SVM Assignment

Question 1

How is Soft Margin Classifier different from Maximum Margin Classifier?

Solution

The maximum margin classifier attempts to find the separating plane between classes by correctly classifying all the points. Therefore, even a small change in the points the separating plane changes drastically and it is an unstable solution.

In Soft margin classifier, instead of using a hard margin we allow the algorithm to miss classify some points by using a hyper parameter. Therefore, small perturbations in the data will no change the hyper plane and provides a more stable and correct solution which doesn't overfit the data.

Question 2

What does the slack variable Epsilon (ε) represent?

Solution:

The slack variable epsilon represents the classification status of the data point. If epsilon is 0 for a point it means that the point is correctly classified wrt to both hyper plane and the margin. When epsilon lies between 0 and 1, it means that the point is on the correct side of the hyper plane but on the wrong side of the margin and if epsilon is greater than 1 then it means that the point falls on the wrong side of the hyper plane and it is misclassified.

Question 3

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

Solution:

The below is the cost function of the SVM, we try to find the optimal parameters given he maximum error hyper parameter. It tries to find the hyper plane for which all the points are at least at a distance of M from the plane and a relaxation is provided on number of points that can either fall on the wrong side of the margin or wrong side of the boundary. The slack variable epsilon takes care of this which was explained in the previous question.

C which is the sum of all epsilons (slack variables) tells how many misclassifications are allowed for the algorithm to build a model from the data. It helps in building a more stable and robust model and also avoiding overfitting.

$$\sum y_i (\bar{w} \cdot \bar{x}_i) \geq M(1 - \xi_i)$$

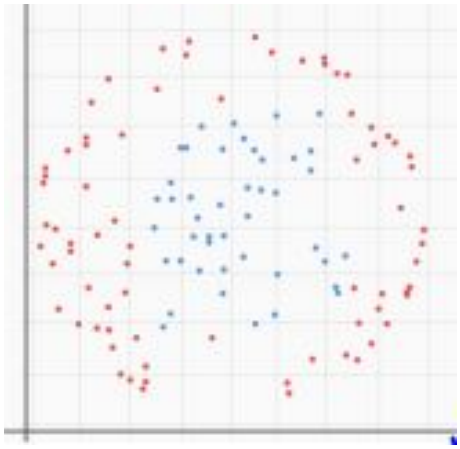
$y_i \Rightarrow$ class of the point i

$\bar{w} \Rightarrow$ coefficients of the hyperplane

$M \Rightarrow$ minimum of Maximum distance (Margin)

$\xi_i \Rightarrow$ slack variable.

Question 4



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

Solution:

We can clearly see that the given data cannot separated into two classes by hyperplane and we can see that the boundary between classes is circular. Therefore, while building SVM model we will use a 'rdf' kernel which will transform that data into a new feature space such where the data can be separated by a hyperplane and find the corresponding hyperplane.

Question 5

What do you mean by feature transformation?

Solution:

Sometimes the output variable may not be linearly related to the existing features and hence you cannot use linear methods for such cases. Therefore, the features are transformed so that using some functions so that the newly transformed features are linearly related to the output variable and hence the whole problem can be transformed using a linear method in the newly transformed feature space.

The process of creating new features from existing features is called as feature transformation.