Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer:

SoftMargin Classifier	Maximum Margin Classifier
Allows errors to be made while fitting the model to training dataset.	On a contrary allows zero errors
Produce more Generalisable model on new data	Produce less Generalisable model ,but perform well on training dataset leads to overfit
It is proposed by Vapnik to solve this problem by slacked variables	It works only on data which is completely linearly separable without any errors like outliers or noise .

Question 2

What does the slack variable Epsilon (ε) represent?

Answer:

Usage of Slack variable is to control misclassifications.
It represents where the observation is location relatively to the margin and hyperplane.
The slack variable is represented by (ε)

ε	
0	For an observation which are at a distance of more than M which is safe distance from the hyperplane
Between 0 and 1	The data point is correctly classified but violates the margin
Greater than 1	If the data point is incorrectly classified ,it violates the hyperplane

The correct Classification is better when the value of slack is lower or equals to 0.

Question 3

How do you measure the cost function in SVM? What does the value of C signify?

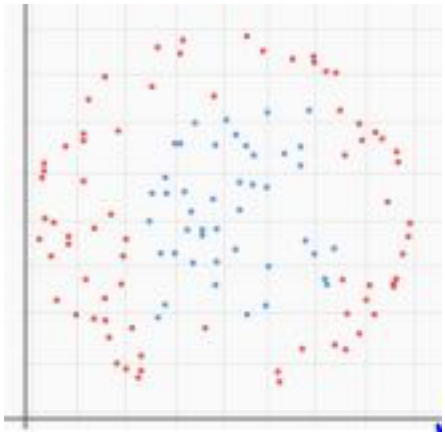
Answer:

The sum of all the epsilons(ε) of both the hyperplanes and least sum of all the epsilons will give us the best one. The summation of all the epsilons of each data points is the cost and it is denoted by C, i.e.

$$\sum \varepsilon i \leq C$$

- The C parameter helps the SVM optimisation
- how much you want to avoid misclassifying each training example.
  - For large values of C, the optimisation will choose a smaller-margin hyperplane ,if that hyperplane does a better job of getting all the training points classified correctly.
  - In Contrast, a very small value of C will cause the optimiser to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.
  - For very tiny values of C, you should get misclassified examples, often even if your training data is linearly separable.

Question 4



Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

Answer:

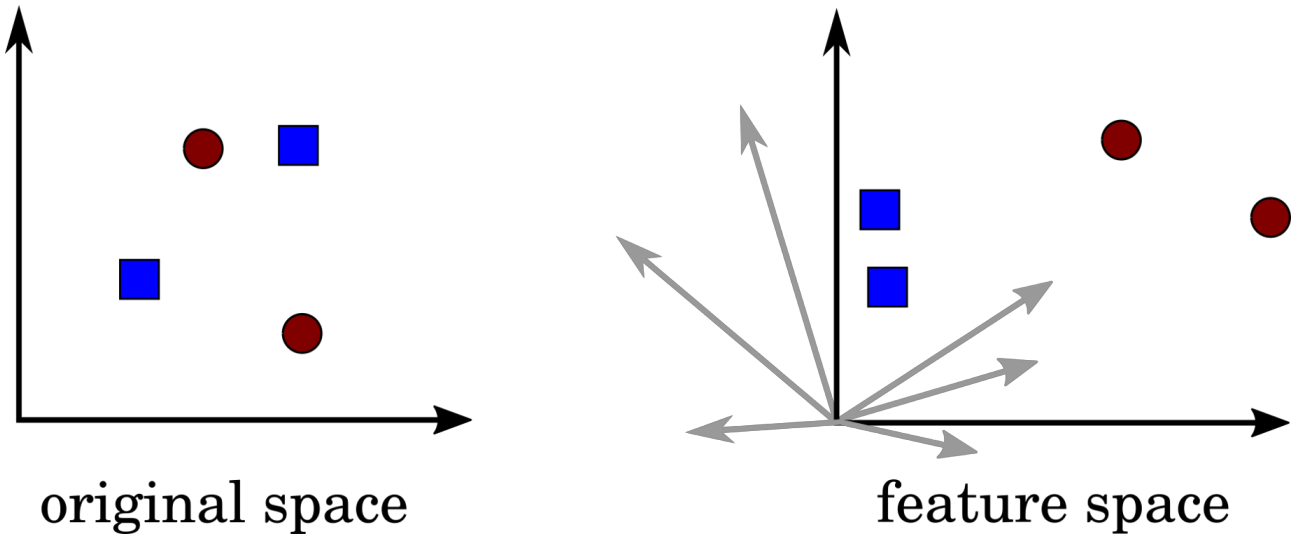
- We can classify data with RBF kernel using SVM.
- We can further tune hyperparameter to optimize separation & optimized accuracy & precision

Question 5

What do you mean by feature transformation?

Answer:

There is always such kind of data which seems to be inseparable and we assume that the data is not linearly separate .However , it is incorrect to assume that because with SVMs ,it is easy to classify it. Data is not always supposed to be separated on the two dimension.
If the initial data is in 2D , we can easily transform it before applying SVM to it .



Transforming to Higher Dimensional data

For example, we can do what is called a polynomial mapping by applying the function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ defined by:

$$\phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2)$$

This is feature transformation and the downside of it is ,it require lot of computational cost ,thats why Kernels are used by replacing the dot product of two examples by a kernel function. It can be linear, polynomial eg RBF etc