ML Lab 03 Report

Name: Raman Kumar
Course: Machine Learning
Reg. No: <u>BL.EN.P2DSC22009</u>

Q.1 Do you think the classes you have in your dataset are well separated? Justify your answer.

Yes, the classes in my dataset are well separated. There is a little error in the accuracy score but it is minor so it can be neglected. Still the accuracy score I'm getting after classifying the dataset using k-NN model it is in between 0.93 to 0.98.
Also I'm getting the k-NN classifier score for k=1 exactly 1.0 so I can say the they are well separated.

Q.2 Do you think distance between class centroids (mean of vectors in a class) is a good enough measure to test for class separability? Justify your answers. Use diagrams to illustrate your arguments.

Yes, by using good enough distance measure metric we can simply classify or separate the classes. By initiating the centroids we can simply assign the data points to nearest centroid. Then we can calculate the proximity measure by using Euclidean, Manhattan or Jaccard similarity. That way the association of data points to centroids will lead to boundary or partitions of classes.

Q.3 Explain the behavior of the kNN classifier with increase in value of k. Explain the scenarios of over-fitting and under-fitting in kNN classifier.

- There is no best method to get the best value of 'k' in kNN classifier. We can only find that by trial and test using train-validation-test split or cross validations. But the behaviour of kNN classifier will change a little with the increase of value of k.
- The classification boundary will be a little smoother if the value of k will be large. But there will be risk of underfitting also in large value of k.
- But if it will be too small then it can cause overfitting.

- My model on kNN gave good accuracy on k = 1 and 2. When the value of 'r' increased till 15 it was 95%.

model accuracy