CS:5101 Machine Learning

Term 3 (Dec 2020 - Feb 2021)

Programming Assignment - 3

Classification Techniques
Naive Bayes

Due Date-14/12/2020

Follow the instructions given below carefully:

1. You are allowed to use ONLY the python libraries in today's assignment for modelling classifier

2. You must submit your code in a python .ipynb notebook with naming format as follows:
   Firstname_Lastname_assignment3.ipynb

3. For each question, create a separate text block containing the question followed by a code block containing the solution.

4. Your code must be properly commented explaining each step clearly.

5. If any of the above instructions are not followed, penalty will be there for the same.

6. Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for assignment 3.

Question You are provided with two three class dataset namely linearly separable and non-linearly separable. Learn a Naive Bayes classifier for both the dataset separately and report your results with observation as mentioned below. Assume data follows gaussian distribution. Data points are two dimensional.

- Your code should input train and test data from each of the corresponding files and learn a Naive bayes classifier. Class1_train implies train data with label class 1 and similarly for rest of files.

- Give following outputs in the python notebook itself with proper headings mentioning dataset name
  1) decision region plots learnt by the classifier (Plots should have train data superimposed in the figure and should have proper labelings and title)
  2)accuracy values for train data and test data
  3) confusion matrix for train and test data
  separately for linearly separable and non-linearly separable data.

- At last in a text block you should mention your observation about the decision surface that can be produced by Naive Bayes classifier on each dataset

- evaluation scheme:
  -2 mark- Naive Bayes(code)
  -1.5 mark - test accuracy, confusion matrix
  -1.5 mark - plot decision boundaries/regions, observation

NB: Try implementing logistic regression classifier on the same datasets and observe the performance and decision regions obtained for logistic regression classifier (this is a practise exercise which is not required to submit).