SVM ASSIGNMENT

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

Ans: Maximal marginal classifier (i.e hyperplane) perfectly separates the two classes, it has a rather limited applicability: it cannot classify data points if they are partially intermingled.

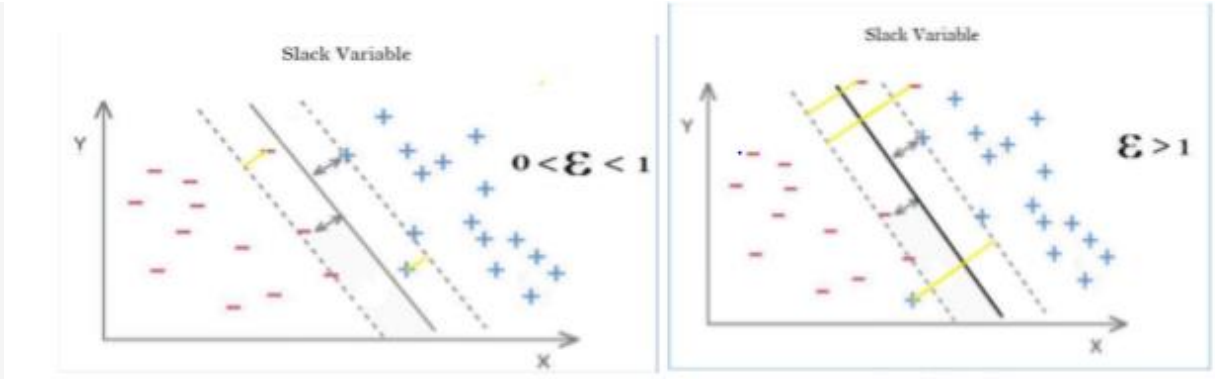
Where as the Soft Margin Classifier overcomes the drawbacks of the Maximal Margin Classifier by allowing certain points to be misclassified. We can control the amount of misclassifications using the cost of misclassification 'C', where C is the maximum value of the summation of the slack variable epsilon i.e $(\sum \epsilon i \leq C)$

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

Ans: Slack variance represent an unused quality of resource, it is added to less than or equal to type constraint to get an equality constraint.

The value of Epsilon (ε) defines a margin of tolerance where no penalty is given toerrors. The larger Epsilon (ε) is, the larger errors you admit in the solution, by contrast, if Epsilon (ε)→0, every error is penalized.

- for $0 < \xi \leq 1$ point is between margin and correct side of hyperplane. This is a margin violation
- for $\xi > 1$ point is misclassified



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

Ans: There is no rule of thumb to choose a C value, it totally depends on your testing data. The only option is to try on bunch of different values and choose the value which gives you the lowest misclassification rate on testing data.

Using **gridserachCV**, can provide a list of different value parameter and can imply which value is best.

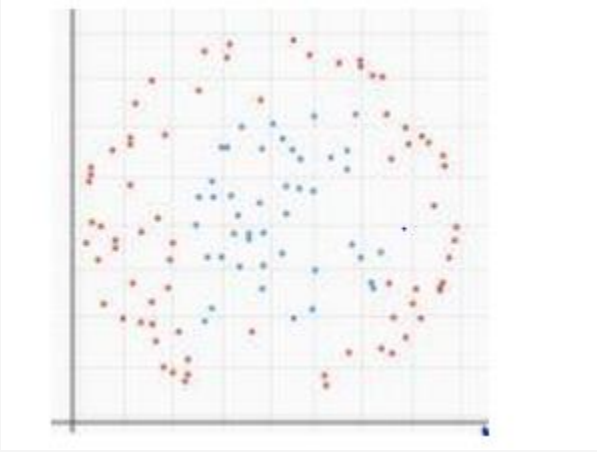
TheC parameter tells the SVM optimization how much you want to avoid misclassifying each training example.

Larger value of parameter C => small margin

Small value of parameter C =>Large margin

Question 4 Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

Ans: For given dataset where ref and blue points represent the two classes, to classify the data we use SVM :



First, we have to find a line that separates two classes.

A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal here is to find the line passing as far as possible from all points.

The operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples.

Question 5 What do you mean by feature transformation?

Ans: Feature Transformation (FT) refers to family of algorithms that create new features using the existing features. These new features may not have same interpretation as the original features, but they may have more discriminatory power in a different space than the original space. This can also be used in feature reduction. FT may happen in many ways, by simple/linear combination of original features or using non-linear functions.

Some common technique of FT are:

- Scaling or normalizing feature within a range, say 0 to 1.
- Principal component analysis and its variants.
- Random projection.
- Neural networks.
- SVM also transform features internally.
- Transforming categorical features to numerical.

Pallawi Jyoti