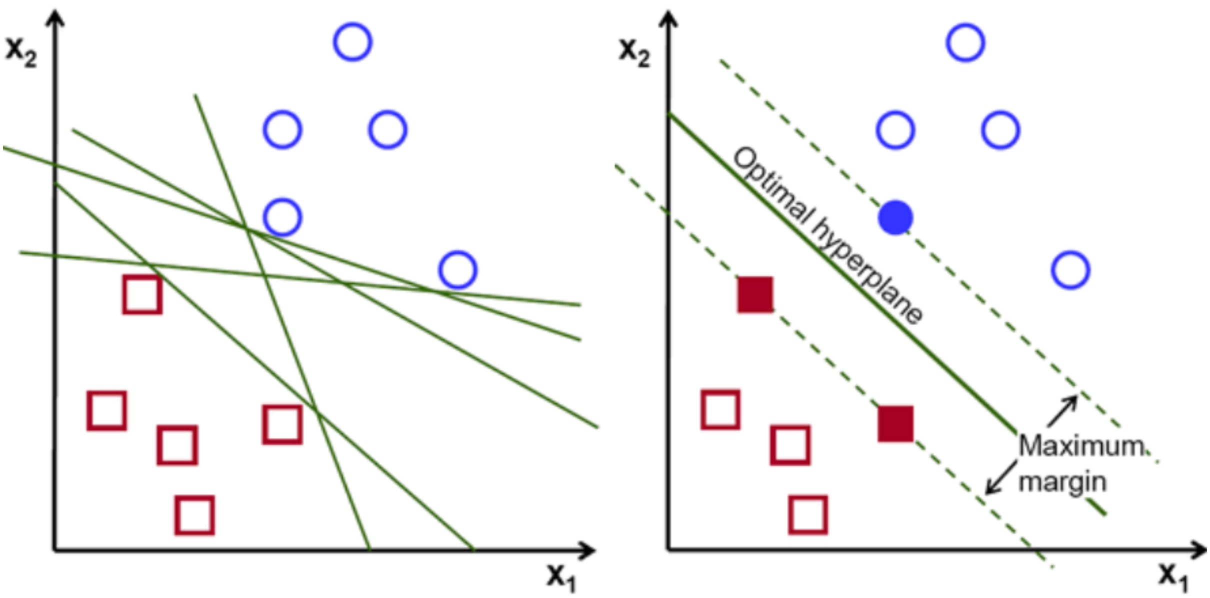
<u>Support Vector Machine (SVM)</u>

Support Vector Machine (SVM) is used to create the best fit line or decision boundary that can segregate the n number of features. This decision boundary is known as hyper-plane.

Hyperplane not only separates the two classes but also tries to maintain the maximum distance between the most extreme points (support vectors) of the two classes.

It is like the global separator for both of the classes which are separated individually by support vectors. Where it also maximizes distance from both sides.



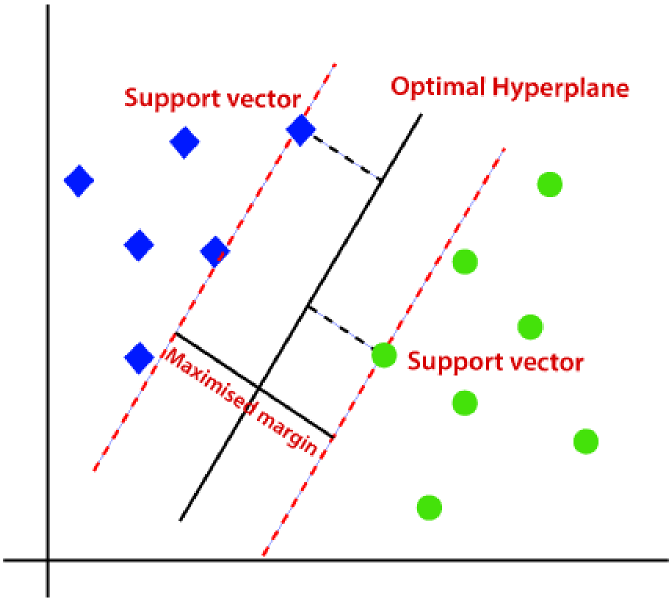
If we will plot clearly separable line it is called hard margin.

