**Name: Avirneni vinay kumar**

**ID: 700734549**

**Course: Machine learning**

**Github Link:** [**https://github.com/vinay779/Assignment-5.ML**](https://github.com/vinay779/Assignment-5.ML)

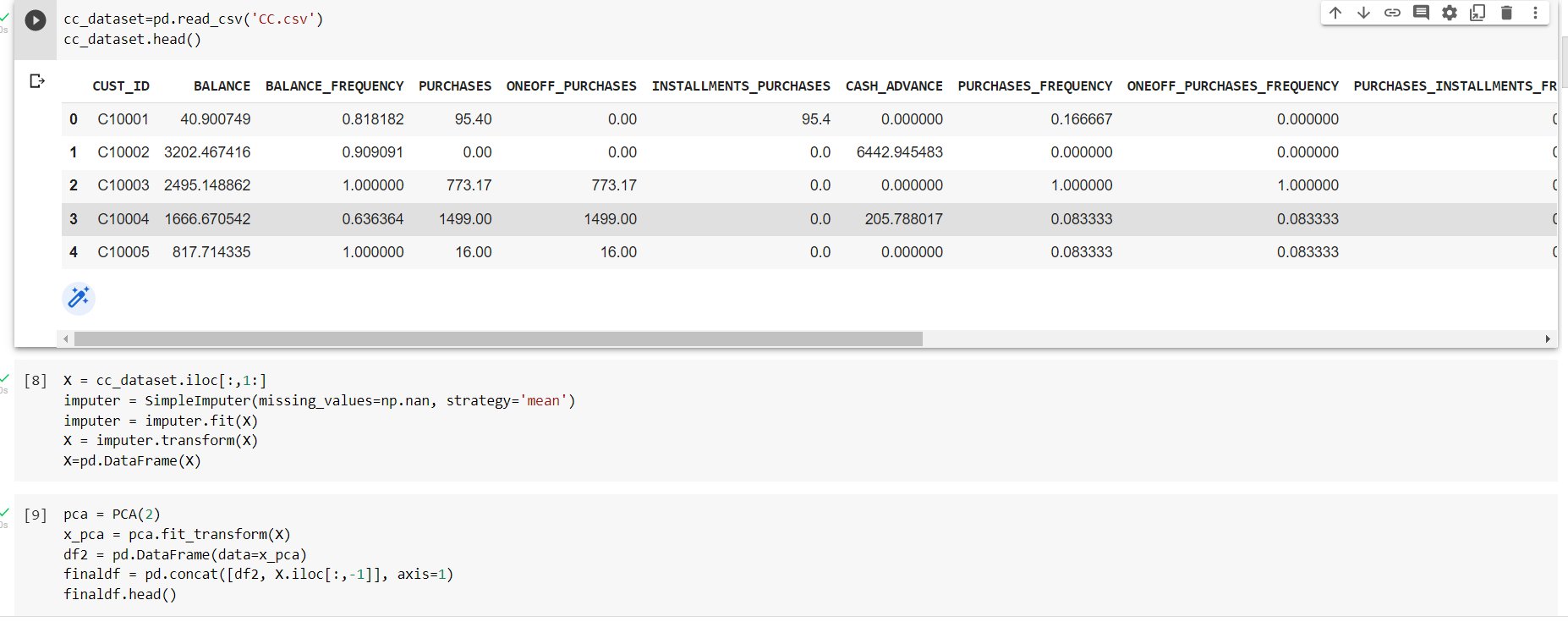
**Video Link:** [**https://youtu.be/5VFcYo2CAsg**](https://youtu.be/5VFcYo2CAsg)

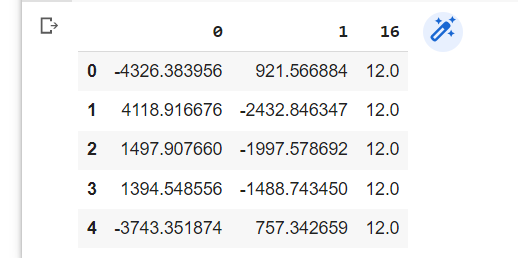
1. **Question1: Principal Component Analysis**

