What is Support Vector Machine?

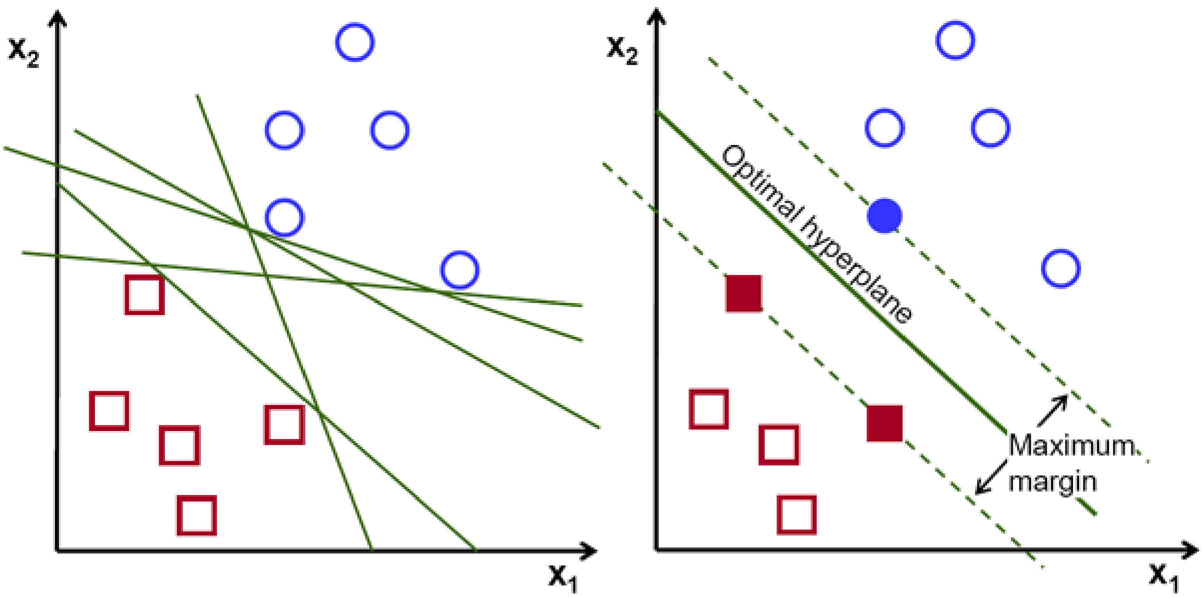
The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

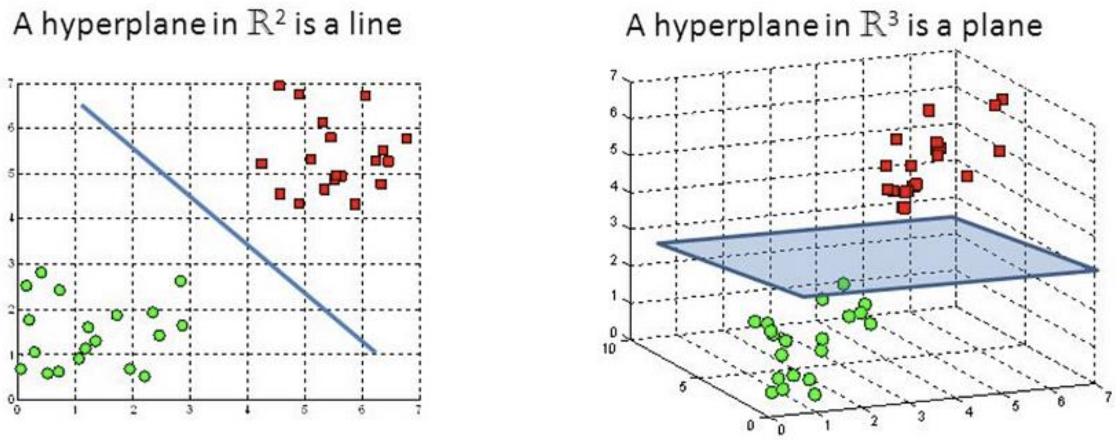
The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as **support vectors**, and hence algorithm is termed as **Support Vector Machine.**



