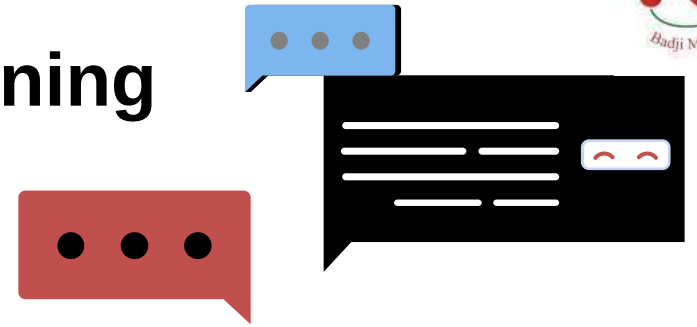
Chapter 4 : Supervised learning



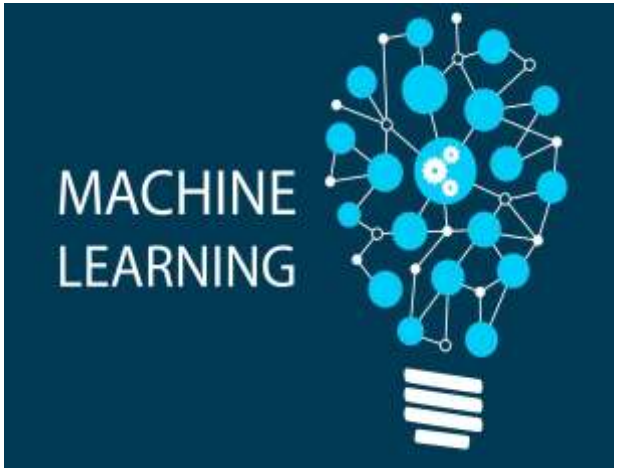


Table of contents

1

Supervised learning?

Supervised learning:

- **Definition:** Supervised learning is a type of machine learning where the model is trained on labeled data, meaning the input data is paired with the correct output.

- **Objective:** The goal is to learn a mapping function from inputs to outputs, which can then be used to predict outcomes for new, unseen data.

1

How to formulate a supervised learning problem ?

Supervised learning problem:

- Supervised learning problems are formulated using a labeled dataset, which consists of input-output pairs. The inputs are often referred to as features, and the outputs are called labels or targets. The process involves:
- Defining the Problem: Determine what you want to predict (e.g., classifying emails as spam or not spam, predicting house prices, etc.).

Supervised learning problem:

- Key Components:
- Input Features: The variables used to make predictions.
- Output Labels: The target variable to be predicted.
- Training Data: A dataset with known input-output pairs used to train the model.
- Test Data: A separate dataset used to evaluate the model's performance.

Steps to Supervised learning:

1. Define the problem (e.g., classification or regression).

Collect and preprocess data.

2. Choose a model (e.g., linear regression, decision trees, etc.).

3. Train the model using the training dataset;

4. Evaluate the model using the test dataset.

5. Deploy the model for inference on new data.

Steps to Supervised learning:

- Example Problems:

1. Predicting house prices (regression).
2. Classifying emails as spam or not spam (classification).

3

Algorithms of supervised learning?

Regression Algorithms:

Regression algorithms are used to predict continuous numerical values. They model the relationship between input features and a continuous target variable.

1. Linear Regression :

Models the relationship between input features and a target variable using a linear equation (e.g., y=mx+by).

Regression Algorithms:

Key Concepts:

Input Features (X): Independent variables used to make predictions.

Target Variable (Y): The continuous value to be predicted.

Model: A linear equation that maps input features to the target variable:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon$$

β_0: Intercept (bias term).

Regression Algorithms:

$\beta_0, \beta_1, \ldots, \beta_n$: Coefficients (weights) representing the change in Y per unit change in Xi.

ϵ: Random error term (unexplained variability).

Use case:

- Predicting house prices based on features like size, location, and number of bedrooms
- Estimating sales revenue based on advertising spend.

Regression Algorithms:

Advantages:

- Simple and interpretable.
- Computationally efficient.

Limitations:

- Assumes a linear relationship between features and target.
- Sensitive to outliers.

2. Polynomial Regression

Extends linear regression by adding polynomial terms (e.g., y=ax2+bx+cy=ax2+bx+c) to model nonlinear relationships.

Regression Algorithms:

Use Cases:

- Predicting growth rates that follow a nonlinear trend.

Advantages:

- Can model more complex relationships than linear regression.

