SUPPORT VECTOR MACHINE

INTRODUCTION:

SVM is one of the most popular supervised machine learning algorithms, Which is used for classification as well as regression problems. However, Primarily it is used for classification problems in 'Machine Learning'.

GOAL:

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/vectores that help in creating the hyperplane. These extreme cases are called 'Support Vectors' and hence algorithm is termed as "Support Vector Machine".

It can easily handle multiple continuous and categorical variables. SVM generates optimal hyperplanes in an iterative manner. Which is used to find a Maximum Marginal Hyperplane (MMH) that best divides the datasets into classes.

SUPPORT VECTORS:

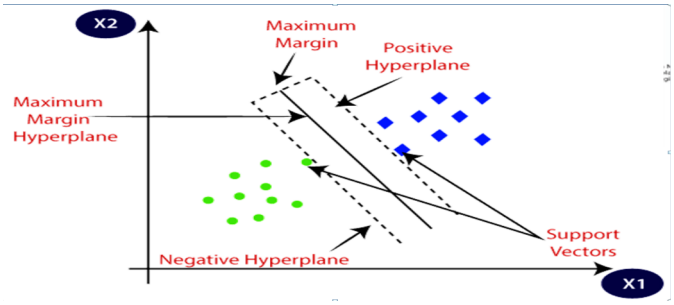
Support vectors are the data points which are closest to the hyperplane. These points will define the separating line better by calculating margins. These points are made relevant to the construct of the classifier.

HYPERPLANE:

A hyperplane is a decision plane which separates between a set of objects having different class membership.

MARGIN:

A margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vector or closed points. If the margin is larger in between the classes then it is considered a good margin, a smaller margin is a bad margin.

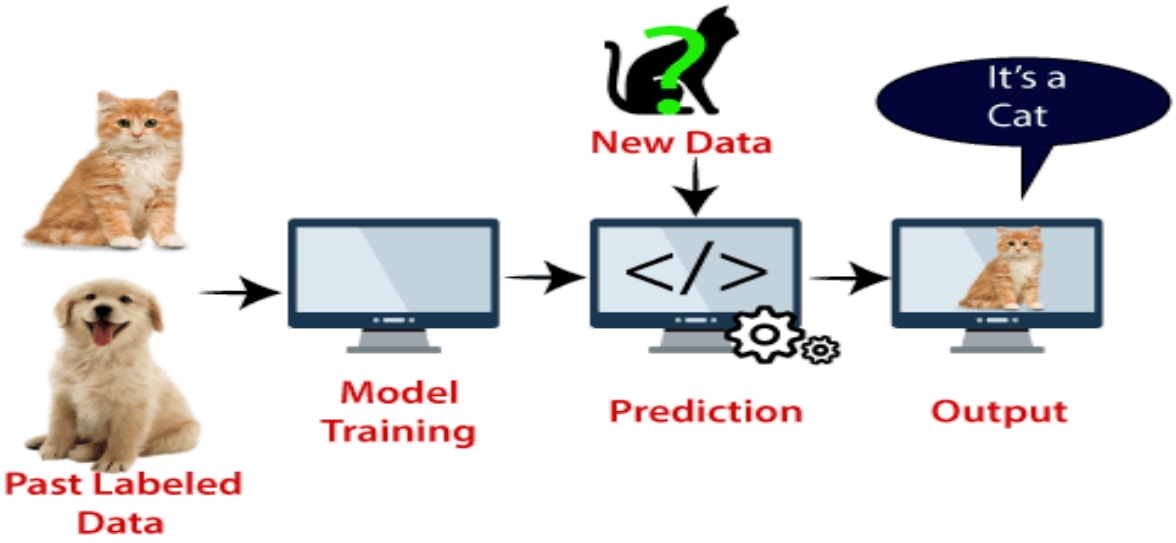


WORKING PRINCIPLE:

EXAMPLE:
Suppose we have a strange cat that also has some features of a dog, So if we want a model that can accurately identify whether it is cat or dog. Such a model can be created by using "SVM Algorithm".

First train our model with lots of images of cats or dogs. So that it can learn about different features of cats and dogs and then we test it with this strange creature. SV creates a decision boundary between these two data and chooses extreme points/vectors, it will see the extreme case of cat and dog.

On the basis of the support vector, It will classify it as a cat.



.

DEALING WITH NON-LINEAR AND INSEPARABLE PLANES:

Some problems can't be solved by using "Linear Hyperplane". So SVM uses a "Kernel Trick" to transform the input space to higher dimensional space. The data plotted on the X-axis and Z-axis (Z = x^2 = y^2.Now you can easily segregate these points using linear separation.

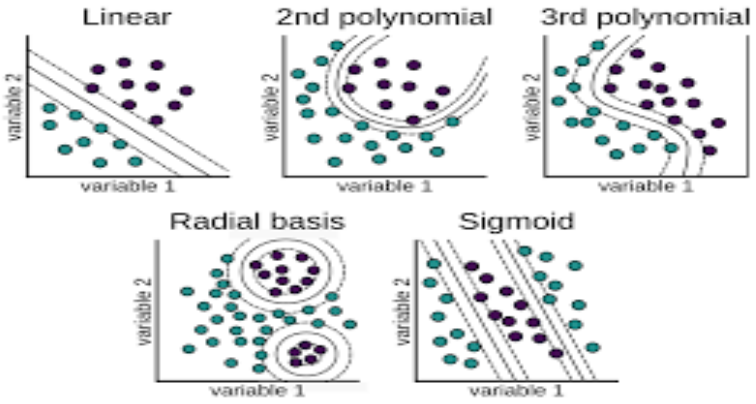
KERNELS:
Kernel is used to transform the low-dimensional space into high dimensional space or required form. This technique is called "Kernel Trick" converts non-separable problems into separable problems. Kernels tricks helps you to build a more accurate classifier

POLYNOMIAL KERNEL:

It is a more generalized form of the linear kernel. The polynomial kernel can distinguish curved or nonlinear input space.

$$K(x,x1) = 1+sum(x*x1)^d$$

Where 'd' is the degree of the polynomial d=1 is similar to the linear transformation.



RADIAL BASIS FUNCTION:

RBF is a popular kernel function commonly used in Support Vector Machine classification. RBF can map an input space into infinite dimensional space.

$$K(x,x1) = \exp(-gamma * sum(x-x1)^2)$$

Here gamma is a parameter which ranges from 0 to 1. A higher value of gamma will perfectly fit the training dataset but which causes over-fitting. Gamma =0.1 is considered to be a good default value.

TUNING HYPERPARAMETERS:

---

1 KERNELS
2 REGULARIZATION (C)
3. GAMMA

REGULARIZATION:

In python sklearn ‘C’ parameter is used to maintain regularization.C is the penalty parameter, which represents misclassification or error term. It tells the SV optimization how much error is bearable. This is how you can trade-off between decision boundary and misclassification term.

GAMMA:

Low values of gamma will loosely fit the training dataset, whereas higher values of gamma will exactly fit the training dataset. Which causes overfitting .

Other Words, the low value of gamma considers only nearby points in calculating the separation line, while the value of gamma considers all the data points in the calculation of the separation line.

PROS AND CONS:

PROS: SVM offers good accuracy and performs faster prediction compared to naive bayes. SVM works very well with a clear margin of separation and with high dimensional space.

CONS: SVM is not suitable for large data sets because of high training time. It works poorly with overlapping classes and is also sensitive to the type of kernel used.