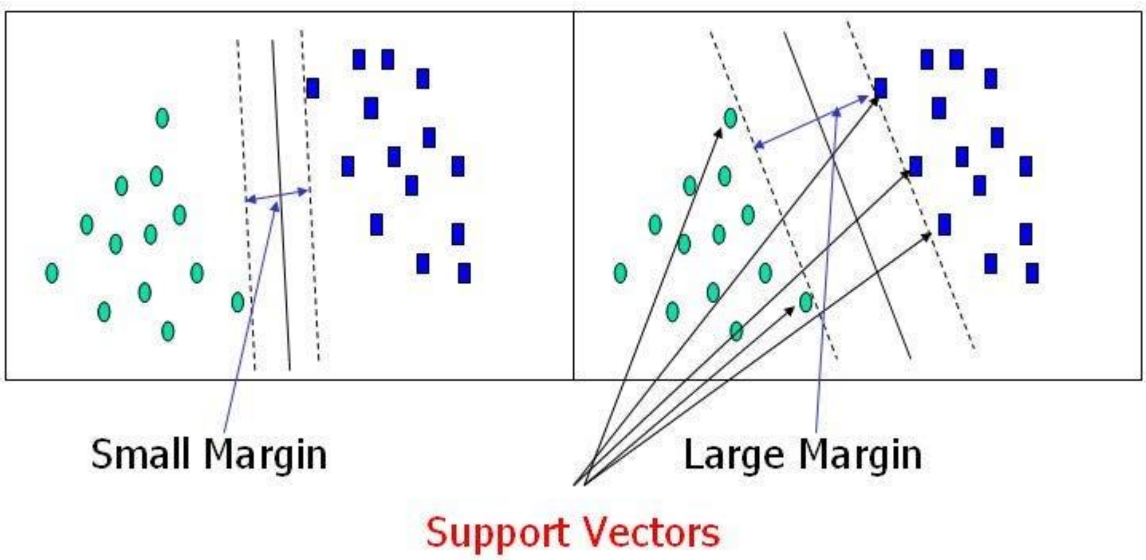
Possible Hyperplanes

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides that future data points can be classified with more confidence.

