PART II-- SUBJECTIVE QUESTIONS

Q1) How is soft margin different from maximum margin classifier?

Ans 1) -a) Soft Margin Classifier-- allows some amount of misclassification.

Maximum margin classifier-- doesn't allow any misclassification.

b) Soft Margin Classifier—produces more generalizable model with some errors that work fine with new unseen data.

Maximum margin classifier—the model is produced is not so much generalizable and doesn't work well with new unseen data.

c) Soft Margin Classifier—solves the problem of outliers or noise by the introduction of slack variable.

Maximum margin classifier—works only when the data is linearly separable without any errors.

Q2) What does slack variable epsilon represents.?

Ans 2) It is used to control misclassification. It tells you what the distance of observation from the margin is.

For the points which are at a distance >M, i.e. at a safe distance from the hyperplane, **ϵ**, the slack variable =0

For points which are correctly classified but falls inside the margin, slack variable, **0<ϵ<1.**

For points incorrectly classified, i.e. which violates the hyperplane, slack variable, **ϵ>1.**

Q3) How does you measure the cost function in SVM? what does the value of C signify.?

Ans 3) C or the cost function--It represents the summation of the epsilons **ϵ** of all the data points , ∑ **ϵi≤C**

If C is large, the slack variables (epsilons(ϵ)) can be large, i.e. you allow a larger number of data points to be misclassified or violate the margin; and if C is small, you force the individual slack variables to be small, i.e. you do not allow many data points to fall on the wrong side of the margin or the hyperplane.

Q4) Given the above dataset where red and blue points represent the two classes, how will you use SVM to classify the data?

Ans4) This type of nonlinear data can be solved using **rbf kernel** further we can tune hyperparameters to get desired accuracy and other metrics.

Q5) What do you mean by feature transformation?

Ans 5) The process of transforming the original attributes to

New feature space is called 'Feature Transformation', as the number of attributes increases, there is an exponential increase in the number of dimensions in the transformed feature space. The feature transformation is computationally costly, so kernel is used for the same.

.