Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Answer:

- Maximum Margin Classifier does not allow for any misclassification of data points and hence in certain circumstances may lead to overfitting
- In order to determine the best classifier, Soft Margin Classifier may allow certain points to be misclassified and therefore avoid overfitting.

Question 2

What does the slack variable Epsilon (ε) represent?

Answer:

A slack (epsilon) variable is used to control misclassifications. It indicates the permissible error in SVM. Each data point has a slack value ranging from 0 to infinity associated with it, depending on where the point is located.

- If the value of slack variable (epsilon) is 0, it means that the point has been correctly classified and it lies outside the margin
- If the value of slack variable (epsilon) is between 0 and 1, it means that the point has been correctly classified and it lies inside the margin
- If the value of slack variable (epsilon) is greater than 1, the data point has been wrongly classified.

Therefore, lower values of slack are better than higher values.

Question 3

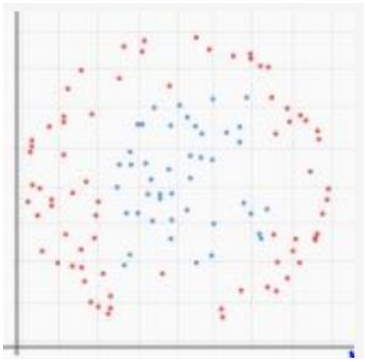
How do you measure the cost function in SVM? What does the value of C signify?

Answer:

C denotes the cost function. The cost function C is the sum of the slack variables for all the data points.

- If the value of C is large, the slack variables can be large. This means that the model allows a large number of data points to be misclassified or to violate the margin. In this case the model has a high bias, i.e. the model is flexible, more generalisable, and less likely to overfit.
- If the value of C is small, the individual slack variables are forced to be small, thereby the margin is narrow and there are very few misclassifications. It means that we do not allow many data points to fall on the wrong side of the margin or the hyperplane. In this case the model has a high variance, i.e. the model is less flexible, less generalisable, and more likely to overfit.

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:

The dataset in the above figure cannot be classified into two classes (red and blue) using a linear hyperplane. Hence we will need to do the following:

- Transform the data from the 2-D attribute space to a linear data in 3-D feature space.
- Since the data now is linear, we can apply Maximum Margin Classification or Soft Margin Classification as applicable.

Question 5

What do you mean by feature transformation?

Answer

The process of transforming the original attributes into a new feature space is called Feature Transformation. An increase in the number of attributes will result in an exponential increase in the number of dimensions in the transformed feature space.

For e.g., if we have four variables in the data set, then considering a polynomial transformation with degree 2, we would end up making 15 features in the new feature space as shown in the figure below.

$$a_1X^2 + a_2Y^2 + a_3Z^2 + a_4W^2 + a_5XY + a_6YZ + a_7ZW + a_8WX + a_9WY + a_{10}ZX + a_{11}X + a_{12}Y + a_{13}Z + a_{14}W + C = 0$$