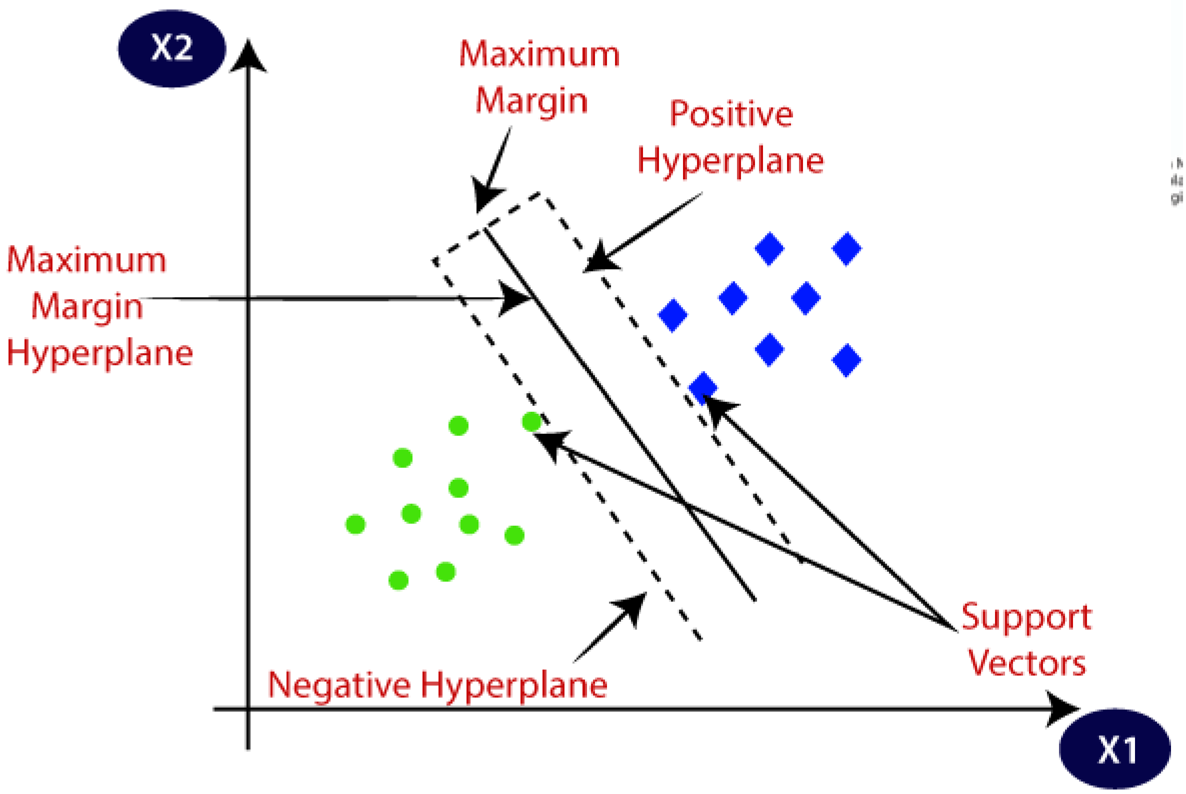


Hyperplanes in 2D and 3D feature space

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. For 1-dimension it can be a point. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane, and for 3 or more dimensions it is a hyperplane. It becomes difficult to imagine when the number of features exceeds 3.



Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Since these vectors support the hyperplane, hence called a **Support vector.** Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM. The **hyperplane** with maximum margin is called the **optimal hyperplane.**



Types of SVM

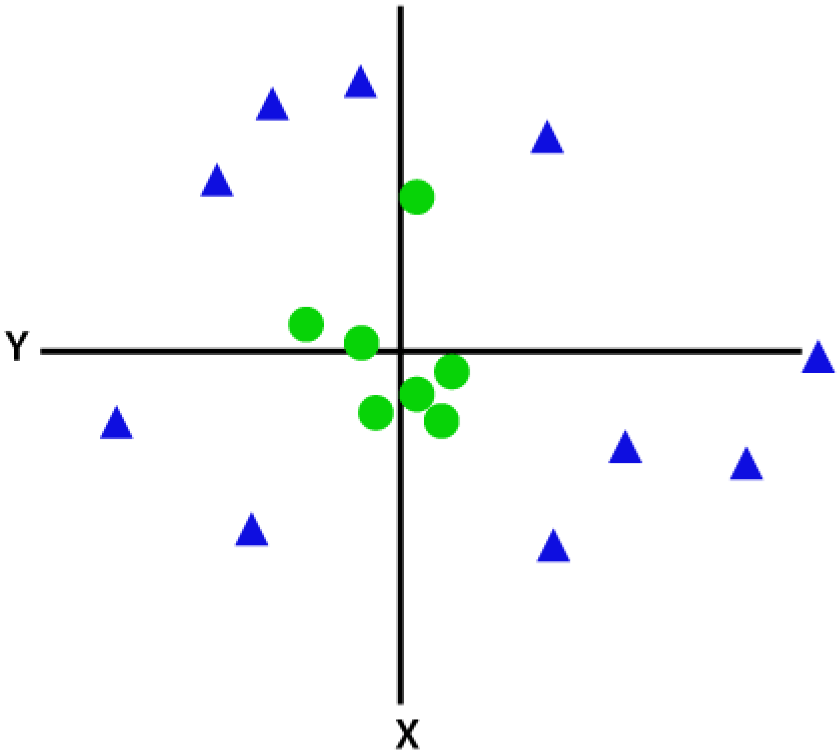
SVM can be of two types:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as **Linear SVM classifier**.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such

data is termed as non-linear data and classifier used is called as **Non-linear SVM classifier.**

Non-Linear SVM:

