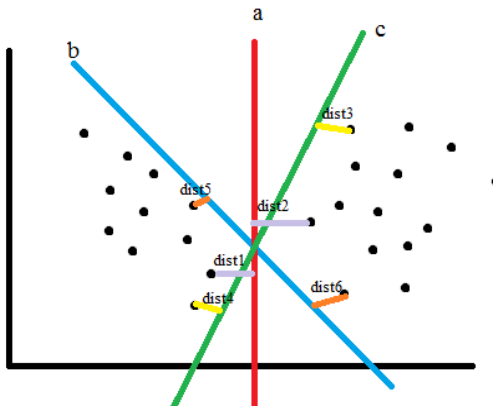
SVM

Introduction

The SVM finds the line or the hyper plan between two classes which maximizes the distance to the nearest point. This distance is called Margin. For example: if we have two classes looks like this, and we have three lines. Which line will be the best line to separate to classes, a, b, or c according to the SVM algorithm?



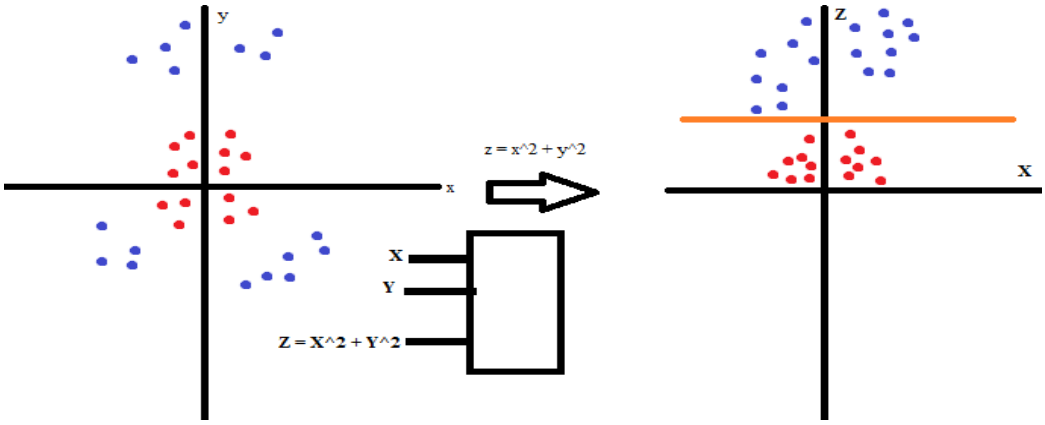
Of course, the three lines separates the dataset, but the line (a) is the best since it has the largest margin. Line (a) has the largest margin (dist1, dist2) to the nearest points. Compared to line (b) which it is margin to the two classes are (dist5, dist6) and line (c) which its margin to the two classes are (dist3, dist4).

Sometimes, we need a kernel to separate between classes in more efficient way.

What is the kernel?

The kernel is the function which is used to convert the problem from nonlinear separable problem into linear separable problem by using a function, this function is called a kernel.

For example, if we have the following two classes on the left side, they can't be linearly separable, but after adding the function or the kernel which will be the input to the model, it will be linearly separable as in the right side.



