EE5180 Programming Assignment -2
Bayes Classifier
Release Date : 18th March,2023
Submission Date: 30th March,2023

Notes:

1. Please use moodle discussion threads to post your doubts and check it before posting if the same question has been asked earlier. Please feel free to reach out to Jeffin(EE19B085) or Rohith(EE19B114) if you have any questions about the assignment.
2. **What to Submit? :** Submit a single zip file in the moodle named as PA2_Rollno.zip containing reports named "Rollno_Report.pdf" and folders containing corresponding codes. If you are using a jupyter notebook it would suffice to include the observations in the notebook itself and submit a pdf version of the notebook.
3. Read the problem fully to understand the whole procedure.
4. Do not use any direct library functions to implement the classifiers(such as sklearn.naive_bayes).
5. Any plagiarism/cheating will be dealt very, very strictly.
6. You should thoroughly know what you are doing, and it will be asked in your viva.
7. Late submissions will be evaluated for reduced marks, and for each day after the deadline, we will reduce the weightage by 10%.

Questions

1. There are two datasets for this question:

   - dataset1 datasets folder
   - dataset2

   Hint: Use "$df = pd.read\_csv(r'path', delim\_whitespace = True, header = None)$" to read the files into a dataframe.

2. This is a binary classification task.

3. Each dataset contains datapoints and their corresponding class labels. Perform a 80-20 train-test split.

4. Construct Bayes classifiers under assumptions below, and use ML estimators to compute and return the results on test set for both the datasets.

   a)**Naive Bayes Classifier:** Assume $X|Y = 1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X|Y = 2 \sim \mathcal{N}(\mu_2, \Sigma_2)$. (where Σ_1 and Σ_2 are diagonal)

   b)**Bayes Classifier:** Assume $X|Y = 1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X|Y = 2 \sim \mathcal{N}(\mu_2, \Sigma_2)$.

5. Plot the decision boundary and contours for both Naive Bayes and Bayes classifiers for both the datasets and interpret the results

6. Show the confusion matrix for both Naive Bayes and Bayes classifiers for both the datasets and interpret the results