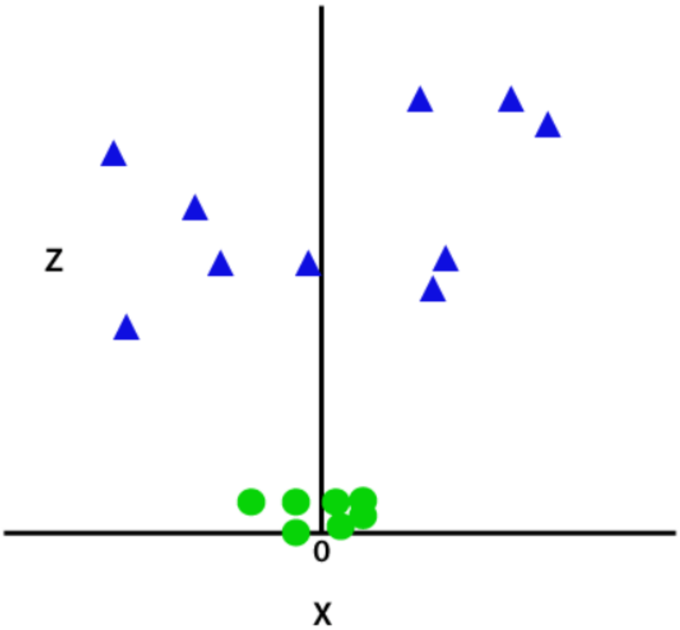
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



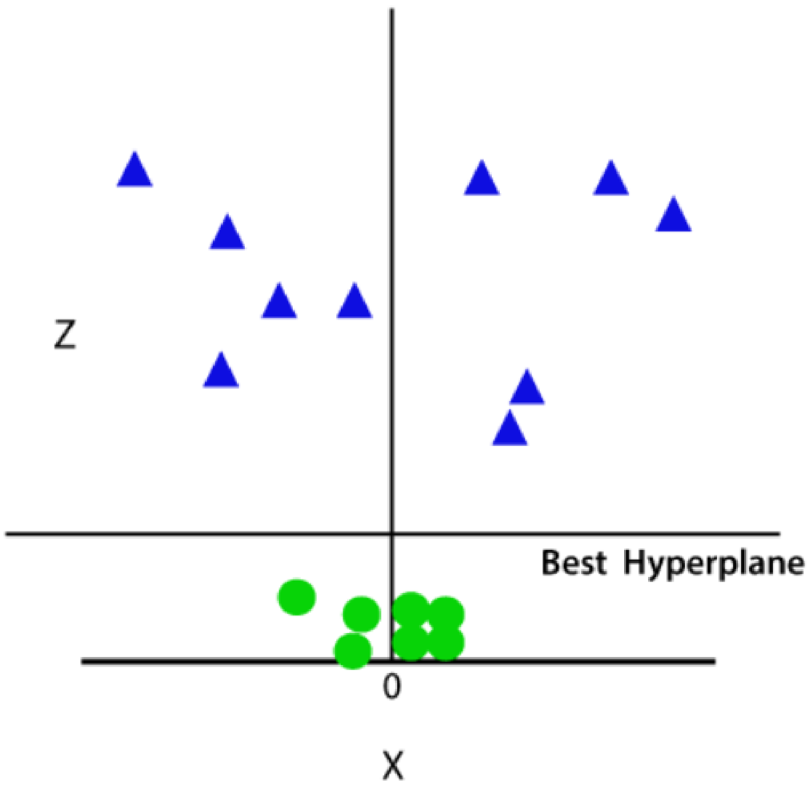
So, to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$Z= x^2 + y^2$

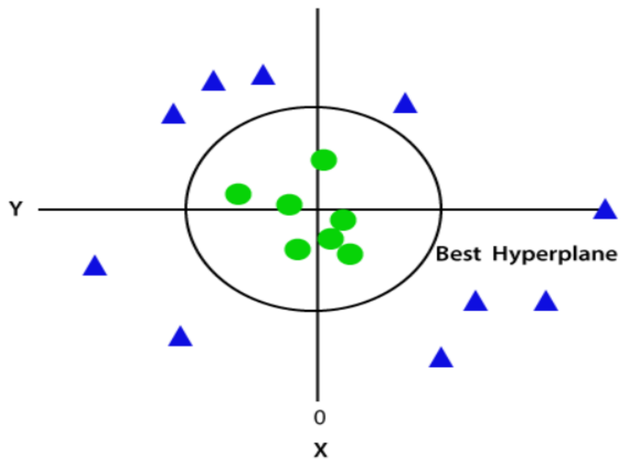
By adding the third dimension, the sample space will become as below image:



