

## Data Science Lab (CS 356)

### Assignment 10

Date: 04.04.2022

Due Date: CSE B: 11.04.2022

CSE A: 13.04.2022

#### *Instructions to submit the lab assignment*

- a. Add proper comment lines for each important step of the code.*
  - b. All the codes should be in same file.*
  - c. Name each file as rollnumber\_assignmentnumber.pdf.*
  - d. Upload the program file in google classroom.*
- 

1. Implement Linear Discriminant Analysis (LDA) from scratch using Python
  - a. Dataset: <https://archive.ics.uci.edu/ml/datasets/Iris>
  - b. Group the training data into its respective classes . [Form a dictionary and save data grouped by classes it belongs to]
  - c. Calculate mean vector of given training data of K-dimensions excluding the target class and calculate class-wise mean vector for the given training data
  - d. Calculate scatter matrices needed to maximize the difference between means of given classes and minimize the variance of given classes.
  - e. Calculate eigen values of M and get eigen vector pairs for first n (needed ) dimensions.
  - f. Selecting Linear Discriminants for the new features subspace
    - i. Sorting eigen vectors by decreasing eigenvalues
    - ii. Choosing k eigen vectors with the largest eigenvalues
  - g. Transforming the samples onto the new subspace.
2. Implement Logistic Regression by using LDA for the following dataset. [ sklearn Library can be used for LDA and Logistic Regression Computation]
  - a. Dataset: Uploaded in Google Classroom
  - b. Split the dataset
  - c. Feature Scaling
  - d. Apply LDA
  - e. Train the model with Logistic regression
  - f. Compute the Confusion matrix