Hard margin basically tries to fit a decision boundary that maximises the distance between the support vectors of the two classes.

In hard margin we won’t find errors, we will get clearly separable data points.

Drawbacks of Hard margin

1. It is very sensitive to outliers
2. It only works on data that is linearly separable



It is extremely rare to have a dataset that is linearly separable.

SVM separate non-linearly separable data using **Soft Margin**

It is of two types

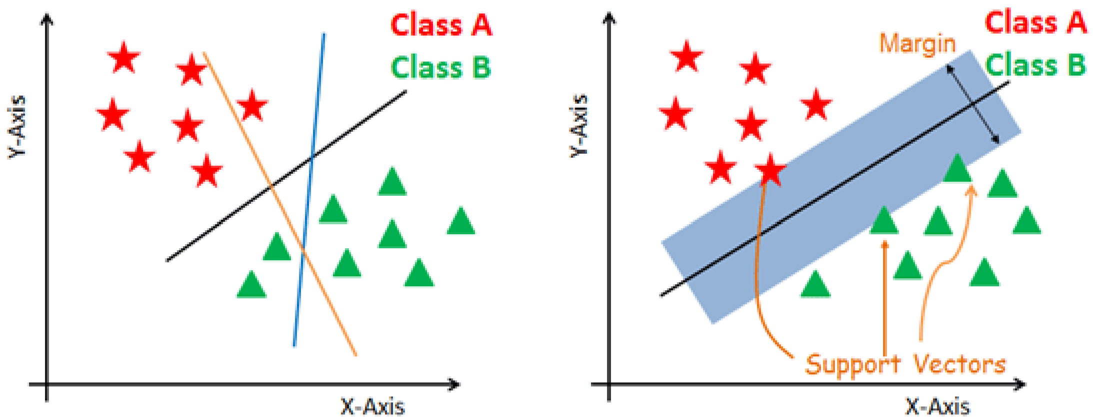
- Classification → SVC Support Vector Classifier
- Regression → SVR Support Vector Regressor

- **Support Vectors:**

The data points or vectors that are the closest to the hyperplane and affect the position of the hyperplane or, in other words, the points located on the edge are called as Support Vector. Deleting the support vectors will change the position of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Since these vectors support the hyperplane, hence called a Support vector.

Support vector machine is highly preferred by many as it produces significant accuracy.

To separate the two classes with data points, we could choose many possible hyperplanes. Our objective is to find a plane with maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provide us the benefit that future data points can be classified more accurately.



- **<u>Hyper-Plane</u>**

Hyper plane is another plane just like best fit line which crosses near the support vectors. It should be equidistance from the decision boundaries.

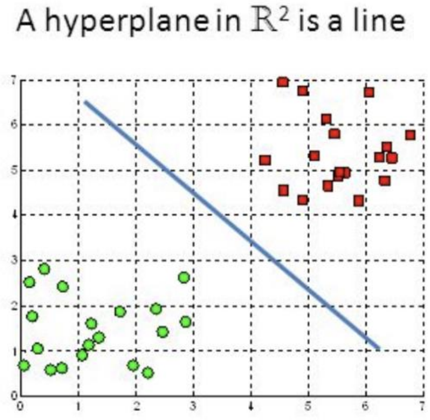
Hyper-plane is also called the decision boundaries that segregate the data points of both classes.

