BLG 454E Learning From Data (Spring 2018)

150130032-Baran Kaya Homework III

29.04.2018

1 Question 1.a

The main reason for reducing dimensionality is getting rid of the insignificant features for calculations. Using only dominant features for calculations is faster and makes results more accurate.

2 Question 1.b

We can calculate the result of the data that have all features and then we can calculate the result of the feature reduction data and compare them. Or we can just assume the rate of performance by feature numbers. (Feature reduction / All features)

3 Question 1.c

After PCA class labels are not like the left figure. Performance of the classification is reduced because it calculates classes with less than the left figure's number of features.

4 Question 1.d

Performance of the classification can be reduced. Because less data means poorer results. However, speed of the algorithm will be faster than the whole data.

5 Question 1.e

For calculating PCA eigenvalues and eigenvectors of covariance are used. For reducing the diamension of feature space k eigenvectors needed. I used 2 largest eigenvalues eigenvectors to create 2D feature space. Figure 1 and 2 show the eigenvectors.

After using the eigenvectors and multipling data with these vectors, result become 2D feature space. Figure 3 shows the result of data after PCA.

6 Question 2.a/b

Codes in the file.

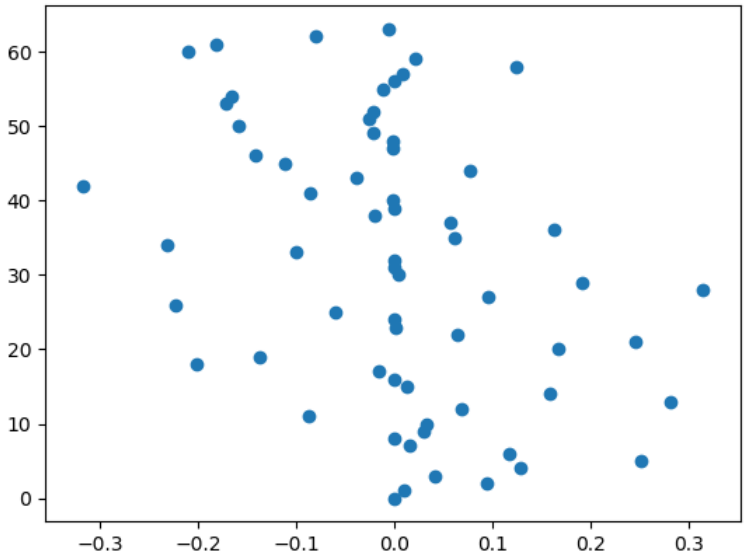

Figure 1: Largest Eigenvalue's Eigenvector Graph

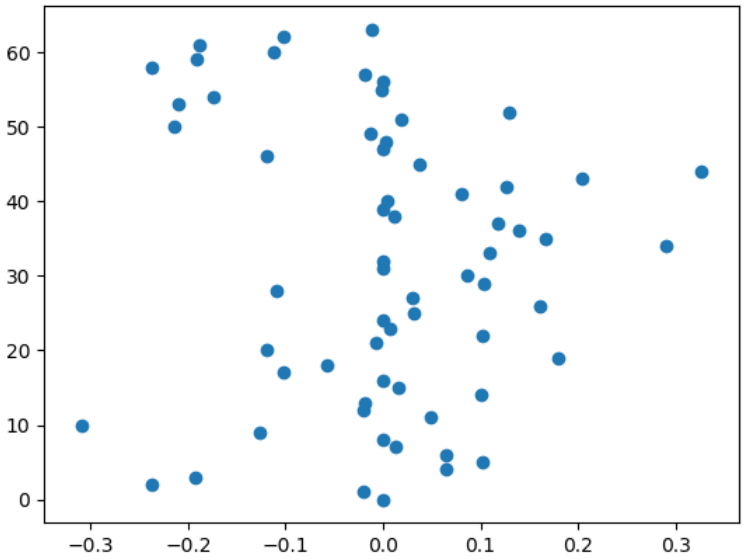

Figure 2: Second Largest Eigenvalue's Eigenvector Graph

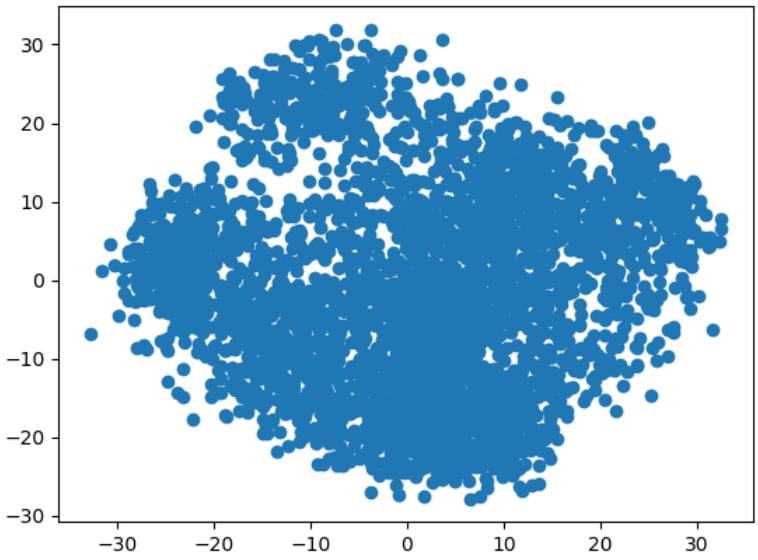

Figure 3: Data After PCA (2D)