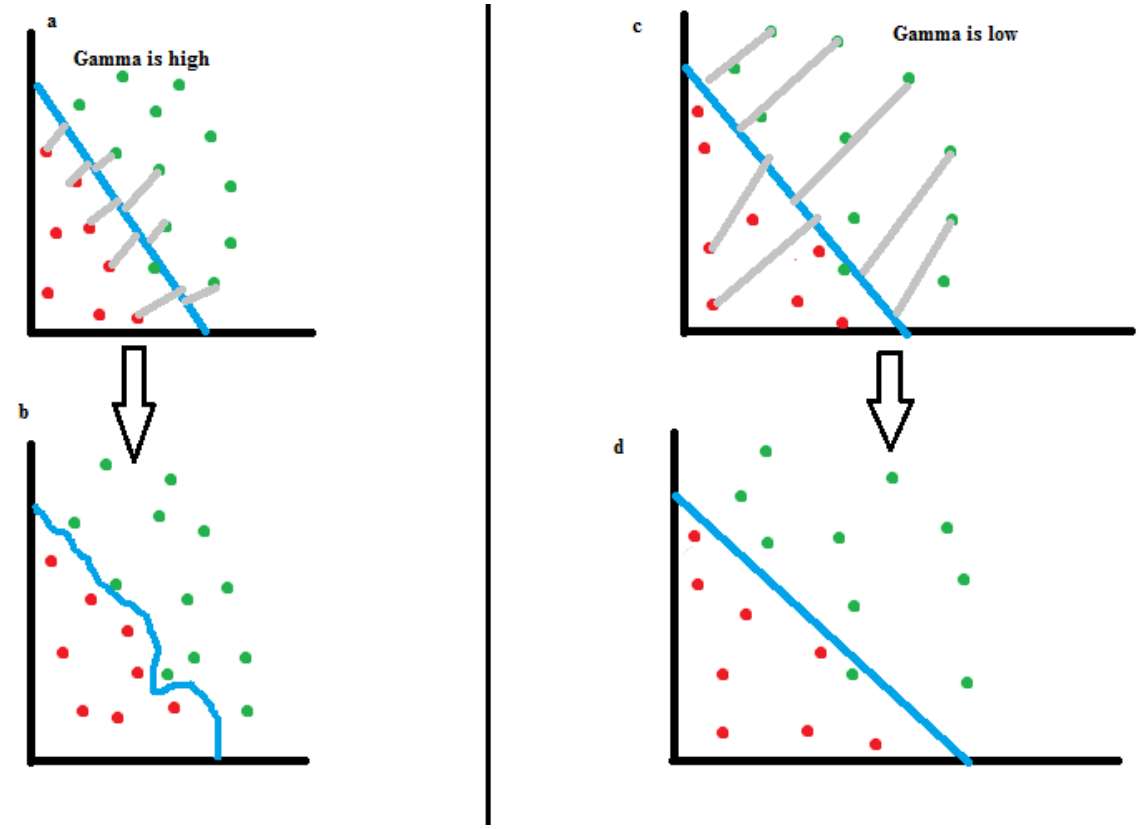
So, the kernel will change the input space into larger input space. If you take the solution and go back into the original space, then the separator will be a non-linear.

The kernel gives the SVM the powerful to solve complex problems, since it enable the SVM to draw the non-linear separator between classes.

SVM parameters:

- Kernel: which is used to solve complex/ non-linear classifications.
- C parameter: which is used to control the tradeoff between smooth decision boundary and classifying training points correctly. If the c parameter has a large value, then it will classify the training points correctly. But if the c value is a small value, then it gives us a smooth decision boundary.
- Gamma parameter: gamma defines if the decision boundary will be dependent on the near points or far points in the dataset. To understand more: here in the below figure if the gamma has high value, the near points have a high weight and influence the decision boundary (see Figure a), so it pulls the decision boundary tries to fit the training points (see Figure b). But if the gamma has a low value, then the far points have a high weight and influence the decision boundary (see Figure c), so the decision boundary looks like straight.

  Note: All of above parameters can led to overfitting. Overfitting means that the decision boundary tries to fit the training points literally. This cause the boundary to fail in testing data. So we should avoid overfitting during the mode training by tuning these parameters.



SVM Pros and Cons:

Pros:

- The SVM is good enough to find the line which maximizes the margin and tolerate or ignore the outlier.
- Also, it is effective in high dimensions.
- Also, it is versatile because of its kernels.

Cons:

- SVM fails in large dataset, since its training time is cubic in the size of data.
- They don't work in lots of noise when the classes overlapping. So, in this case the Naïve Bayes will be better.

Note: If there are large dataset and large features, there might be an overfitting. Also, it will be slow.

Note: I have a surprise for you, you can try both of c, kernel parameters on the SVM and emails data from the following link

https://github.com/MarwaMohammad/Machine_Learning_course/blob/main/svm/svm_author_id.py