Assignment - Support Vector Machines

Problem Statement - II

Question 1
How is Soft Margin Classifier different from Maximum Margin Classifier?
Answer:

Soft Margin Classifier: The soft margin classifier permits the misclassification of certain points to ensure that the model works fine on unseen data.

Maximum Margin Classifier: The line or the hyperplane which maintains the maximum equal distance from the line or hyperplane is called as Maximum Margin Classifier. Maximum Margin Classifier performs well on training data but fails to perform on testing data.

Question 2
What does the slack variable Epsilon (ε) represent?
Answer:

To identify the support vector classifier and overcome misclassification issue slack variable Epsilon (ε) plays a vital role. The range of slack variable Epsilon is between 0 and +infinity. It is good to have a lower value of slack variable.

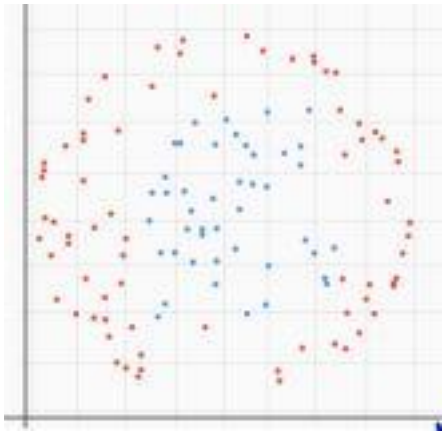
- When ε = 0 it means no misclassification and the observations are on correct side of margin.
- When ε > 1 it means observation is on the incorrect side of margin.
- When ε within 0 and 1 it means observation classifies correctly but violates the margin.

Question 3
How do you measure the cost function in SVM? What does the value of C signify?
Answer:

SVM algorithm tries to build a hyperplane such that the it has maximum distance between the point of class one to other. Summing all the epsilons (ϵ) of hyperplane and selecting the one that gives least sum of epsilons(ϵ). The value of C (cost) signifies the summation of all epsilon of each data point.

- If the value of C is large, the epsilons (ϵ) can be large signifying higher number of data points can be misclassified making the model flexible and avoid overfit.
- If the value of C is small, the epsilons (ϵ) can be small signifying no data points can be misclassified making the model less flexible and prone 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:

To classify the red and blue data points in two classes we need to use kernel trick which enables the linear SVM model to segregate nonlinearly separable data points. We can convert nonlinear data to linear ones by applying functions to original attributes, so we convert original space to transformed space. Transforming the original space to feature space is known as feature transformation. Feature transformation increases the number of features making the model computationally expensive. Using the Linear kernel, Polynomial Kernel and Radial basis function (RBF) kernel we can find the best hyperplane separating two classes.

Question 5
What do you mean by feature transformation?
Answer:

It is important to convert nonlinear boundaries to linear boundaries by applying specific functions to the original attributes. The original space is called the attribute space and the transformed space is called the feature space. Transforming the original space to feature space is known as feature transformation.