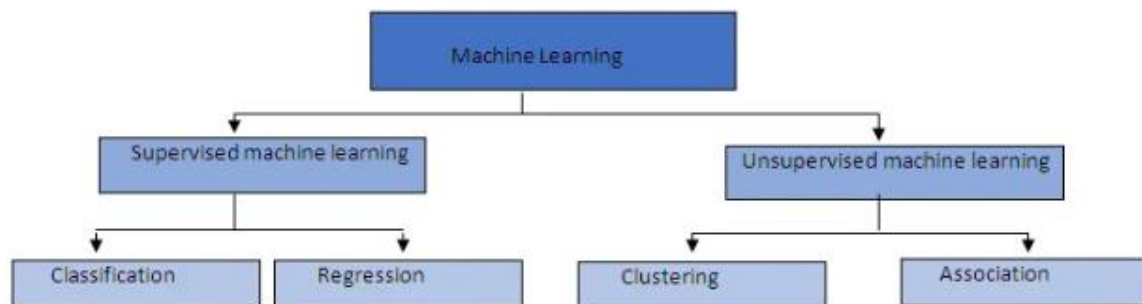
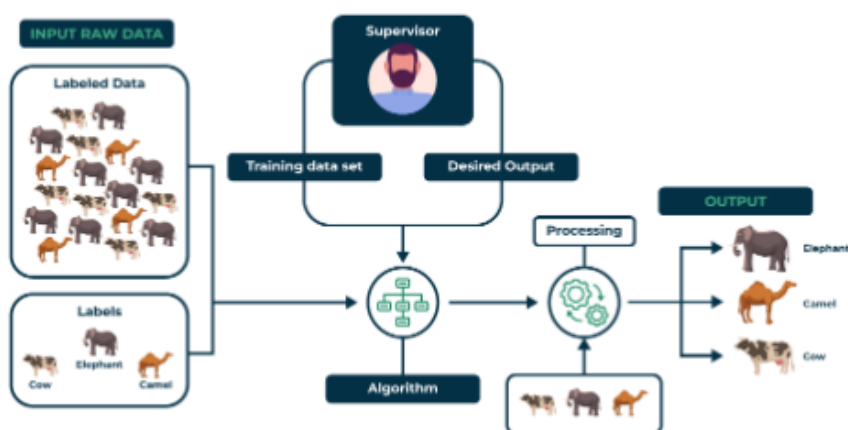


# Understanding Supervised Learning Algorithms: A study about Classification and regression



## What is supervised learning?

Supervised learning is a machine learning concept, where the input is labeled and gives corresponding output. In supervised learning, learning happens with the help of supervision. It aims to develop a model that can predict the output for new, unseen data. Supervised learning has a wide range of applications in different fields including recommendation systems, natural language processing, finance, healthcare, and marketing. Students learn under the supervision of the teacher is also an example of supervised learning.



Some examples of supervised learning algorithms are:

**Decision Tree:** This supervised learning technique can be used in classification and regression problems. It is used in classification problems. This classifier is a tree-structured with internal nodes standing for dataset attributes, branches for decision rules, and leaf nodes for each outcome. It

consists of two nodes, a Decision node, and a leaf node. Decision node. It is easy to understand because it is tree-structured.

**Support Vector Machine:** One of the most widely used supervised learning techniques for both classification and regression issues is support vector machine or SVM. But it's mostly applied to machine learning classification challenges. The algorithm aims to create the best line or decision boundary that can separate the data points into the correct category. The boundary for separating classes is referred to as a hyperplane. The hyperplane is created by choosing the extreme points or vectors.

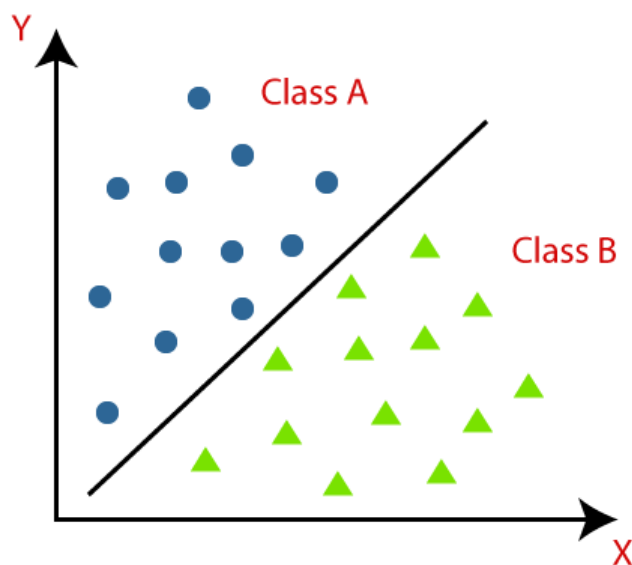
**Random Forests:** It is a machine learning algorithm that belongs to the supervised learning technique. **It is used for classification as well as regression problems.** This ensemble learning technique multiple classifiers to solve complex problems. It contains multiple decision trees and thus improves accuracy and prevents overfitting.

