Lab -8

PRML
AY 2020-21 Trimester - III

April 26 , 2021

Dimensionality reduction and Feature Selection

1. Using the data set, execute a PCA analysis using at least two dimensions of data (note that the last column should not be used here). In your report / code , discuss/include the following items. (20 marks)
    a. Standardize the data. (5 marks)
    b. How many eigenvectors are required to preserve at least 90% of the data variation? (5 marks)
    c. Look at the first eigenvector. What dimensions are the primary contributors to it (have the largest coefficients)? Are those dimensions negatively or positively correlated? (5 marks)
    d. Show a plot of your transformed data using the first two eigenvectors. (5 marks)
2. For the aforementioned dataset perform Linear discriminant analysis
    a. Compare the results of PCA and LDA. (5 marks)
    b. Plot the distribution of samples using the first 2 principal components and the first 2 linear discriminants. (5 marks)
    c. Learn a Bayes classifier using the original features and compare its performance with the features obtained in part (b). (5 marks)
3. Perform feature selection using any 2 methods studied in class and do the classification for the dataset using a classification algorithm of your choice. Do the following tasks:
    a. Preprocess the data and perform exploratory data analysis. (5 marks)

b. Identify the features having high significance using both of the methods. (5 marks)
c. Calculate and compare the accuracy and F1 score by both the methods and with the classifier learned using all the features (without doing feature selection), and analyze which method performs the best and why. (5 marks)
d. Use Pearson Correlation and compute correlated features with a threshold of 70%. (5 marks)

Please submit the necessary codes (Notebook) containing your output, and a PDF explaining and analyzing (e.g., what design choices would lead to better prediction) the steps for all the parts in both the questions along with necessary plots/figures.

Note: No submission will be accepted after the final deadline.

Demo reference

References:

- PCA Explanation
- PCA Library
- Feature Selection
- Correlation Plot