ASDS 5303 Project 5

Due Friday Dec. 6th, 2024

Answer the following questions. The goal of this lab is to build a apply the PCA for the dataset 'Wine'.

Note: You need to submit two files for this assignment, a pdf report and the original code file in R. Miss the original code file will have 50% reduction of your score. Late submission is not accepted.

Download the dataset, Wine.

1. Load the Wine dataset and normalize the features. Why is feature normalization important when applying PCA?
2. Apply PCA to reduce the dimensionality of the Wine dataset. Keep enough principal components to explain 90% of the variance. How many principal components are required to capture 90% of the variance?
3. Plot the explained variance ratio and the cumulative explained variance. What does this tell you about the dataset?
4. Plot the data in the first two principal component spaces. Can you see clear separation between the three wine cultivars?
5. Reconstruct the original dataset from the reduced PCA components. What information is lost when reducing the dimensionality?