Hude Hude hh3024

Problem Set 5 - Advanced Econometrics

This MATLAB script performs various tasks related to K-means clustering and Principal Component Analysis (PCA) on the Iris dataset.

1. K-means Clustering

- **Part 1a**: K-means clustering is applied to the raw Iris data, and performance metrics (precision, recall, F1-score, accuracy) are calculated.
- **Part 1b**: The same process is repeated after normalizing the data, and results are compared to the unnormalized case.

2. Principal Component Analysis (PCA)

- **Part 2a**: PCA is performed on the Iris dataset, and the variance explained by the first three principal components is calculated.
- **Part 2b**: The data is visualized in the spaces defined by the 1st, 2nd, and 3rd principal components.

3. K-means Clustering on Principal Components

- **Part 3a**: K-means clustering is performed on the first two principal components, both with and without normalization.
- **Part 3b**: K-means clustering is performed on the first three principal components, and performance metrics are computed for normalized and unnormalized data.