

predicts discrete categories or classes, assigning input data to predefined groups such as spam/not spam or specific object types.

Mahcine Learning Notes

Ahmed Yasser

2025

Contents

Some Important Definitions	3
Supervised Learning	3

Some Important Definitions

- θ = Parameters of the model
- X = Inputs or features
- y = Output or target variable
- $h(x)$ = Hypothesis function
- m = Number of training examples (Number of rows in the training set)
- n = Number of features
- $(X^{(i)}, y^{(i)})$ = The i -th training example

Supervised Learning

Supervised learning is a type of machine learning where the model is trained on labeled data. The goal is to learn a mapping from inputs to outputs based on the provided labels. The model can then make predictions on new, unseen data.

Supervised Learning Types

Supervised learning can be divided into two main types:

1. **Classification:** Predicts discrete categories or classes, assigning input data to predefined groups such as spam/not spam or specific object types.
2. **Regression:** Predicts continuous numerical values or quantities within a range, such as prices, temperatures, or heights.

here are some common supervised learning algorithms used for regression tasks:

- Linear Regression
- Polynomial regression
- Lasso Regression (L1 Regularization)
- Ridge Regression (L2 Regularization)
- Decision Trees
- Random Forests
- Support Vector Regression (SVR)

Linear Regression

Linear regression is a supervised learning algorithm used for regression tasks. It models the relationship between input features and a continuous output variable by fitting a linear equation to the observed data. The goal is to find the best-fitting line (or hyperplane in higher dimensions) that minimizes the difference between the predicted values and the actual values.

Linear Regression Equation

The hypothesis function for linear regression is represented as:

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \sum_{j=0}^n \theta_j x_j$$

Where:

- $h(x)$ is the predicted output (hypothesis function)
- θ_0 is the bias
- $\theta_1, \theta_2, \dots, \theta_n$ are the weights (coefficients) for each feature
- x_1, x_2, \dots, x_n are the input features

The equation can also be expressed in vector form as:

$$h(x) = \theta^T X$$

Where:

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_n \end{pmatrix} \text{ is the vector of parameters (weights) and } X = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \text{ is the vector of input features.}$$

Cost Function

The cost function for linear regression is used to measure how well the model fits the training data. It calculates the difference between the predicted values and the actual values. The most common cost function used in linear regression is the Mean Squared Error (MSE):

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2$$

Where:

- $J(\theta)$ is the cost Function
- m is the number of training examples
- $h(x^{(i)})$ is the predicted value for the i -th training examples
- $y^{(i)}$ is the actual value for the i -th training examples
- θ is the vector of parameters (weights)
- $x^{(i)}$ is the input features for the i -th training examples