

Prior Knowledge

Supervised and unsupervised learning:

Supervised and Unsupervised learning are the two techniques of machine learning. But both the techniques are used in different scenarios and with different datasets.

Supervised Learning:

Supervised learning uses labeled input and output data, Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.

Unsupervised Learning:

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.

Classification:

Classification is the type of supervised machine learning. For any given input, the classification algorithm help in the prediction of the class of the output variables, there can be multiple type of classification are - binary classification, multi-class classification. Types of classification - K- Nearest Neighbour Logistic regression Decision tree Random forest Naive Bayes SVM (Support vector machine)

Regression :

Regression is a statistical technique that relates a dependent variable to one or more independent (explanatory) variables. A regression model is able to show whether changes observed in the dependent variable are associated with changes in one or more of the explanatory variables.

Types of Regression

- Logistic Regression
- Linear Regression
- Polynomial Regression
- Random Forest Regression
- Gradient Boosting Regression

Clustering : Clustering is unsupervised machine learning algorithm, it is used to group data point having similar characteristics as cluster.

Clustering is divided into two groups

1. Hard clustering : In hard clustering, the data point is assigned to one of the clusters only.
2. Soft clustering : It provides a probability likelihood of a data point to be in each of the clusters.

Data Visualization:

The process of finding trends and correlations in our data by representing it pictorially is called Data Visualization. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. To perform data visualization in python, we can use various python data visualization modules such as Matplotlib, Seaborn etc.

Matplotlib and Seaborn:

Matplotlib and Seaborn are python libraries that are used for data visualization. They have inbuilt modules for plotting different graphs. While Matplotlib is used to embed graphs into applications, Seaborn is primarily used for statistical graphs.

Python Flask :

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.