Support Vector Machine (SVM)

What is a Support Vector Machine?

It's a powerful supervised machine learning algorithm that works best on smaller datasets but on complex ones. SVM can be used for both classification and regression problems.

SVM try to find a hyperplane that best separates two classes by finding the maximum margin between the hyperplanes that means maximum distances between the two classes.

Types of Support Vector Machine

Linear SVM

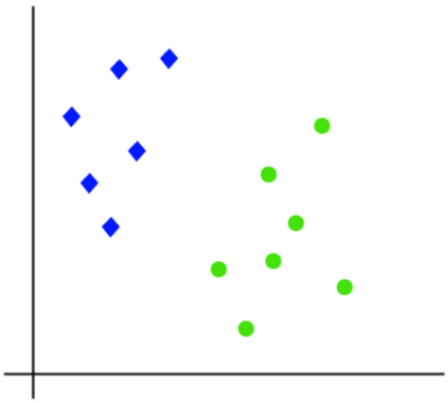
We have linear Support Vector Machine case when the data is perfectly linearly separable only then we can use Linear SVM. It means that the data points can be classified into two different classes by using a single straight line (example → 2D)

Non-Linear SVM

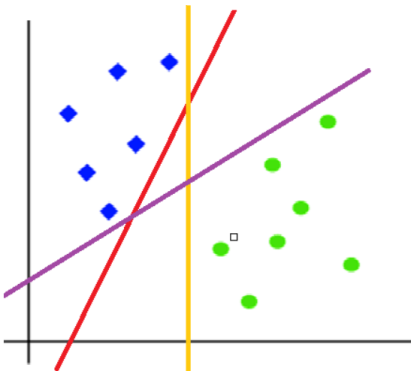
We have Non-Linear Support Vector Machine when the data is not linearly, that means when the data points cannot be separated into two different classes by using a straight line (if 2D) then we can use some advanced techniques like kernel tricks to classify them.

How does SVM work?

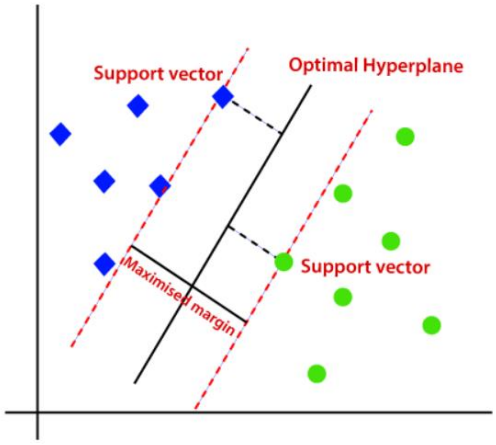
Suppose we have a dataset that has two classes (green and blue). We want to classify that the new data point as either blue or green.



Since we are plotting the data points in a 2-dimensional graph we call this decision boundary a **straight line** but if we have more dimensions, we call this decision boundary a **"hyperplane"**



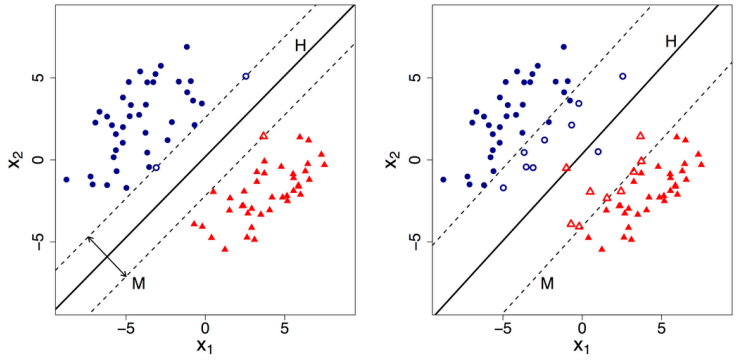
The best hyperplane is that plane that has the maximum distance from both the classes. This is done by finding different hyperplanes which classify the labels in the best way then it will choose the one which is farthest from the data points or the one which has a maximum margin.



