## Project Report

**Submitted by:**

Sri Chaitra Kareddy

sxk180037

Input:

Data Files:

* Iris.data – contains comma separated 5 values in one line and 150 lines. Each row is considered as one data point of 5 features
* Wine.data – contains comma separated 13 values in one line and 178 lines. Each row is considered a data point of 13 features

Label Files

* Iris.labels – contains 150 lines having the values of 1, 2 and 3
* Wine.labels – contains 178 lines having values of 1, 2 and 3

Output:

* A comma separated ﬁle containing the n×2 matrix of the reduced data – considered as each data point having only 2 features
* A comma separated ﬁle containing the two vectors v1,v2 as a matrix of size 2×m – considered as the two eigen vectors which transformed and reduced the data into only two features

Programs:

Part I:

* + Make sure the python file and the input data are in the same directory
* pca1.py
  + Read the ﬁles for the matrices XT and the labels y. Performs PCA on these to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 pca1.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 pc1.py iris.data iris.labels pca1IrisVectors.csv pca1IrisReducedData.csv
* pca2.py
  + Read the ﬁles for the matrices XT and the labels y. Performs PCA with mean subtraction on these to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 pca2.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 pc2.py iris.data iris.labels pca2IrisVectors.csv pca2IrisReducedData.csv
* scatter1.py
  + Read the ﬁles for the matrices XT and the labels y. Finds the different scatter groups and calculates the within groups scatter W. Uses this to calculate the last two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 scatter1.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 scatter1.py iris.data iris.labels scatter1IrisVectors.csv scatter1IrisReducedData.csv
* scatter2.py
  + Read the ﬁles for the matrices XT and the labels y. Finds the different scatter groups and calculates the between groups scatter B. Uses this to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 scatter2.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 scatter2.py iris.data iris.labels scatter2IrisVectors.csv scatter2IrisReducedData.csv
* scatter3.py
  + Read the ﬁles for the matrices XT and the labels y. Finds the different scatter groups and calculates the between groups scatter B and the within groups scatter W. Uses these to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 scatter3.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 scatter3.py iris.data iris.labels scatter3IrisVectors.csv scatter3IrisReducedData.csv

Part II:

* reducedim1.py
  + Read the ﬁles for the matrices XT and the labels y. Performs PCA on these by considering all data to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n×2. Writes to file the vectors v1, v2 and the matrix D
  + Execution: python3 reducedim1.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 reducedim1.py iris.data iris.labels reducedim1IrisVectors.csv reducedim1IrisReducedData.csv
* reducedim2.py
  + Read the ﬁles for the matrices XT and the labels y. Reduces the dimensions by considering groups and trying to maximize the ratio of their between class scatter B and within class scatter W. Uses these to calculate the top two vectors v1,v2, each of size m. Computes the projections of XT on v1,v2 and use them to create the matrix D of size n
  + Execution: python3 reducedim2.py <input data file> <input labels file> <output vectors file name>.csv <output reduced data file name>.csv
  + Example: python3 reducedim2.py iris.data iris.labels reducedim2IrisVectors.csv reducedim2IrisReducedData.csv