Limitations:

- Prone to overfitting if the polynomial degree is too high.

Linear Regression : Steps

- **Steps for Linear Regression**
- **Step 1:** Define the Problem
- Identify the dependent variable (target) and independent variables (features)
- Example: Predict house prices (target) based on house size (feature).

Linear Regression : Steps

Size (sq. ft.)	Price ($)
1,500	300,000
2,000	400,000
2,500	500,000
3,000	600,000
3,500	700,000
4,000	750,000

Linear Regression : Steps

- **Step 2:** Collect and Prepare the Data
- Gather a dataset containing both the features and the target variable.
- Visualize the Data

Step 3: Explore and Visualize the Data

- Use scatter plots to visualize the relationship between the features and the target.
- Check for linearity, outliers, and trends.
- Example: Plot house size (X-axis) vs. price (Y-axis).

Linear Regression : Steps

Step 4: Split the Data

- Divide the dataset into training and testing sets.
- Example: Use 80% of the data for training and 20% for testing.

Step 5: Define the Linear Regression Model

- The model equation for simple linear regression is:
-

$$y = b + m \times x$$

Linear Regression : Steps

- Where:

y: Dependent variable (target).

x : Independent variable (feature).

b: Y-intercept.

m: Slope.

- **Step 6: Train the Model**
- Use the training data to find the best values for b and m.
- Example: Predict house prices for the test set.

Linear Regression : Steps

- **Step 7:** Make Predictions
- Use the trained model to predict the target variable for the test data.
- Example: Predict house prices for the test set.
- **Step 8:** Evaluate the Model
- Use evaluation metrics to assess the model's performance:
- Mean Squared Error (MSE):

$$\text{MSE}=\frac{1}{N}\sum_{i=1}^{n}(y_i-\hat{y}_i)2$$

Linear Regression : Steps

Root Mean Squared Error (RMSE):

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i-\hat{y}_i)\,2}$$

Where:

y_i: Actual value.

$\hat{y}_i$: Predicted value.

Linear Regression : Steps

- **Step 9:** Interpret the Results

Analyze the model’s coefficients (b and m) to understand the relationship between the features and the target.

- Example: If m=150, it means that for every additional square foot, the house price increases by $150.

Linear Regression : Steps

- **Step 10:** Deploy the Model

Use the trained model to make predictions on new, unseen data.

- Example: Predict the price of a new house based on its size.

Linear Regression Equation

- The equation for simple linear regression is

$$y = b + m \times x$$

- Where:
- y: Dependent variable (target, e.g., house price).
- x: Independent variable (feature, e.g., house size).
- b: Y-intercept (the value of y when x=0).
- m: Slope (how much y changes for a unit change in x).

Linear Regression Equation

Explanation of b and m

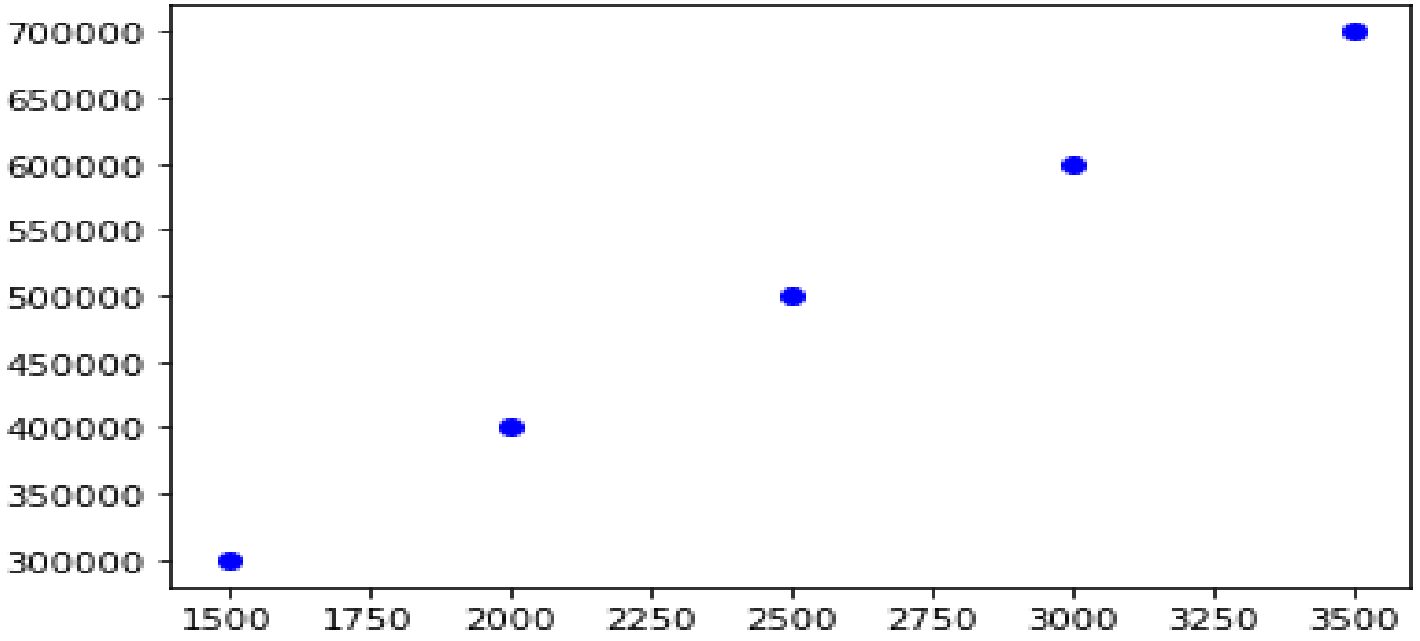
- Y-intercept (b):
- Represents the value of y when x=0.
- In the house price example:
- If b=50,000, it means that when the house size is 0 square feet, the predicted price is $50,000.
- Note: This is a mathematical artifact and may not have a practical meaning in real-world scenarios (e.g., a house cannot have 0 square feet).

