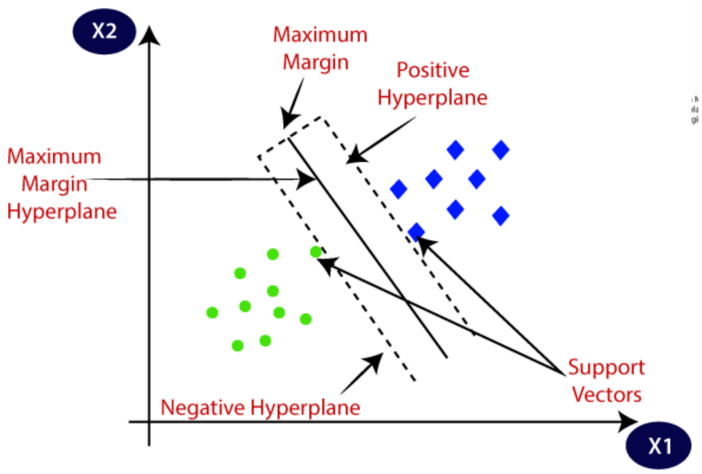
SVM

- Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression.
- Though we say regression problems as well ,it's best suited for classification.
- 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
- 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.
- SVM algorithm can be used for Face detection, image classification, text categorization, etc.



Support Vectors

- Data Points that are closest to the hyperplane are called support vectors.
- Separating line will be defined with the help of these data points.
- Since these vectors support the hyperplane, hence the name support vectors

Hyperplane

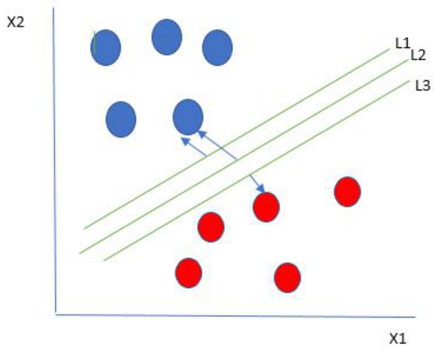
- As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
- The dimension of the hyperplane depends upon the number of features
- If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane

Margin

- It may be defined as the gap between two lines on the closest data points of different classes.
- It can be calculated as the perpendicular distance from the line to the support vectors.
- In SVM large margin is considered a good margin

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.
- So from the above figure, we choose L2.

Computing the margin
In book

Linear SVM
- Linear SVM is used for linearly separable data
- 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
- The classifier used is called as Linear SVM classifier.

Working
- The working of the SVM algorithm can be understood by using an example.
- Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2.
- We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue.

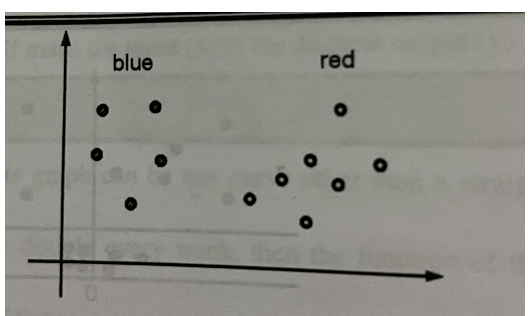
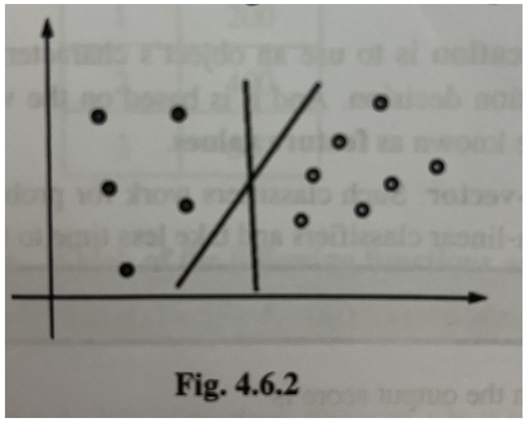
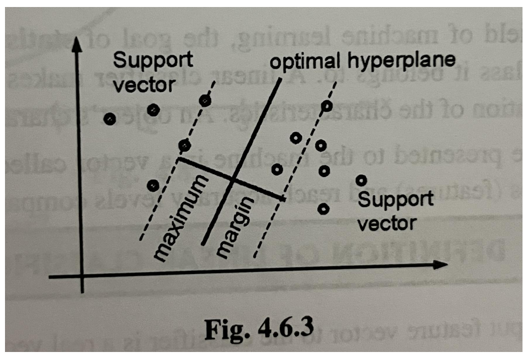


Fig. 4.6.1

- So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes



- SVM algorithm helps to find the best line or decision boundary(hyperplane); SVM algorithm finds the closest point(support vectors) of the lines from both the classes. And the goal of SVM is to maximize the distance between the vectors and the hyperplane known as the margin.
- The hyperplane with maximum margin is called the optimal hyperplane.

