SUPPORT VECTOR MACHINES

What is SVM?

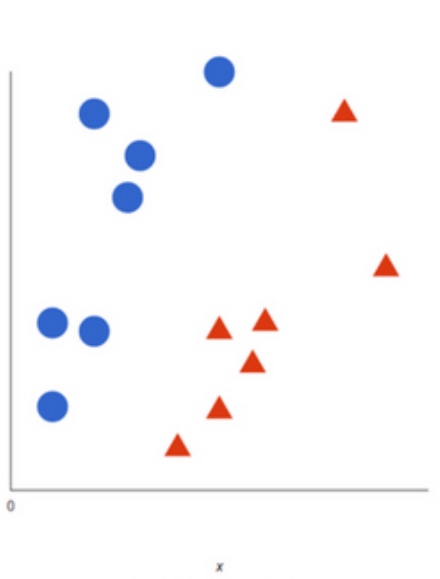
Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression. However, it is mostly used in classification problems.
In this algorithm, we plot each data item as a point in n-dimensional space (where n is the number of features) with the value of each feature being the value of a particular coordinate.

How is the data classified?

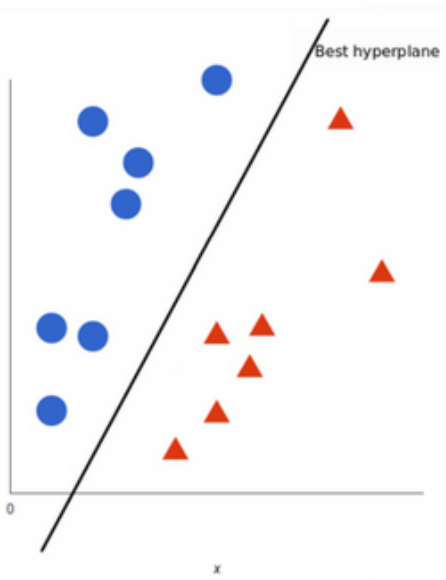
We perform classification by finding the hyperplane that differentiates the two classes very well. In other words the algorithm outputs an optimal hyperplane which categorizes new examples.

What is a optimal Hyper-Plane?

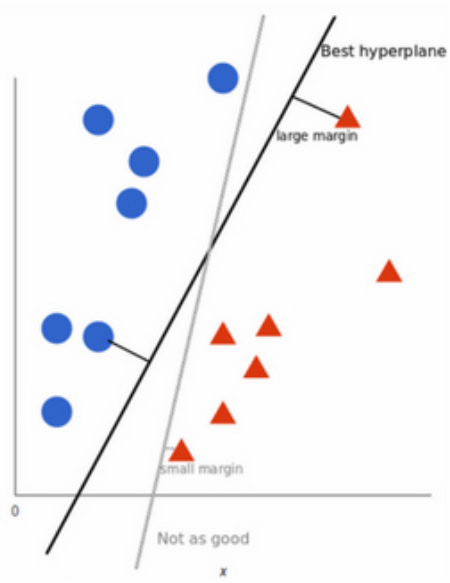
For SVM, it's the one that maximizes the margins from both tags. In other words: the hyperplane whose distance to the nearest element of each tag is the largest.



Our labeled data

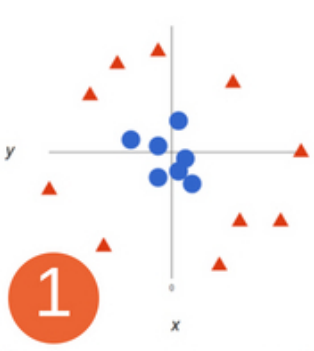


In 2D, the best hyper-plane is simply a line



Not all hyper-planes are created equal

Nonlinear data

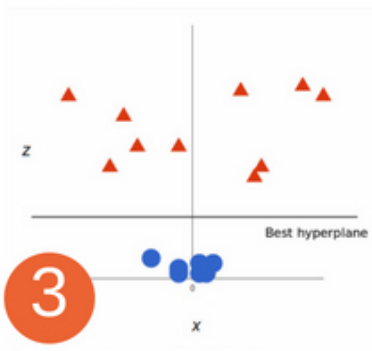


1

In above case we cannot draw a linear boundary.We will now add a third dimension. We create a new z dimension, and we rule that it be calculated a certain way that is convenient for us: $z = x^2 + y^2$ (equation for a circle).

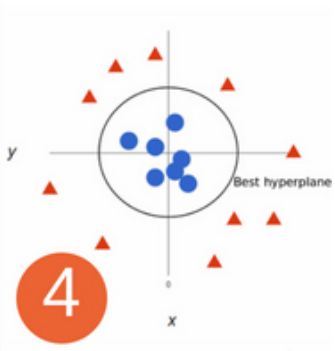
2

This will give us a three-dimensional space. From a different perspective, the data is now in two linearly separated groups. All values for z would be always positive because z is the squared sum of both x and y.



3

Since we are in three dimensions now, the hyper-plane is a plane parallel to the x axis at a certain z. We choose the hyperplane, one that maximizes the margins from both of the Classes.



4

Now we map back to 2 dimensions.Our decision boundary is a circumference which separates both tags using SVM. We got a circle as a hyper-plane.

TUNING PARAMETERS

KERNEL

The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays role. Polynomial and exponential kernels calculates separation line in higher dimension. This is called kernel trick.

REGULARIZATION

For large values of this parameter, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. Conversely, a very small value of it will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.

GAMMA

The gamma parameter defines how far the influence of a single training set reaches. With low gamma, points far away from the possible separation line are considered in calculation for the separation line. Where as high gamma means the points close to possible line are considered in calculation.

MARGIN

A margin is a separation of line to the closest class points.
A good margin is one where this separation is larger for both the classes. A good margin allows the points to be in their respective classes without crossing to other class.