Odd Semester 2023-2024
Programming Assignment 1
CS401-Introduction to Machine Learning

Date: 08/09/2023

Note: Each group should take the data assigned to the group only

Classification Tasks:

Dataset: 2-dimensional data of 2, 3 or 4 classes:

(a) Linearly separable data set
(b) Nonlinearly separable data set
(c) Overlapping data set

Classifier to be built:

1) *Nearest neighbor classifier*
2) *K nearest neighbor classifier (Value of K to be chosen experimentally)*
3) *Reference template based classifier*
   a. Mean vector as reference template for a class
   b. Mean vector and covariance matrix as reference template for a class
4) *Bayes classifier-Unimodal Gaussian density*
   a. Covariance matrix for all the classes is the same and is $\sigma^2 I$
      i. You can obtain the same covariance matrix for all the classes by taking the average of covariance matrices of all the classes. You can obtain same variance by averaging all the variances.
   b. Full covariance matrix for all the classes and is same for all the classes
      i. Same covariance matrix for all the classes may be obtained by taking a verage of covariance matrices of all the classes
      ii. Same covariance matrix for all the classes by computing the covariance matrix of training data of all the classes combined.
   c. Covariance matrix is diagonal and is different for each class
   d. Full covariance matrix for each class and is different

Report should include the results of studies presented in the following each dataset:

1. Plot of training data with mean displayed in different color
2. Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data
3. Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.
3. Decision region plot for every pair of classes with the respective training data superimposed
4. Decision region plot for all the classes together with the training data superposed
5. Decision region plot for all the classes together with the testing data superimposed

Report should also include your observations about the performance and the nature of decision surface for each classifier, and for each dataset.

Submission guideline:

- The report should be submitted in PDF form only preferably prepared using LateX. Report should be named as Group<num>_Assignment1_IML.pdf. Eg. Group01_Assignment1_IML.pdf
  - The report must be communicated to me through an email with the subject name "CS401-IML Programming Assignment 1-Group <num>". One submission per group must be done keeping the other member in CC.
- The code and report should be bundled in a single zip file and uploaded to Google classroom (one submission per group). Name the zip file as Group<num>_Assignment1_IML.zip. Eg. Group01_Assignment1_IML.zip

Deadline for submission of report: 11:55 PM, Thursday, 22nd September 2023