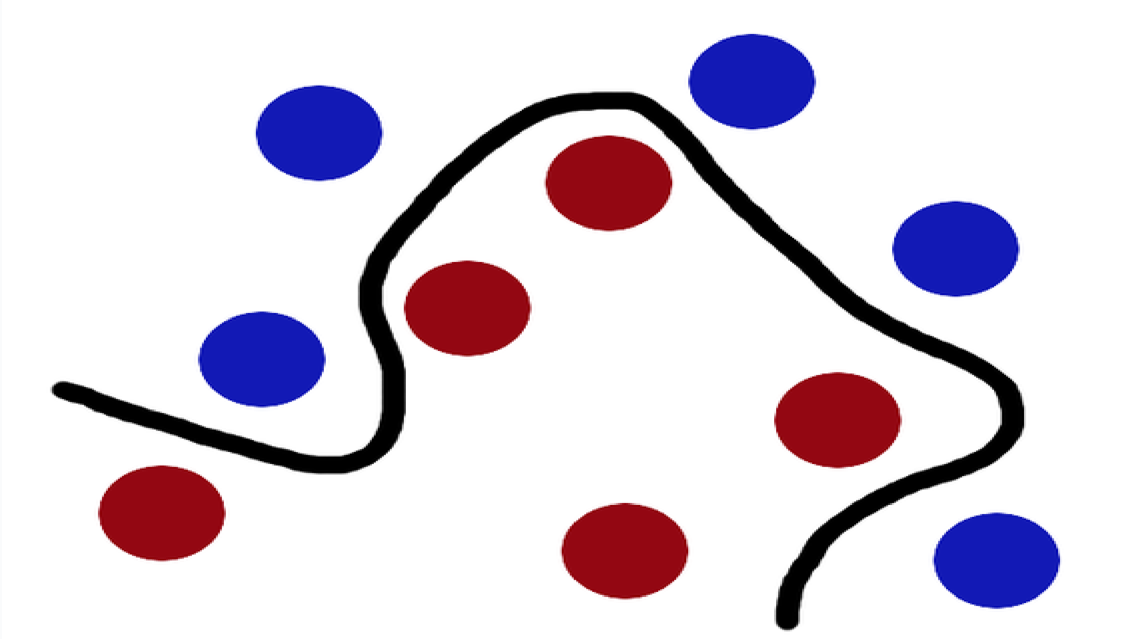
So now, SVM will divide the datasets into classes in the following way. Consider the below image:



Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with z=1, then it will become as:



Hence, we get a circumference of radius 1 in case of non-linear data.



In machine learning, a “**kernel**” is usually used to refer to the kernel trick, a method of using a linear classifier to solve a non-linear problem. It entails transforming linearly inseparable data like above figure to linearly separable ones. The kernel function is what is applied on each data instance to map the original non-linear observations into a higher-dimensional space in which they become separable.

What is a Kernel Function?

This function transforms the n-dimensional input space to an m-dimensional space so that we can do the required calculations in a higher dimension efficiently. We have seen before that a non-linear curve in the lower dimension becomes a linear curve in a higher dimension.

Mathematical definition: *K(x, y) = <f(x), f(y)>. Here K is the kernel function, x, y are n dimensional inputs. f is a map from n-dimension to m-dimension space. Usually, m is much larger than n.*