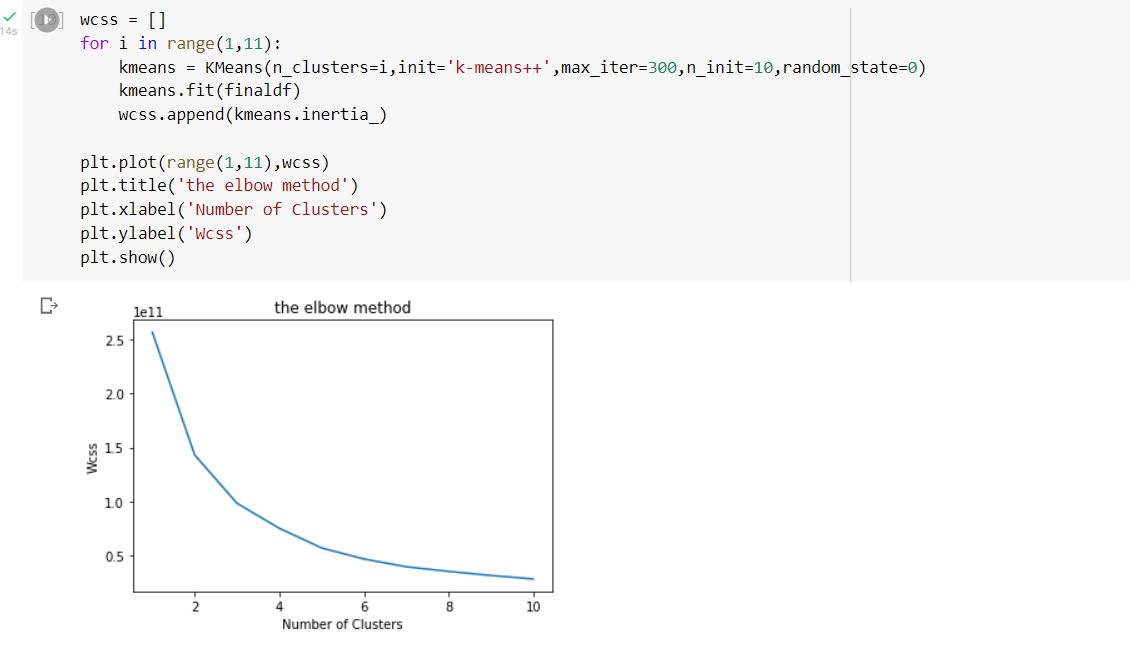
**a. Applying PCA on CC dataset.**

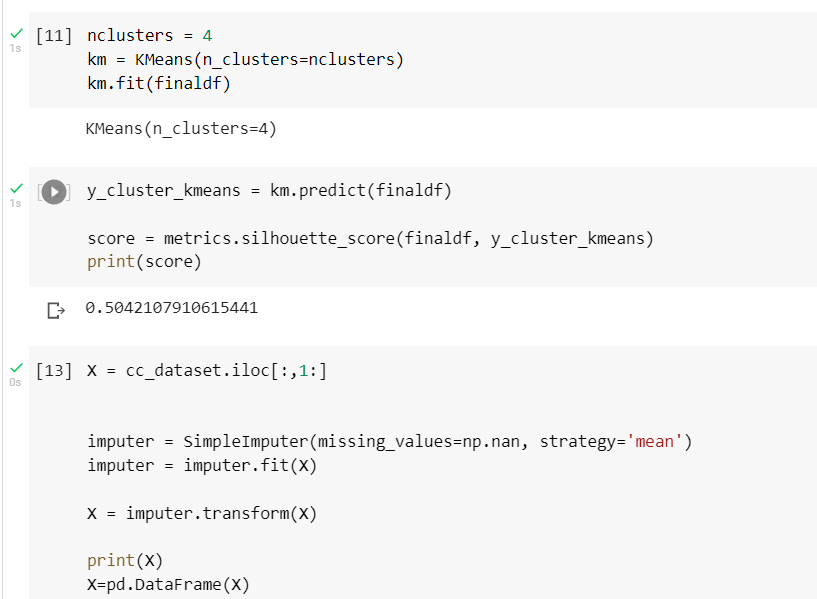
**b. Applying k-means algorithm on the PCA result and report your observation if the silhouette score has improved or not?**

