Data Science Lab (CS 356)

Assignment 9

Date: 28.03.2022 **Due Date: 04.04.2022**

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 Principal Component Analysis from scratch in Python for the following dataset and show the following steps below.
   a. Dataset: https://archive.ics.uci.edu/ml/datasets/iris
   b. Scale the dataset.
   c. Calculate the covariance matrix for the features in the dataset.
   d. Calculate the eigenvalues and eigenvectors for the covariance matrix.
   e. Sort eigenvalues and their corresponding eigenvectors.
   f. Plot the principal components and percentage of explained variances.
   g. Choose first k eigen vectors
   h. Transform the original matrix.

2. Implement PCA and Logistic Regression for the following dataset by performing the required steps.
   a. Dataset: https://www.kaggle.com/datasets/dileep070/heart-disease-prediction-using-logistic-regression
   b. Loading the dataset
   c. Scale the dataset
   d. Select the principal components
   e. Build the Logistic regression model with the transformed dataset.