V.S.B. ENGINEERING COLLEGE, KARUR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

IBM NALAIYA THIRAN

PRIOR KNOWLEDGE

Date	17 November 2022
Team ID	PNT2022TMID33300
Project Name	Fertilizers Recommendation System for Disease Prediction

PRIOR KNOWLEDGE

MACHINE LEARNING ALGORITHMS

Machine Learning algorithms are the programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences on their own. Different algorithms can be used in machine learning for different tasks, such as simple linear regression that can be used for prediction problems like stock market prediction, and the KNN algorithm can be used for classification problems. In this topic, we will see the overview of some popular and most commonly used machine learning algorithms along with their use cases and categories.

TYPES OF MACHINE LEARNING ALGORITHMS

Machine Learning Algorithm can be broadly classified into three types:

1. Supervised Learning Algorithms
2. Unsupervised Learning Algorithms
3. Reinforcement Learning algorithm

The below diagram illustrates the different ML algorithm, along with the categories:

1) SUPERVISED LEARNING ALGORITHM

Supervised learning is a type of Machine learning in which the machine needs external supervision to learn. The supervised learning models are trained using the labeled dataset. Supervised learning can be separated into two types of problems when data mining: classification and regression:

- **Classification** problems use an algorithm to accurately assign test data into specific categories, such as separating apples from oranges. Or, in the real world, supervised learning algorithms can be used to classify spam in a separate folder from your inbox. Linear classifiers, support vector machines, decision trees are all common types of classification algorithms.
- **Regression** is another type of supervised learning method that uses an algorithm to understand the relationship between dependent and independent variables. Regression models are helpful for predicting numerical values based on different data points, such as sales revenue projections for a given business. Some popular regression algorithms are linear regression, logistic regression and polynomial regression.

2) UNSUPERVISED LEARNING ALGORITHM

It is a type of machine learning in which the machine does not need any external supervision to learn from the data, hence called unsupervised learning. Unsupervised learning models are used for three main tasks: clustering, association and dimensionality reduction:

- **Clustering** is a data mining technique for grouping unlabelled data based on their similarities or differences. For example, K-means clustering algorithms assign similar data points into groups, where the K value represents the size of the grouping and granularity. This technique is helpful for market segmentation, image compression, etc.
- **Association** is another type of unsupervised learning method that uses different rules to find relationships between variables in a given dataset. These methods are frequently used for market basket analysis and recommendation engines, along the lines of "Customers Who Bought This Item Also Bought" recommendations.
- **Dimensionality reduction** is a learning technique used when the number of features (or dimensions) in a given dataset is too high. It reduces the number of data inputs to a manageable size while also preserving the data integrity. Often, this technique is used in the pre-processing data stage, such as when autoencoders remove noise from visual data to improve picture quality.

3) REINFORCEMENT LEARNING

In Reinforcement learning, an agent interacts with its environment by producing actions, and learn with the help of feedback. In reinforcement learning, a program, a so-called agent, should independently develop a strategy to perform actions in an environment. For this purpose, positive or negative reinforcements are conveyed, which describe the interaction interactions of the agent with the environment. In other words, immediate feedback on an executed task. The program should maximize rewards or minimize punishments. The environment is a kind of simulation scenario that the agent has to explore.

List of Popular Machine Learning Algorithm:

- Linear Regression Algorithm
- Logistic Regression Algorithm
- Decision Tree
- SVM
- Naïve Bayes
- KNN
- K-Means Clustering
- Random Forest
- Apriori
- PCA

1. **Linear Regression:**

Linear regression is one of the most popular and simple machine learning algorithms that is used for predictive analysis. Here, predictive analysis defines prediction of something, and linear regression makes predictions for continuous numbers such as salary, age, etc.

2. **Logistic Regression:**

Logistic regression is the supervised learning algorithm, which is used to predict the categorical variables or discrete values. It can be used for the classification problems in machine learning, and the output of the logistic regression algorithm can be either Yes or NO, 0 or 1, Red or Blue.

3. **Decision Tree Algorithm:**

A decision tree is a supervised learning algorithm that is mainly used to solve the classification problems but can also be used for solving the regression problems. The internal node is used to represent the features of the dataset, branches show the decision rules, and leaf nodes represent the outcome of the problem.

