AIMIA

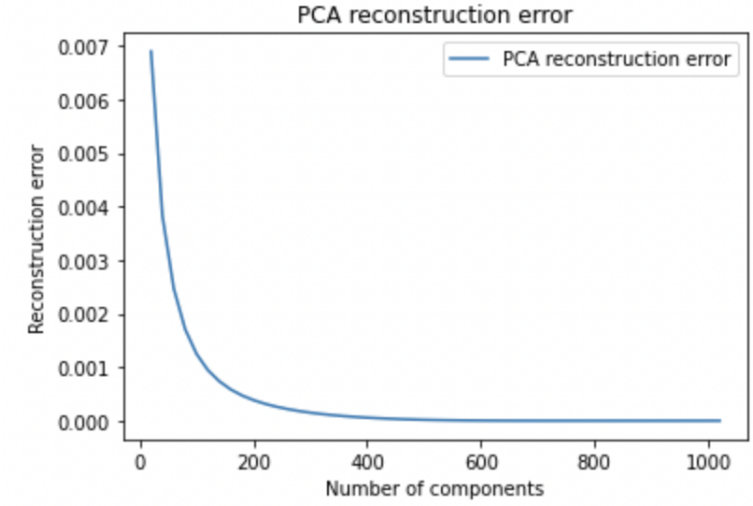
Homework 2

Abhishek Kumar Chaudhary

November 28, 2022

Problem 1:

The reconstruction mean square error (loss) after performing PCA is given below:
50 components: 0.00304
100 components: 0.00125
Reconstruction error as a plot of number of principal components:



Problem 2:

Choosing PCA components by elbow method. Above 300 components there is not much gain. Choosing number of components as 300 for further analysis. PCA projected embeddings is split into 70% training, 10% validation and 20% testing.
Grid search is done for both linear and rbf kernels of SVM with 5 fold cross validation and best hyper-parameter is chosen(can be seen in attached ipynb file.

Class-wise accuracy and F1 score in following order **[Normal Mild Severe]**:

- **Linear kernel SVM**:
  - Test:
    - Accuracy: [0.955 0.941 0.852]
    - F1 score: [0.938 0.946 0.898]

  - Train:
    - Accuracy: [0.972 0.953 0.939]
    - F1 score: [0.953 0.964 0.969]

- **RBF kernel SVM**:
  - Test:
    - Accuracy: [0.958 0.968 0.926]
    - F1 score: [0.961 0.965 0.926]
  - Train:
    - Accuracy: [1 1 1]
    - F1 score: [1 1 1]