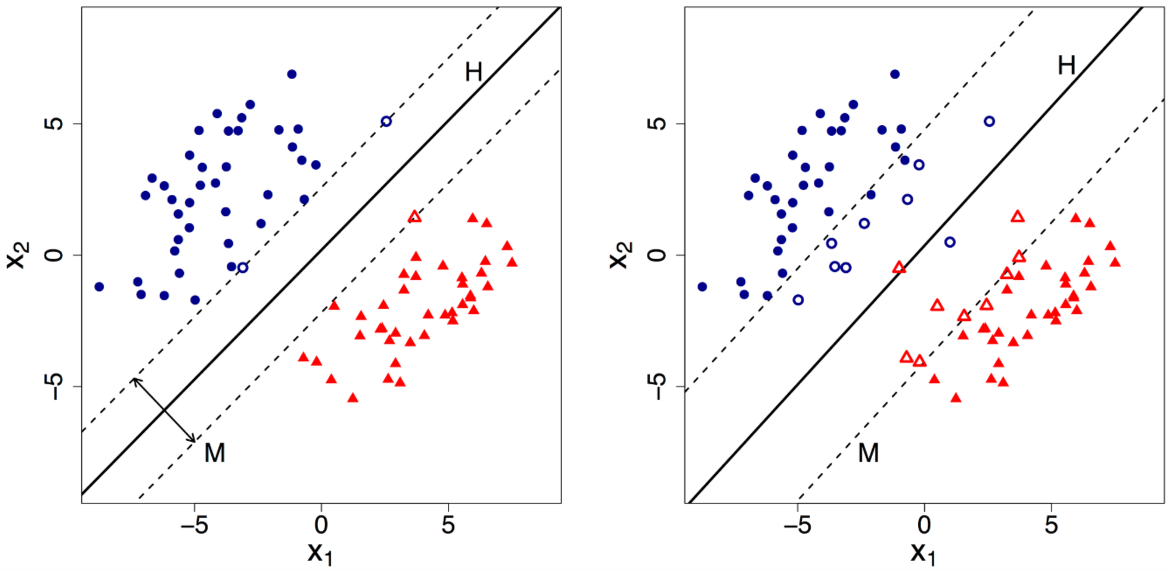
The Kernel Trick

A method where Non-Linear data is projected onto a higher dimension space such that it is easier to classify the data where it could be linearly divided by a plane. The trick is that we are not transforming the data points, we are achieving this mathematically for our convenience.

Linear SVM vs Non-Linear SVM

Linear SVM	Non-Linear SVM
The data points are separated using a single line	The data points are hard to separate using a single line so other shapes are used as a decision boundary.
Data is classified with the help of a hyperplane.	Kernels are used to classify data points.
difficult to classify more than two labels.	Used for classifying more than two labels.

Types of Margin : Hard vs Soft Margin



HP1(hyperplane 1) **HP2(hyperplane 2)**

Suppose we are given 2 Hyperplanes, one with 100% accuracy (HP1) on the left side and another with >90% accuracy (HP2) on the right side. Which one would you think is the correct classifier?

Most of us would pick the HP2 thinking that it is having the maximum margin. But it is the wrong answer. But Support Vector Machine would choose the HP1 though it has a narrow margin. Because though HP2 has maximum margin but it is going against the constraint that: ***each data point must lie on the correct side of the margin and there should be no misclassification.*** This constraint is the ***hard constraint.***

We now clearly can say the left-hand side diagram, HP1 is hard-margin SVM whereas in the right-hand side diagram, HP2 is soft -margin SVM. **Hard margin SVM** does not allow any misclassification to happen. **By default, Support Vector Machine implements Hard margin SVM**. It works well only if our data is linearly separable.

Soft margin SVM allows some misclassification to happen by relaxing the hard constraints of Support Vector Machine. In case our data is non-separable/ nonlinear data, Soft Margin SVM can be applied. **Soft margin SVM is implemented with the help of the Regularization parameter (C).**

Regularization parameter (C): It tells us how much misclassification we want to avoid.

– *Hard margin SVM generally has large values of C.*

– *Soft margin SVM generally has small values of C.*

QUESTIONS FROM SVM ALGORITHM

- Solve numerical example of linear SVM.
- Solve numerical example of non-linear SVM.
- Explain in brief working principle of SVM.
- What are the advantages and disadvantages of SVM algorithm.
- What do you mean by hard margin and soft margin?
- What are support vectors?
- Difference between linear and non-linear SVM.
- Difference between logistic regression and SVM.