**Support Vector Machine Algorithm**

**Support Vector Machine (SVM)** is a supervised machine learning algorithm used for both classification and regression. It is best suited for classification problems. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. And it becomes difficult to imagine when the number of features exceeds three.

SVM is one of the best-known techniques to optimize the expected solution. SVM was introduced by Vapnik as a machine learning model for classification and regression task. The extraordinary generalization capability of SVM, along with its optimal solution and its discriminative power, has attracted the attention of pattern recognition and machine learning communities in the last years. SVM has been used as a powerful tool for solving practical binary classification problems. It has been shown that SVMs are superior to other supervised learning methods. Due to its good theoretical foundations and good generalization capacity, in recent years, SVMs have become one of the most used classification methods.

## Types of Support Vector Machine

#### **Linear SVM**

When the data is perfectly linearly separable only then we can use Linear SVM. Perfectly linearly separable means that the data points can be classified into 2 classes by using a single straight line (if 2D).

#### **Non-Linear SVM**

When the data is not linearly separable then we can use Non-Linear SVM, which means when the data points cannot be separated into 2 classes by using a straight line (if 2D) then we use some advanced techniques like kernel tricks to classify them. In most real-world applications we do not find linearly separable datapoints hence we use kernel trick to solve them.

**Working of SVM Algorithm**

**Selecting the best hyper-plane:**

One reasonable choice as the best hyperplane is the one that represents the largest separation or margin between the two classes. So, we choose the hyperplane whose distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the ***maximum-margin hyperplane/hard margin***.

Chart, scatter chart, bubble chart

Description automatically generated

**Support Vectors:**These are the points that are closest to the hyperplane. A separating line will be defined with the help of these data points.

**Margin:** it is the distance between the hyperplane and the observations closest to the hyperplane (support vectors). In SVM large margin is considered a good margin. There are two types of margins **hard margin** and **soft margin.**



**SVM Classifier**

Equation of the Hyperplane:

**y = wx - b**

**Gradient Descent:**

Gradient Descent is an optimization algorithm used for minimizing the loss function in various machine learning algorithms. It is used for updating the parameters of the learning model.

**w = w - α\*dw**

**b = b - α\*db**

Graphical user interface, text, application

Description automatically generated

**Learning Rate:**

Learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function.

Scatter chart

Description automatically generated

**Implementation:-**

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text

Description automatically generated

**Pros and Cons associated with SVM**

* **Pros:**
  + It works well with a clear margin of separation
  + It is effective in high dimensional spaces.
  + It is effective in cases where the number of dimensions is greater than the number of samples.
  + It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
* **Cons:**
  + It doesn’t perform well when we have large data set because the required training time is higher
  + It also doesn’t perform very well, when the data set has more noise i.e. target classes are overlapping
  + SVM doesn’t directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.