- We can clearly see that SVM tries to maximize the margins and thus called Maximum Margin Classifier.
- The Support Vectors will have values exactly either {+1, -1}.
- The more negative the values are for the Green data points the better it is for classification.
- The more positive the values are for the Red data points the better it is for classification

How to choose the Correct SVM

I'll try to explain with an example that I found online

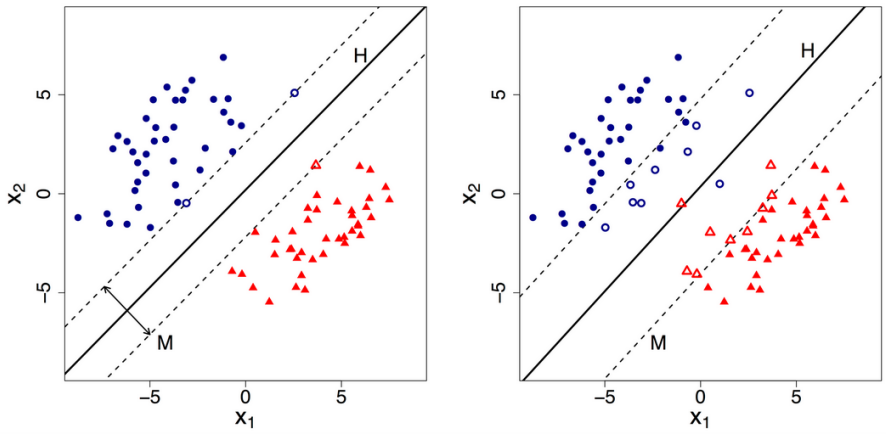


Suppose we are given 2 Hyperplane one with 100% accuracy (HP1) on the left side and another with >90% accuracy (HP2) on the right side.

Most of us would pick the *HP2* thinking that it because of the maximum margin.

Support Vector Machine would choose the *HP1* though it has a narrow margin. Because though HP2 has maximum margin but it is going against the constrain that: ***each data point must lie on the correct side of the margin and there should be no misclassification.***

Difference between hard and soft SVM



In this case HP1 (figure 1) is a **Hard SVM** and the HP2 (figure 2) is a **Soft SVM.**

Hard margin SVM → does not allow any misclassification to happen.

In the previous example our data is non-separable/ nonlinear then the Hard margin SVM will not return any hyperplane as it will not be able to separate the data.

Soft margin SVM → allows some misclassification to happen by relaxing the hard constraints of Support Vector Machine.

Soft margin SVM is implemented with the help of the **Regularization parameter (C).**

Regularization parameter (C): It tells us how much misclassification we want to avoid.

- *Hard margin SVM generally has large values of C.*
- *Soft margin SVM generally has small values of C.*

Relation between Regularization parameter (C) and SVM

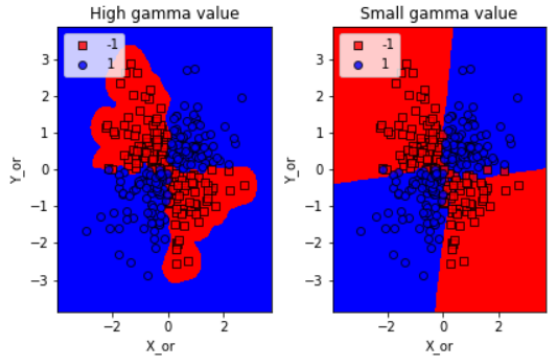
- As the value of C increases the margin decreases thus Hard SVM.
- If the values of C are very small the margin increases thus Soft SVM.

- Large value of C can cause overfitting therefore we need to select the correct value using Hyperparameter Tuning.

Other parameters of SVM

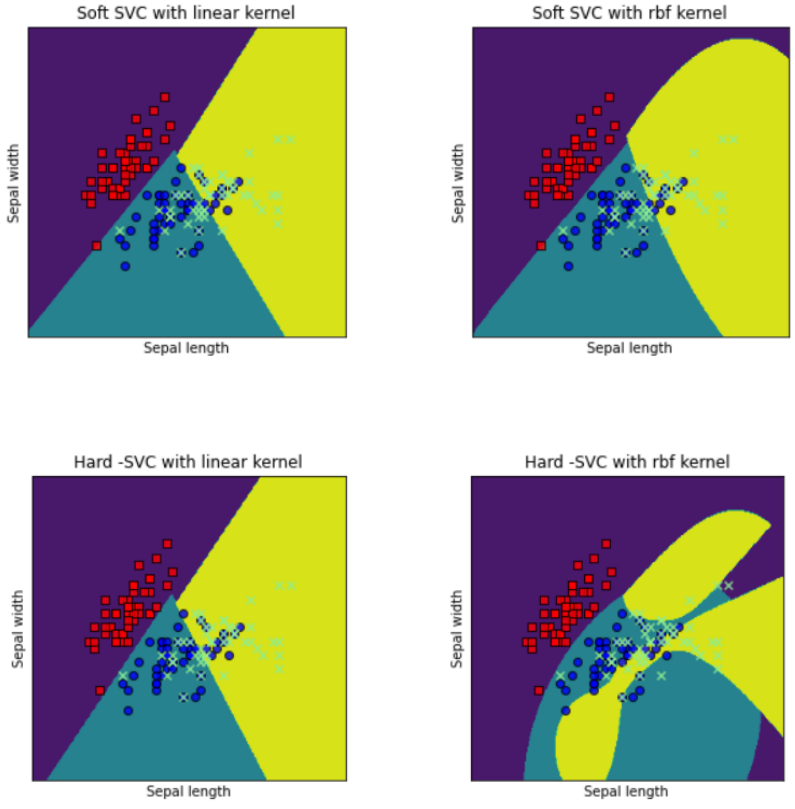
Other significant parameters of Support Vector Machine are the **Gamma** values. It tells us how much will be the influence of the individual data points on the decision boundary.