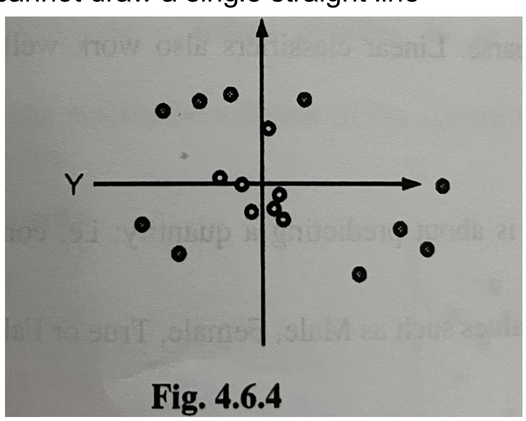


Non linear SVM

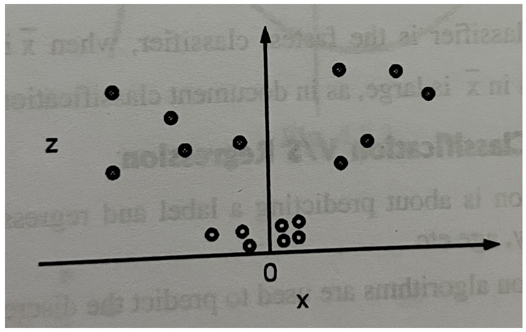
- Non-Linear SVM is used for non-linearly separated data
- If a dataset cannot be classified by using a straight line, then such data is termed as non-linear data
- The classifier used is called Non-linear SVM classifier.

Working

- 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 linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. : z=x2 +y2
- Adding third dimension, sample space looks like



- So now, SVM will divide the datasets into classes in the following way

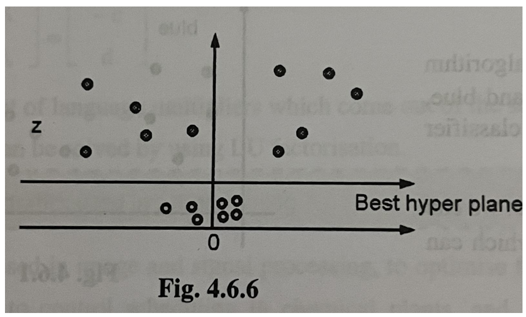


Fig. 4.6.6

- If we convert it in 2d space with z=1, sample space becomes

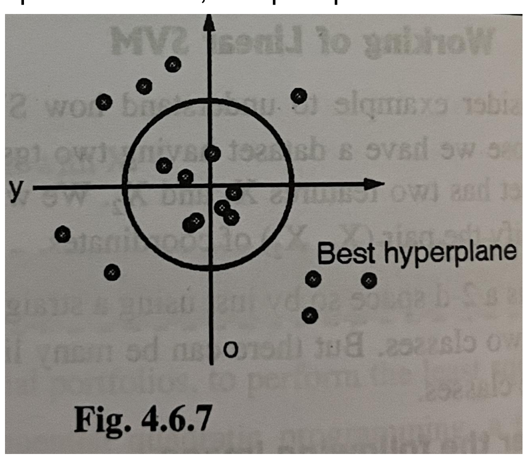


Fig. 4.6.7

Advantages of SVM:

- Effective in high dimensional cases
- Its memory efficient as it uses a subset of training points in the decision function called support vectors
- Different kernel functions can be specified for the decision functions and its possible to specify custom kernels

SVM Kernel

- The SVM kernel is a function that takes low-dimensional input space and transforms it into higher-dimensional space, i.e. it converts non separable problem to separable problem.
- It is mostly useful in non-linear separation problems.
- Simply put the kernel does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined.

Kernel trick

- As it is difficult to classify non linear data, kernel trick helps to give a solution to it
- A Kernel Trick is a simple method where a Non Linear data is projected onto a higher dimension space so as to make it easier to classify the data
- Then it could be linearly divided by a plane.

KT mathematically

In book

Popular kernels

The most important part of using Kernel Trick is to choose the right kernel function because it is very challenging to visualise data in n- dimensional space

- **Fisher Kernel**

Analyses and measures similarity of 2 objects

- **Graph Kernel**

Kernel function that computes inner product on graphs

- **Polynomial Kernel**

Most commonly used with SVM

Purpose of KT

- It offers more efficient and less expensive way to transform data into higher dimensional
- It transforms supervised problem into unsupervised problem
- Reduces computational complexity

Support Vector Regression

- Support Vector Regression (SVR) uses the same principle as SVM, but for regression problems.
- Support Vector Regression is a supervised learning algorithm that is used to predict discrete values.
- In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already been requested from the problem.
- However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated.
- It can be used for both linear and nonlinear regression problems

Terminologies

- Hyperplane:
    - It is a separation line between two data classes in a higher dimension than the actual dimension.
    - In SVR it is defined as the line that helps in predicting the target value.
- Kernel:
    - In SVR the regression is performed at a higher dimension.
    - To do that we need a function that should map the data points into its higher dimension.
    - This function is termed as the kernel.
    - Type of kernel used in SVR is Sigmoidal Kernel, Polynomial Kernel, Gaussian Kernel, etc,
- Boundary Lines:
    - These are the two lines that are drawn around the hyperplane at a distance of ε (epsilon).
    - It is used to create a margin between the data points.
- Support Vector:
    - It is the vector that is used to define the hyperplane or we can say that these are the extreme data points in the dataset which helps in defining the hyperplane.
    - These data points lie close to the boundary.

SVR gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line to fit the data