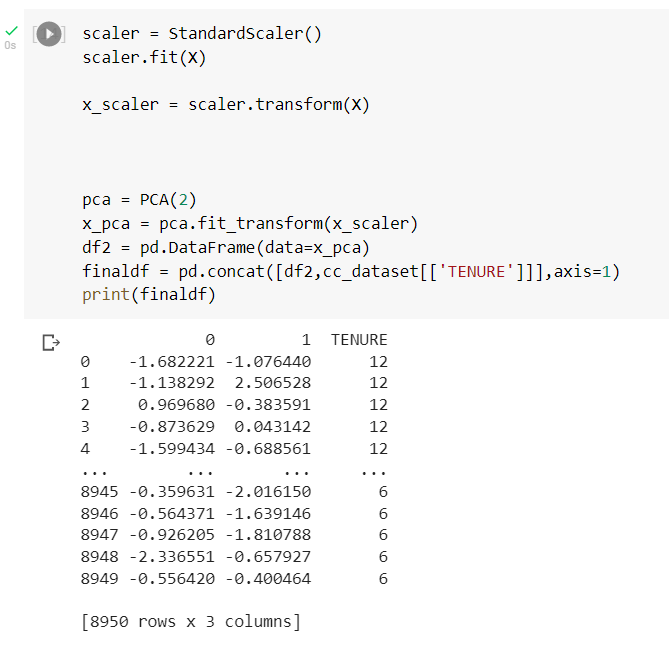
**c. Performing Scaling+PCA+K-Means and report performance.**