- **Large Gamma**: Fewer data points will influence the decision boundary. Therefore, decision boundary becomes non-linear leading to overfitting
- **Small Gamma**: More data points will influence the decision boundary. Therefore, the decision boundary is more generic.

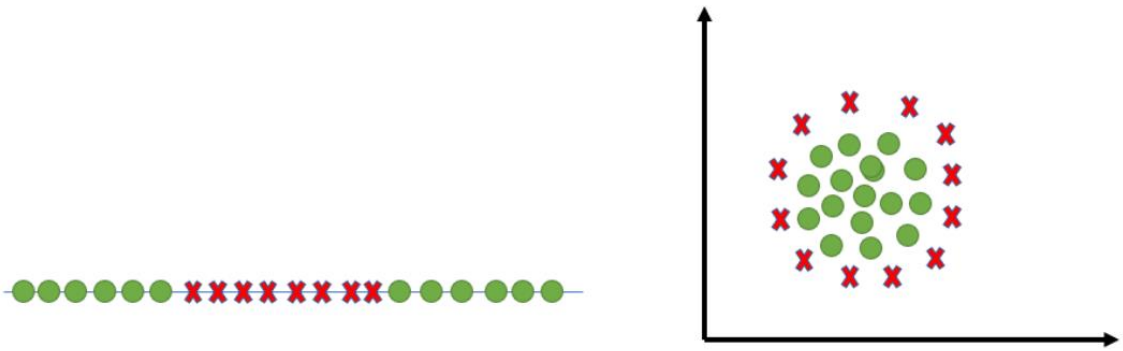


Kernel trick

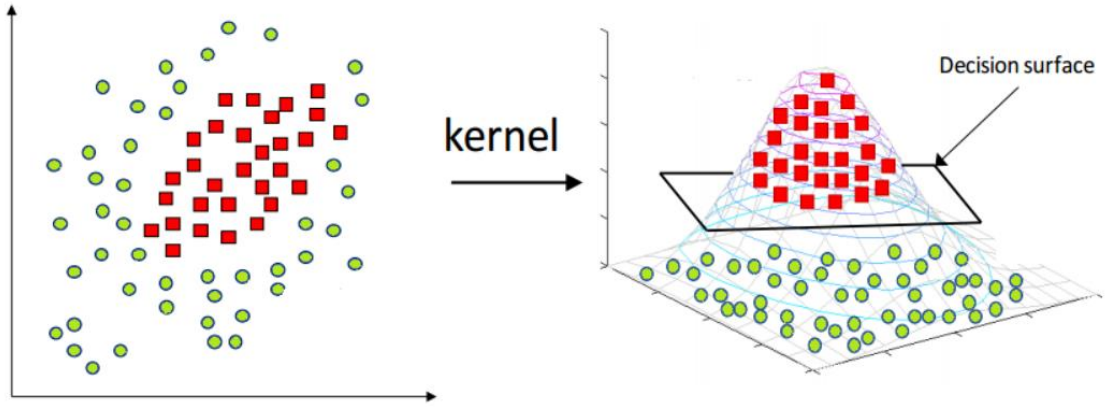
Support Vector Machine deals with nonlinear data by transforming it into a higher dimension where it is linearly separable. Support Vector Machine does so by using different values of Kernel.



We use “Kernel Trick” which makes it easier to classifies the points.



Here we see we cannot draw a single line or say hyperplane which can classify the points correctly. So what we do is try converting this lower dimension space to a higher dimension space using some quadratic functions which will allow us to find a decision boundary that clearly divides the data points.



Type of kernel:

- Polynomial Kernel
- Sigmoid kernel
- RBF kernel
- Bessel function kernel
- Anova Kernel

Advantages

1. SVM works better when the data is Linear
2. It is more effective in high dimensions
3. With the help of the kernel trick, we can solve any complex problem
4. SVM is not sensitive to outliers
5. Can help us with Image classification

Disadvantages

1. Choosing a good kernel is not easy
2. It doesn't show good results on a big dataset
3. The SVM hyperparameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact