

Basic ML

Shrey Batra

What is Machine Learning?

Machine Learning is a field of Computer Science that give systems the ability to “learn” with data, without being explicitly programmed.

It is the process to make the system learn from examples (or without) and find patterns to identify the output based on number of features.

Types of Learning

- Supervised - In this type of learning we have examples which have both input features and their corresponding outputs/labels, from which we tell the algorithm to find patterns with similar outputs.
- Unsupervised - In this type of learning we have examples which have only input features from which we tell the algorithm to find patterns amongst all the data points.

Types of Machine Learning Algorithms

Almost all Machine Learning algorithms can be broken into 2 types -

1. Classification - In which we predict the label/class, which are defined explicitly (eg. Gender, Boolean, etc.)
2. Regression - In which we predict the output, which are defined in a continuous manner (eg. Prices, distances, etc.)

Most Common Supervised Algorithms

1. Naive Bayes
2. SVM
3. Decision Trees
4. Logistic Regression

Naive Bayes

Naive Bayes is a simple but surprisingly powerful algorithm for predictive modeling. It uses Bayes Theorem to calculate the probability of each feature in a document and then predict the class based on the probabilities calculated.

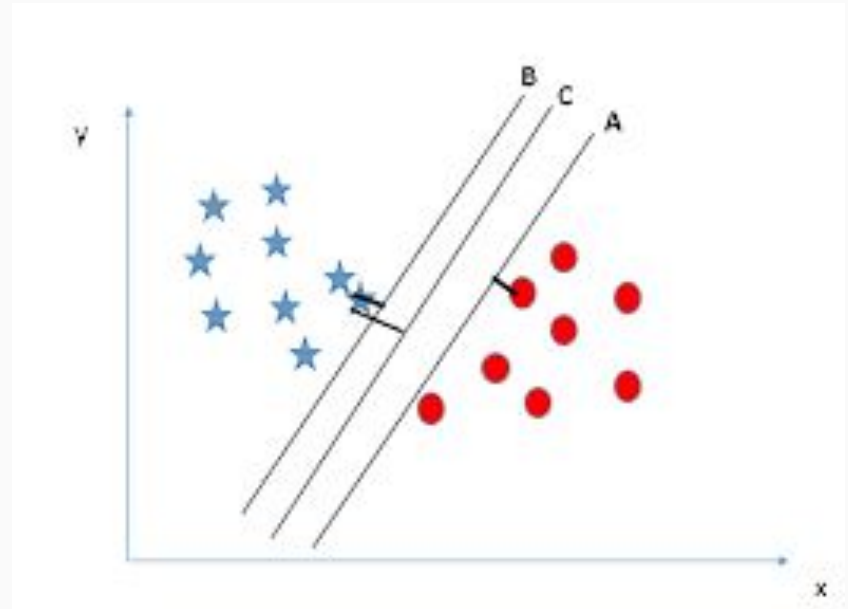
Used mostly in text pattern recognitions, text searches, etc.

Naive, because it just depends on the existence of a feature and not on the length and order of the features. Eg. Chicago Bulls and Bulls Chicago may be different but Naive Bayes would result in same output.

SVM - Support Vector Machine

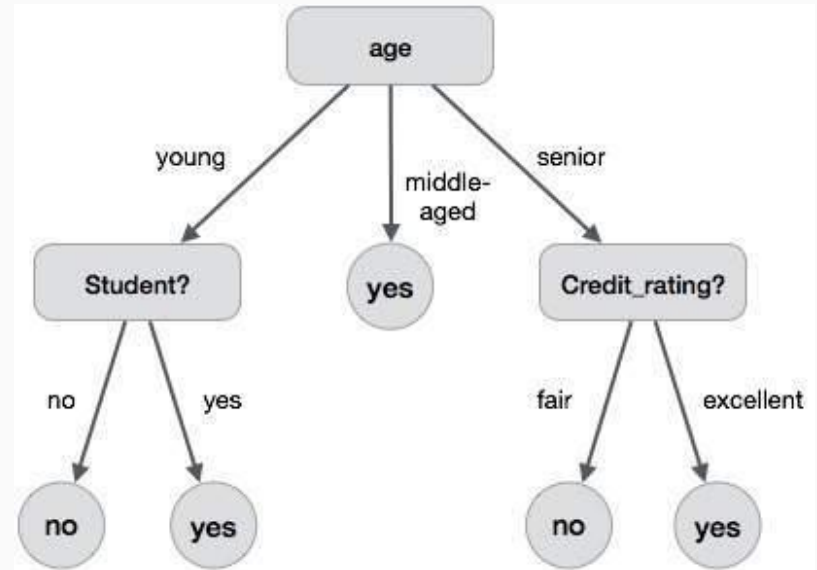
In machine learning, support vector machines (**SVMs**, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

It tries to draw a N dimensional plane to distinguish N different classes where N may be the number of input features.



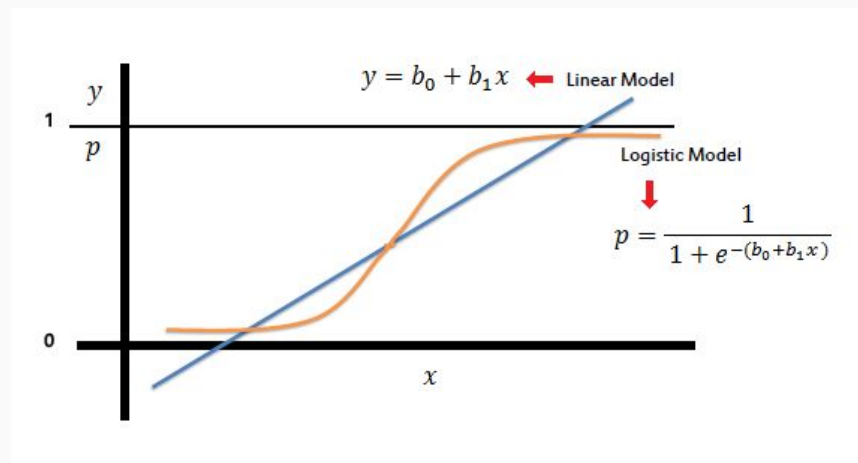
Decision Trees

A **decision tree** is a **decision** support tool that uses a **tree**-like graph or model of **decisions** and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. It just uses If-else statements to break down the decision.



Logistic Regression

It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. It differs from Linear Regression as it uses probability to determine the class/label of the input features. This is based on various Logistic Theorems like sigmoid function, tan inverse function, etc. It weights each feature and finds the optimal weight of each feature in the decision making.



Most Common Un-Supervised Algorithms

1. K - Means
2. K - Median
3. K - Medoid
4. Various other clustering Techniques