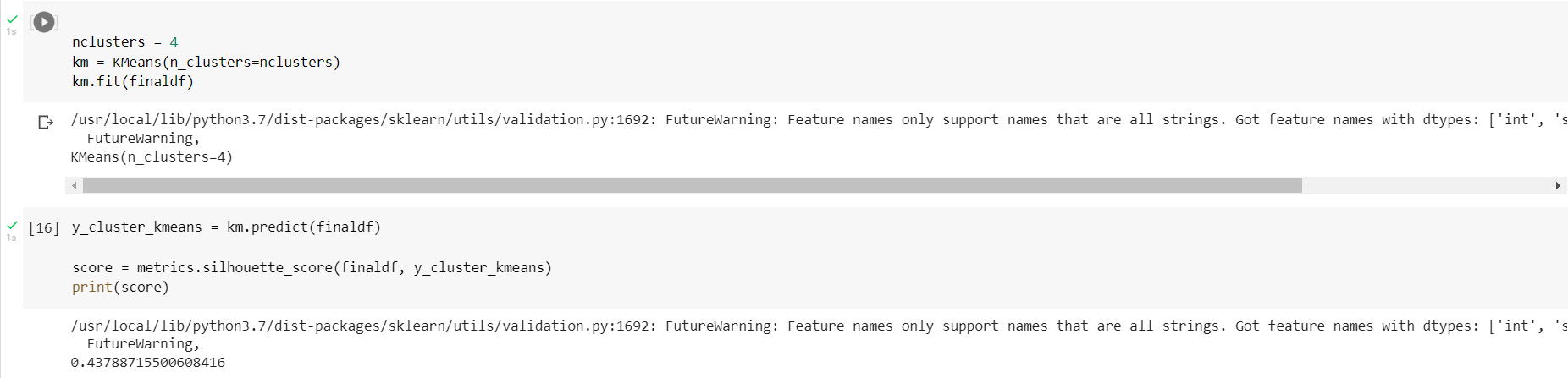
****

****

****

****

****

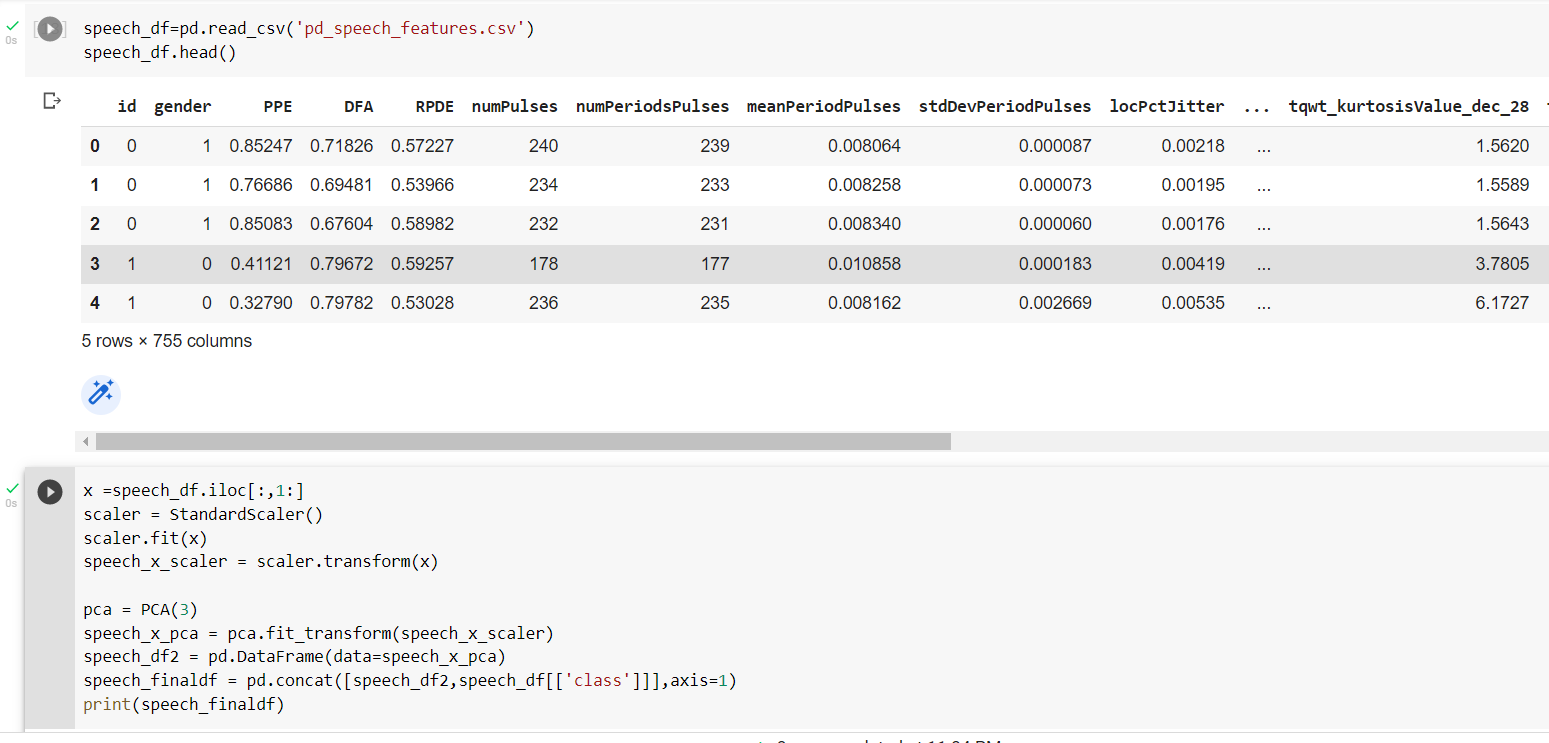
