SUPERVISED LEARNING BASICS

-MUKESH KUMAR

AGENDA

- Definition and Overview
- Key Terminologies: Features, Labels, Training Data, Testing Data
- Difference between Regression and Classification

What is Supervised Learning

- Supervised learning is a type of machine learning algorithm that learns from labeled training data to make predictions or decisions.

- The training data consists of input data paired with the desired output, which acts as a "supervisor" to guide the learning process.

- The goal is for the algorithm to learn a mapping function that can accurately predict the output for new, unseen input data.

KEY TERMINOLOGIES

Features

- Features are the input variables or attributes used by the model to make predictions.
- Each feature represents a specific characteristic or aspect of the data.
- **Example:** In a house price prediction model, features might include the number of bedrooms, square footage, location, etc.

Features

- Features can be:
  - Numerical: Quantifiable values, such as age or height.
  - Categorical: Qualitative values, such as gender or color.
  - Text-based: Data derived from text, such as reviews or comments.

- In a dataset, features are typically represented as columns. For example, in a house price prediction model, features might include the number of bedrooms, square footage, and location

Labels

- Labels are the target variable or output that the model is trying to predict. In supervised learning, each data point in the training set has a corresponding label.

- **Example:** In a house price prediction model, the label would be the actual price of the house.

Training Data

- Training data is the dataset used to train the machine learning model.
- It contains input features along with their corresponding labels, allowing the model to learn the relationship between inputs and outputs.
- **Example:** A dataset containing various features of houses along with their prices is used to train the model.

Testing Data

- Testing data is a separate portion of the dataset used to evaluate the performance of the model.
- The model has not seen this data during training, so it helps in assessing how well the model generalizes to new, unseen data.
- **Example:** After training a house price prediction model, it is tested on a new set of houses (with known prices) to see how accurately it can predict those prices.

REGRESSION VS CLASSIFICATION

Classification

- This task involves predicting discrete labels or categories based on input features.
- The output is categorical, meaning it classifies data points into distinct classes.
- For example, determining whether an email is "spam" or "not spam" is a classification problem.

Classification

- **Objective:** Predict a discrete label or category.
- **Output:** A class label, such as 'spam' or 'not spam', 'cat' or 'dog'.
- **Examples:**
  - **Email Spam Detection:** Classifying emails as spam or not spam.
  - **Image Classification:** Identifying objects within images (e.g., identifying whether an image contains a cat or a dog).

Classification

- **Algorithms:** Logistic regression, decision trees, random forests, support vector machines (SVM), etc.

- **Evaluation Metrics:** Accuracy, precision, recall, F1 score, confusion matrix, etc.

Regression

- This task involves predicting continuous numerical values.
- The output is a real number, which can represent a wide range of values.
- For instance, predicting the price of a house based on its features (size, location, etc.) is a regression problem.

Regression

- **Objective:** Predict a continuous numerical value.
- **Output:** A real-valued number, such as the price of a house, temperature, or salary.
- **Examples:**
  - **House Price Prediction:** Predicting the price of a house based on its features (e.g., size, location).
  - **Stock Price Forecasting:** Estimating the future price of a stock.

Regression

- **Algorithms:** Linear regression, polynomial regression, support vector regression (SVR), etc.

- **Evaluation Metrics:** Mean squared error (MSE), mean absolute error (MAE), R-squared, etc.