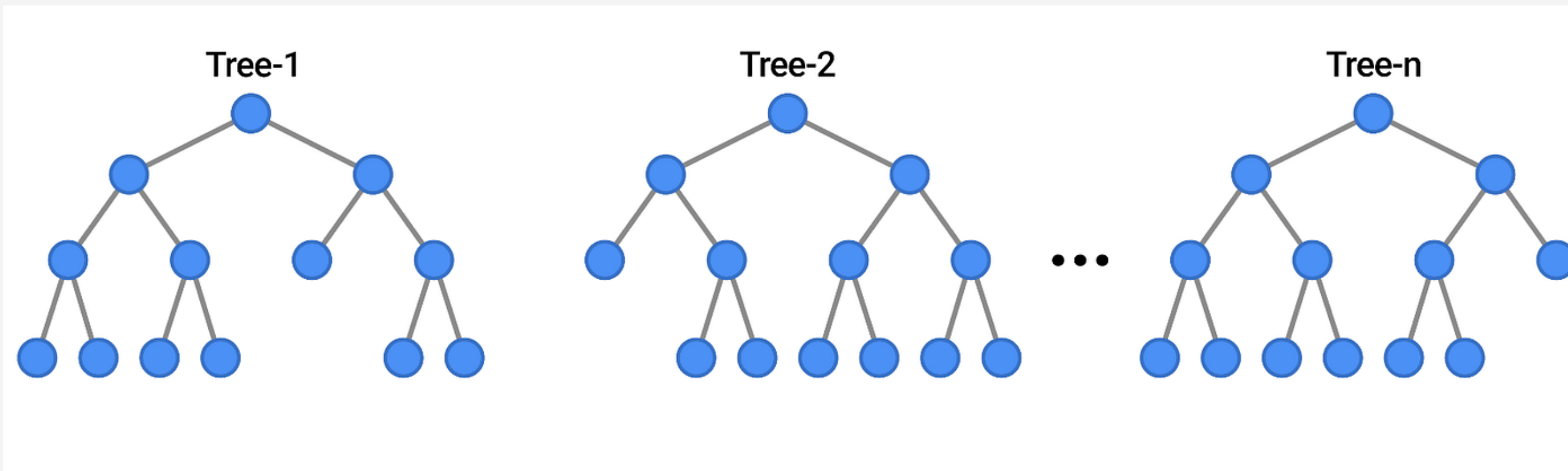
# Supervised Learning

## CART (Classification and Regression Tree)

- Bölmelerin oluşturulması, ağacın bir düğümü (node) ve alt düğümleri (branch) oluşturur.
- Ağacın yaprak düğümleri (leaf node) ise sonuçları belirleyen son düğümlerdir ve sınıf tahminlerini (sınıflandırma) veya hedef değişken tahminlerini (regresyon) içerir.

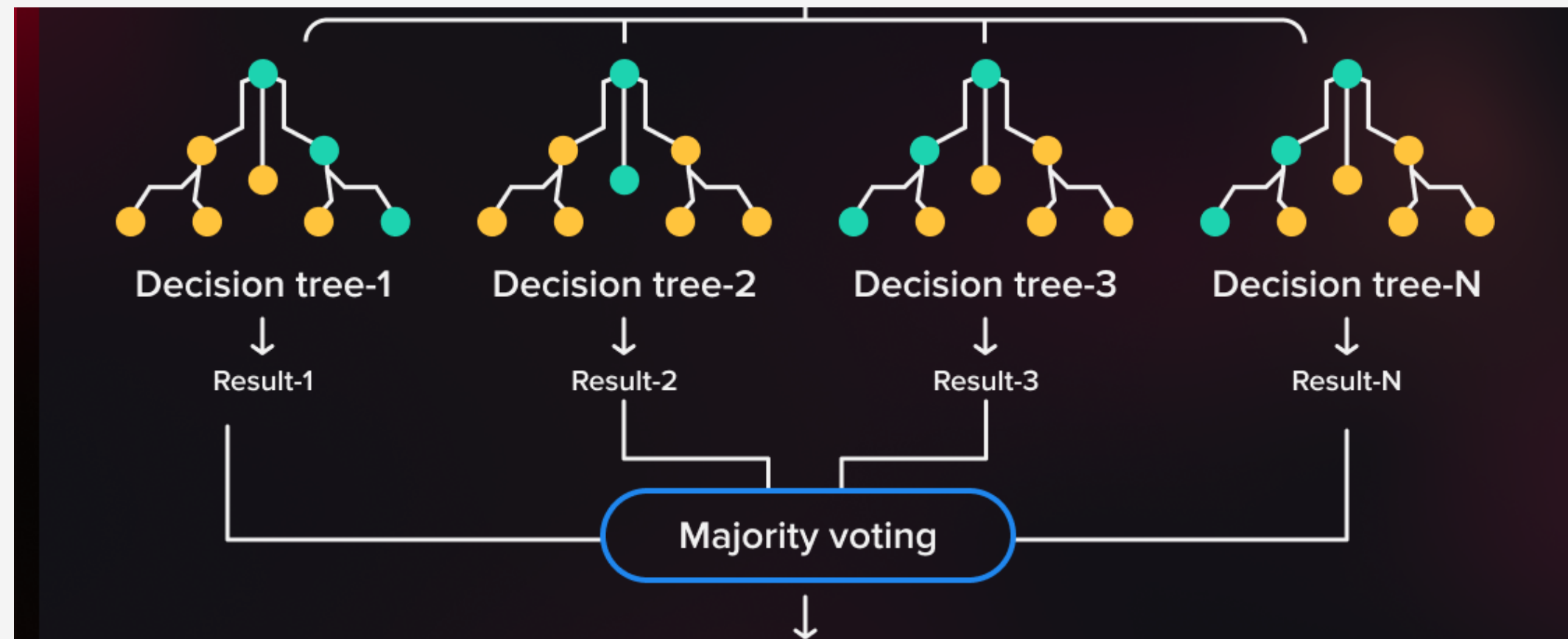
# Supervised Learning

## Random Forest



# Supervised Learning

## Random Forest



# Supervised Learning

LightGMB, XGBoost

- Random Forest, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan güçlü bir makine öğrenimi algoritmasıdır. Bu algoritma, birden fazla karar ağacını (decision tree) bir araya getirerek, daha doğru ve kararlı tahminler elde etmeyi amaçlar.