****

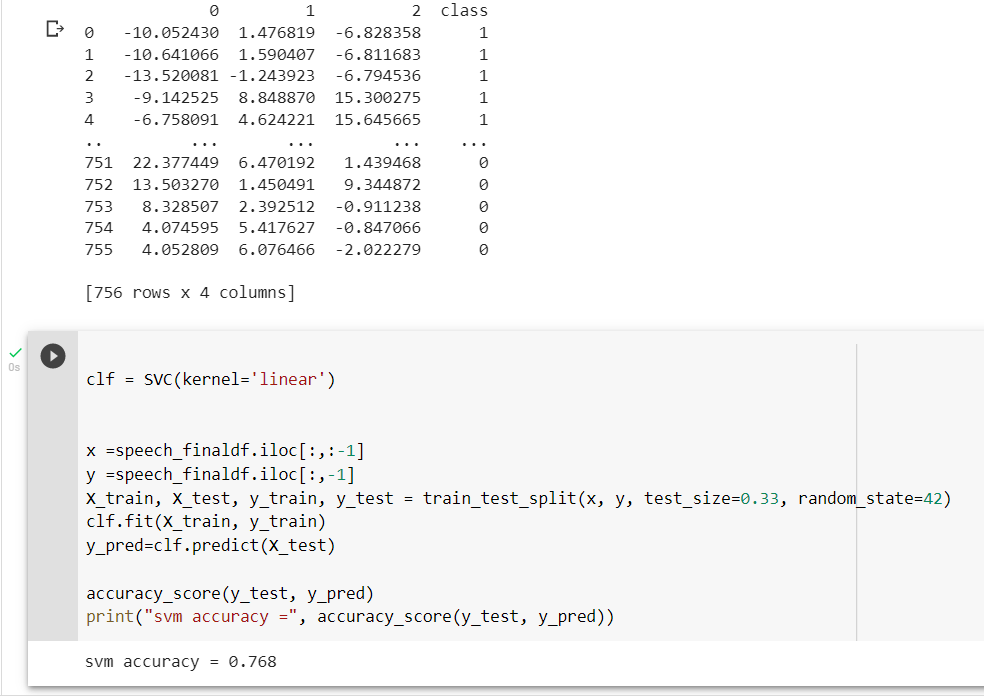
1. **Question 2: Use pd\_speech\_features.csv**