**Naïve Bayes:** Naïve Bayes consists of two words Naïve and Bayes. Naïve gives an assumption that one feature is independent of the other feature. Each feature can identify the object without depending on the other. Bayes depends on the Bayes Theorem. It is also a supervised learning algorithm based on the Bayes Theorem used for solving classification problems. It is used in text classification which has high dimensional training data.

There are two types of supervised learning:

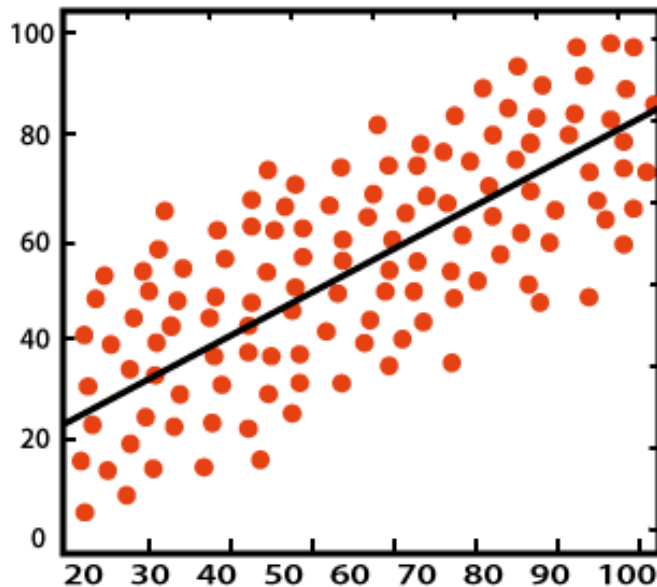
## Classification

In classification, the models are classified into categories. If there are two categories, it is called binary classification. If there are more than two categories, it is called multi-class classification. Classification algorithms are used when the target variable is categorical, which means it gives a discrete set of categories. Classification algorithms divide the output into positive and negative based on the input features. There are different classification algorithms like Support Vector Machine(SVM), Decision Tree, K-Nearest Neighbour(KNN), etc.



## Regression

Regression refers to a supervised learning task where the output is based on predicting continuous value based on input features. Regression's primary goal is to use training data to understand the relationship between the continuous target variable and the input features so that the model can accurately predict new, unknown data.



There are different types of regression:

**Linear Regression:** It is one of the simple algorithms which gives the relationship between continuous variables. Linear regression is called simple linear regression if there is only one input. Linear regression is called multiple linear regression if there are many input variables. In linear regression, the graph will give a linear line.

**Logistic Regression:** This is another approach in supervised learning, which is used to address classification issues. Here we have a dependent variable which is either 0 or 1. Logistic regression works based on probability. It is different from linear regression by how they are used.

**Polynomial Regression:** This approach expands linear regression by adding a polynomial function to the data. This helps in solving more complex relationships. If the data points given are not in a linear fashion, then we can use polynomial regression.

**Support Vector Regression:** It is a type of support vector machine. This is when we use a support vector machine algorithm in regression. It works for continuous variables. It tries to find a function that predicts the continuous output value to the given input value. Both linear and non-linear kernels can be used with SVR. A non-linear kernel is a more complicated function that identifies more subtle patterns in the data than a linear kernel, which is only a dot product of two input vectors. The properties of the data and the complexity of the task determine the kernel to use.

## Difference between Classification and Regression

Regression	Classification
The output must be continuous in nature or real value.	The output must be categorical.
The task of this algorithm map the input value into a continuous output variable.	The task of this algorithm map the input value into a discrete output variable.
Used for continuous data.	Used for discrete data.
Try to find a best-fit line to predict the output.	Try to find a decision boundary that divides the dataset into different classes.
Used to solve regression problems such as weather prediction, house price prediction, etc.	Used for solving classification problems such as identification of spam emails, speech recognition, etc.