4. **Support Vector Machine Algorithm**:

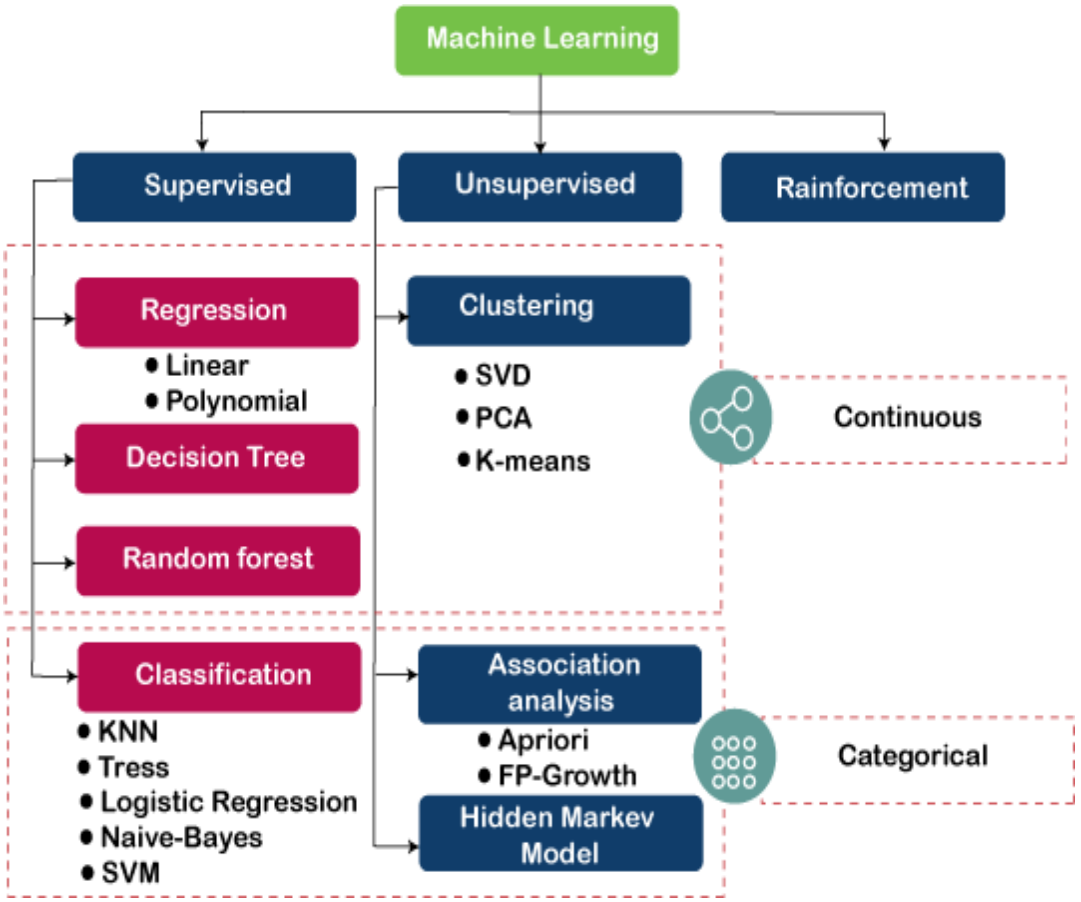
A support vector machine or SVM is a supervised learning algorithm that can also be used for classification and regression problems. However, it is primarily used for classification problems. The goal of SVM is to create a hyper plane or decision boundary that can segregate datasets into different classes. The data points that help to define the hyperplane are known as support vectors, and hence it is named as support vector machine algorithm. Some real-life applications of SVM are face detection, image classification, Drug discovery.

5. K-Nearest Neighbour (KNN):

K-Nearest Neighbour is a supervised learning algorithm that can be used for both classification and regression problems. This algorithm works by assuming the similarities between the new data point and available data points. Based on these similarities, the new data points are put in the most similar categories. It is also known as the lazy learner algorithm as it stores all the available datasets and classifies each new case with the help of K-neighbours. The new case is assigned to the nearest class with most similarities, and any distance function measures the distance between the data points. The distance function can be Euclidean, Minkowski, Manhattan, or Hamming distance, based on the requirement.

6. K-Means Clustering:

K-means clustering is one of the simplest unsupervised learning algorithms, which is used to solve the clustering problems. The datasets are grouped into K different clusters based on similarities and dissimilarities.



Flask:

Flask is a web development framework. It is a framework with a built-in development server and a debugger. Flask framework in itself is different from the other archetypes as it allows web developers to be flexible and to comfortably accommodate the frequently released changes in the software development community. Unlike the Django framework, Flask is very Pythonic. It's easy to get started with Flask, because it doesn't have a huge learning curve. On top of that it's very explicit, which increases readability. To create the "Hello World" app, you only need a few lines of code.