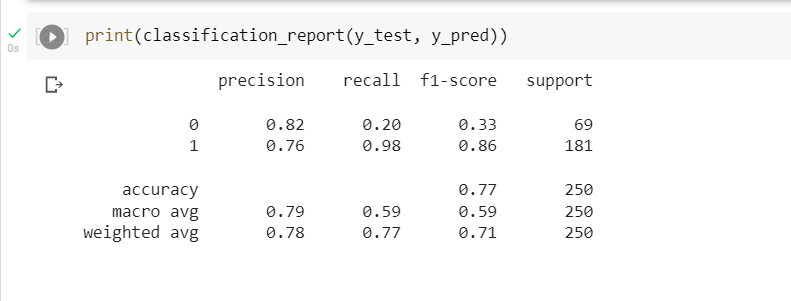
**a. Performing Scaling**

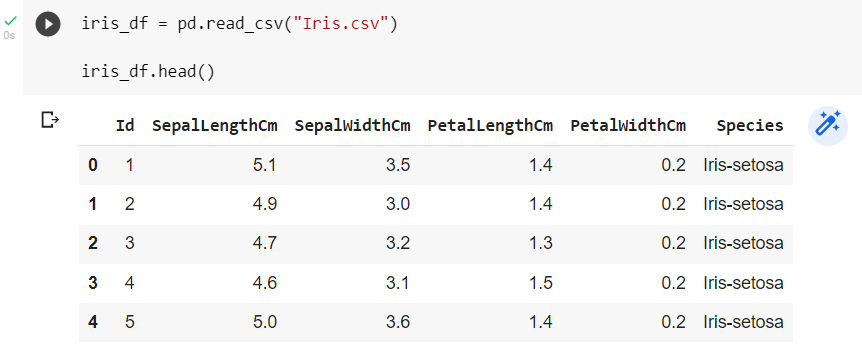
**b. Applying PCA (k=3)**

**c. Using SVM to report performance**

****

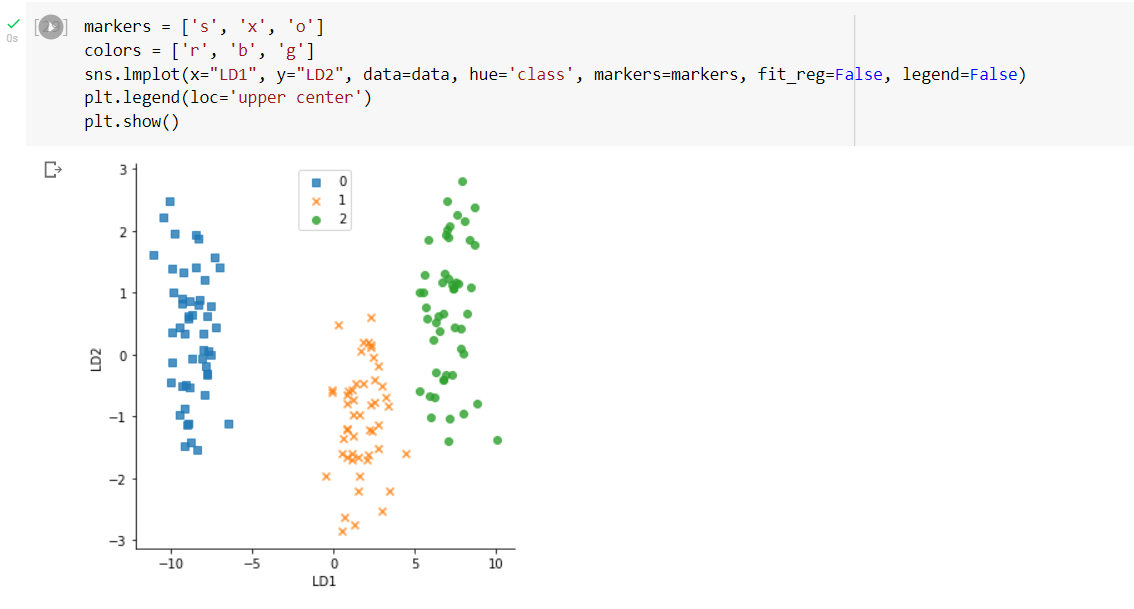
****

****

1. **Question 3: Applying Linear Discriminant Analysis (LDA) on Iris.csv dataset to reduce dimensionality of data to k=2.**

**Graphical user interface, application

Description automatically generated**

****

**Question 4:** **Identifying the difference between PCA and LDA**

**PCA and LDA are two popular dimensionality reduction methods for data with many input features. Both LDA and PCA are linear transformation algorithms, but LDA is supervised whereas PCA is unsupervised, and PCA does not take class labels into account. By locating the largest variance, PCA reduces the number of dimensions in high-dimensional data. The goal of LDA is to find the best feature subspace for class separation.**

**Both approaches rely on dissecting matrices of eigenvalues and eigenvectors, but the core learning approach is significantly different. LDA is supervised, whereas PCA is not. PCA reduces dimensions by investigating the relationships between different features. This is accomplished by creating a new subspace by constructing orthogonal axes - or principle components - with the highest variance direction. PCA generates components based on the direction in which the data has the most variation - for example, the direction in which the data is the most spread out. This component is known as principals and eigenvectors, and it represents a subset of the data that contains the majority of the information - or variance - in our data. LDA, on the other hand, does nearly the same thing, but it includes a "pre-processing" step that computes mean vectors from class labels prior to extracting eigenvalues.**

**PCA: To create the covariance matrix, take the joint covariance - or correlation in some cases - between each pair in the supplied vector.**

**LDA: For each class label, compute the d-dimensional mean vector, Make a scatter matrix for each class as well as a scatter matrix between classes.**