

August-November 2024 Semester
CS616: Statistical Pattern Recognition
CS612: Statistical Pattern Recognition Laboratory
Programming Assignment I

Date: August 23, 2024

Deadline for submission of code and report: Thursday, Sept. 05, 2024, 10:00 PM

Classification tasks

Datasets:

Dataset 1: Linearly separable classes: 3 classes, 2-dimensional linearly separable data is given. Each class has 500 data points.

Dataset 2: Nonlinearly separable classes: 2-dimensional data of 2 or 3 classes that are nonlinearly separable. The number of examples in each class and their order is given at the beginning of each file.

Dataset 3: Real-world data (Vowel data): 2-dimensional data of 3. The real-world data set corresponds to the formant frequencies F1 and F2 for vowel utterances.

Divide the data from each class into training, and test data. From each class, train, and test split should be 70% and 30% respectively.

Assumption: Class-conditional densities are Gaussian

Classifiers to be built for each dataset: Bayes classifier.

1. Covariance matrix for all the classes is the same and is $\sigma^2 \mathbf{I}$.
 - a. You can obtain the same covariance matrix for all the classes by taking the average of covariance matrix of all the classes. You can obtain the same variance by averaging all the variances.
2. Full Covariance matrix for all the classes is the same and is Σ .
 - a. You can obtain the same covariance matrix for all the classes by taking the average of covariance matrix of all the classes.
3. Covariance matrix is diagonal and is different for each class.
4. Full covariance matrix for each class is different.

Presentation of results: Report should include the results of studies presented in the following forms for each classifier and for each dataset.

- 1) Confusion matrix, 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.
- 2) Constant density contour plot for all the classes together with the training data superimposed.
- 3) Decision region plot **superimposed by training data only** for each of the datasets. Give the decision region plot between each pair of classes and also, give the decision region plot all classes together (only if the number of classes are more than 2).
- 4) Inferences on the plots and inferences on the results observed (such as performance, nature of decision surface etc.) for each dataset.

Each group of students must use the dataset identified for that group only.

The expectation of the assignment is to implement from scratch using Python or MATLAB or any other programming language.

Note: You are not supposed to use libraries of Bayes classifier, multivariate Gaussian distribution, likelihood etc.

Report should be in **PDF** form and report by a team should also include the observations about the results of studies.

Instruction:

Upload in Moodle all your codes in a single zip file.

- **Give the name of the code folder as Group<number>_Assignment1_code**
Example: Group01_Assignment1_code.
- **Give the name of the zip file as Group<number>_Assignment1_code.zip**
Example: Group01_Assignment1_code.zip

Upload the report as PDF file.

- **Give the name to the report file as Group<number>_Assignment1_report.pdf**
Example: Group01_Assignment1_report.pdf

We will not accept the submission if you don't follow the above instructions.