Linear Regression Equation

2. Slope (m):

- Represents the change in y for a unit change in x.
- In the house price example:
- If m=150m, it means that for every additional square foot, the house price increases by $150.

Linear regression graph



Linear regression: example

- Problem Statement:

➢ A government wants to understand how the unemployment rate affects consumer spending. The data for the past 5 years is as follows:

Unemployment Rate (x) (%)	Consumer Spending (y) (in $1,000s)
5.0	120
5.5	115
6.0	110

Linear regression: example

- We want to find the equation of the regression line:
- y=b+mx
- where:
- y = Consumer Spending (in $1,000s)
- x = Unemployment Rate (%)
- b = y-intercept
- m = slope

Linear regression: example

- Step 1: Compute the necessary sums

First, calculate the following sums for the given data:

n=5(number of observations)

$\sum x$=5.0+5.5+6.0+6.5+7.0=30.0

$\sum y$=120+115+110+105+100=550

$\sum xy$=(5.0×120)+(5.5×115)+(6.0×110)+(6.5×105)+(7.0×100)

$\sum xy$=600+632.5+660+682.5+700=3275

$\sum x^2$=(5.0)2+(5.5)2+(6.0)2+(6.5)2+(7.0)2∑x2=(5.0)2+(5.5)2 +(6.0)2+(6.5)2+(7.0)2 =25+30.25+36+42.25+49=182.5

Linear regression: example

- Step 2: Calculate the slope (m)
- Use the formula for the slope:
- $m=\frac{n\sum(xy)-(\sum x)(\sum y)}{n\sum(x^2)-(\sum x)^2}$
- Substitute the values:
- $m=\frac{5\times3275-30.0\times550}{5\times182.5-(30.0)2}$

$m=-10$

Linear regression: example

Step 3: Calculate the y-intercept (b)

Use the formula for the y-intercept:

$$b = \bar{y} - m\bar{x}$$

where:

$\bar{x}=\frac{\sum x}{n}$=30/05=6.0

$\bar{y}=\frac{\sum y}{n}$=550/5=110

Substitute the values:

b=110−(−10)×6.0=110+60=170

Linear regression: example

- Final Equation :
- The equation for simple linear regression is:
- $y=170-10x$
- Interpretation :
- b=170: If the unemployment rate (x) is 0%, the predicted consumer spending (y) is $170,000.
- m=−10 : For every 1% increase in the unemployment rate (x), consumer spending (y) decreases by $10,000.

Linear regression: example

- Prediction Example :
- If the unemployment rate is 6.2%, the predicted consumer spending is:
- $y=170-10\times6.2=170-62=108$
- So, the predicted consumer spending is $108,000.

Linear regression: example

- Visualization :
- We can plot the regression line $y=170-10x$ on a graph with the unemployment rate on the x-axis and consumer spending on the y-axis. The line will show a downward slope, indicating that as unemployment increases, consumer spending decreases.