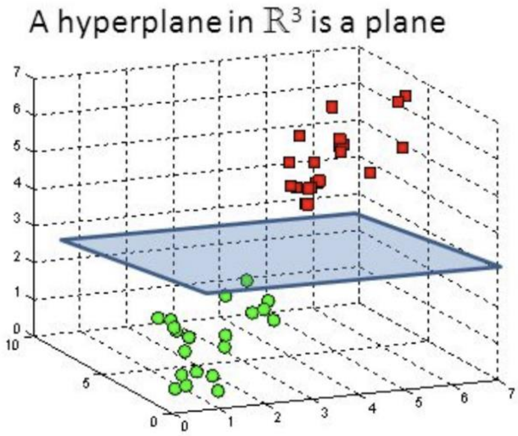
We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

The dimension of the hyperplane depends upon the number of features. 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. Where we add a kernel to separate the points.

The distance between the vectors and the hyperplane is called as **margin**. And the goal of SVM is to maximize this margin.

The **hyperplane** with maximum margin is called the **optimal hyperplane**.





- **<u>Kernel:</u>**

Kernel is a mathematical function that take data as input and converts them into required form.

These are generally used for finding hyper-plane in higher dimension.

The most widely used kernels are Linear, Non-Linear, Radial Basis Function (RBF), Polynomial, and sigmoid. By-default RBF is used as the kernel. Each of these kernels are used depending on dataset.

1. Support Vector Classifier:

Cost function

The cost function that helps maximize the margin.

$$\min \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$

C trades-off margin width
and misclassifications

C = number of error points

ε = distance of error points from marginal plane

Hinge Loss

A cost function is an error rate that tells you how well your model is performing by means of a specific mathematical formula. Hinge loss specially used for classification problem to maximize the margin distance and it leads to better accuracy.

$$\ell(y) = \max(0, 1 - t \cdot y)$$

l = Hinge loss

y = prediction

t = actual target for the prediction, assume t is either +1 or -1

2. SVR (Support Vector Regression)

If data points are linearly separated then we can go for linear regression, we also can support vector regression, but we will avoid it because of time constraints.

The main aim of SVR to minimize the cost function. In support Vector Regression you can plot a non-linear hyper plane with the help of kernel trick which will look like a curve rather than line.

So SVR allows linear as well as non-linear fitting plane/line where as Linear Regression allows only the linearly fitted line.

In SVR we find more generalized model compared to linear regression.

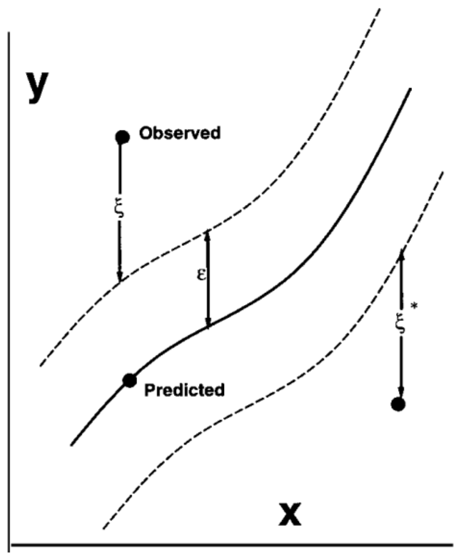
Non-Linear SVM:

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.

So to separate these data points, we need to add one more dimension. for non-linear data, we will add a third dimension z.

it is extremely rare to have a dataset that is linearly separable

To solve this problem we need to add more features, such as polynomial features.



Kernels are mathematical functions that calculates the relationships between non-linearly separable data points and maps them into higher dimensions.

However, kernel functions only calculate the high dimensional relationships between the data points as if they were in a higher dimension; they do not actually do the transformation, meaning that the kernel function does not add any features, but we get the same results as if we id.

This trick(calculating the high dimensional relationships between the data points without actually creating or transforming them) is known as the Kernel Trick.

Cost Function :

$$\min_{\boldsymbol{w},\boldsymbol{x}} \; |\boldsymbol{y_i} - \boldsymbol{w_i}\boldsymbol{x_i}| + \sum € + \; |\varepsilon|$$

Sai Subahsish Rout

Connect with me :

Linked in - https://www.linkedin.com/in/sai-subhasish-rout-655707151/

Github - https://github.com/saisubhasish/Concepts

Thank you

Refer my in github for more such contents.