ASSIGNMENT 6

K-MEANS CLUSTERING

AI LAB

Name: SHARAD SHAHI
Roll No: 22052332
Section: 29
Branch: CSE
Date of Submission: 09.02.2025
GitHub Link: 22052332-ASSIGNMENT

Abstract

This report presents an implementation of the **K-Means clustering algorithm** from scratch using **NumPy and Matplotlib**, without relying on external machine learning libraries. The dataset, preprocessed through **Min-Max scaling normalization**, is clustered for **k=2 and k=3**, with results visualized using scatter plots. Each cluster is represented with distinct colors, and centroids are highlighted. The plots are saved in a specified directory for further analysis. This implementation showcases the effectiveness of **unsupervised learning** in identifying natural groupings within data and provides a foundational understanding of **clustering techniques** in machine learning.

Q. In your report, you should have 2 plots - one for k = 2 and the other for k = 3. In the plot, you need to plot the given dataset and use different colors to show points belonging to a particular cluster.

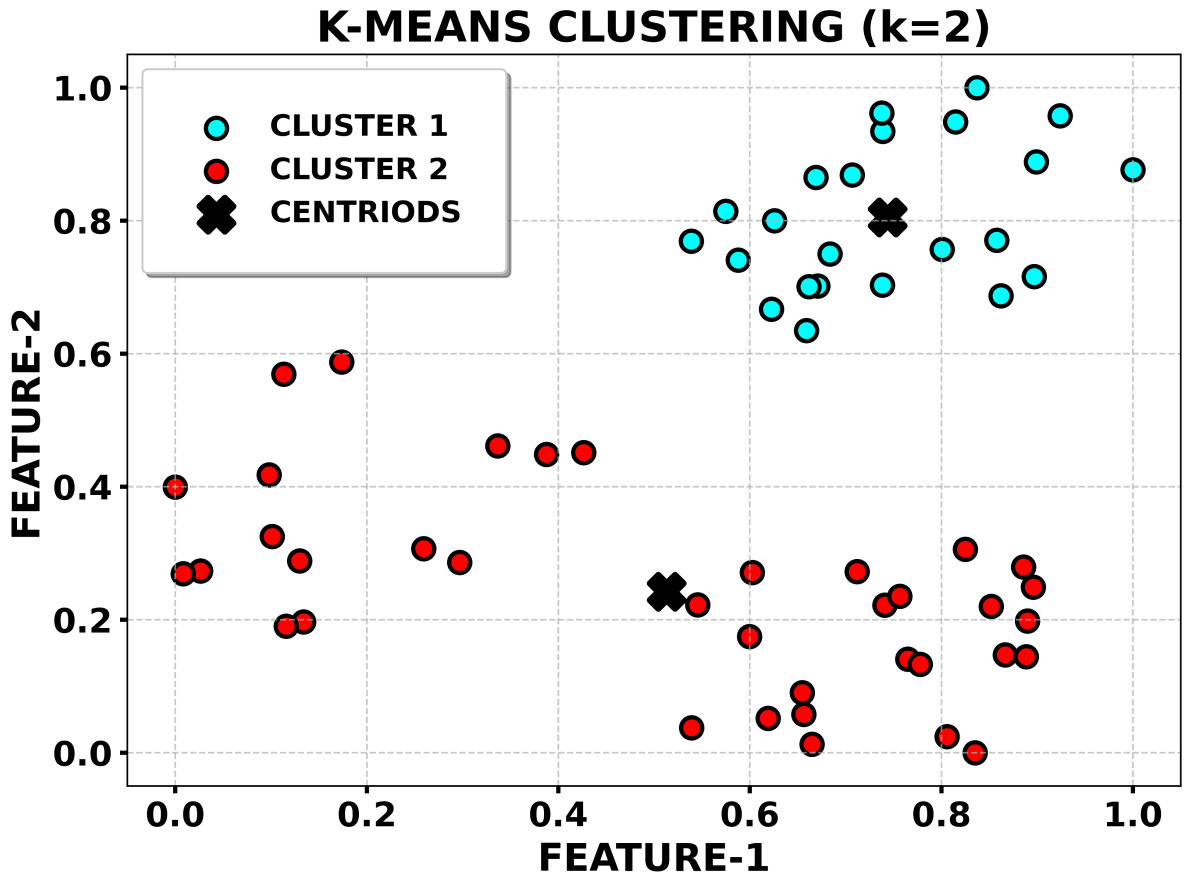

Figure 1: Clustering for k = 2

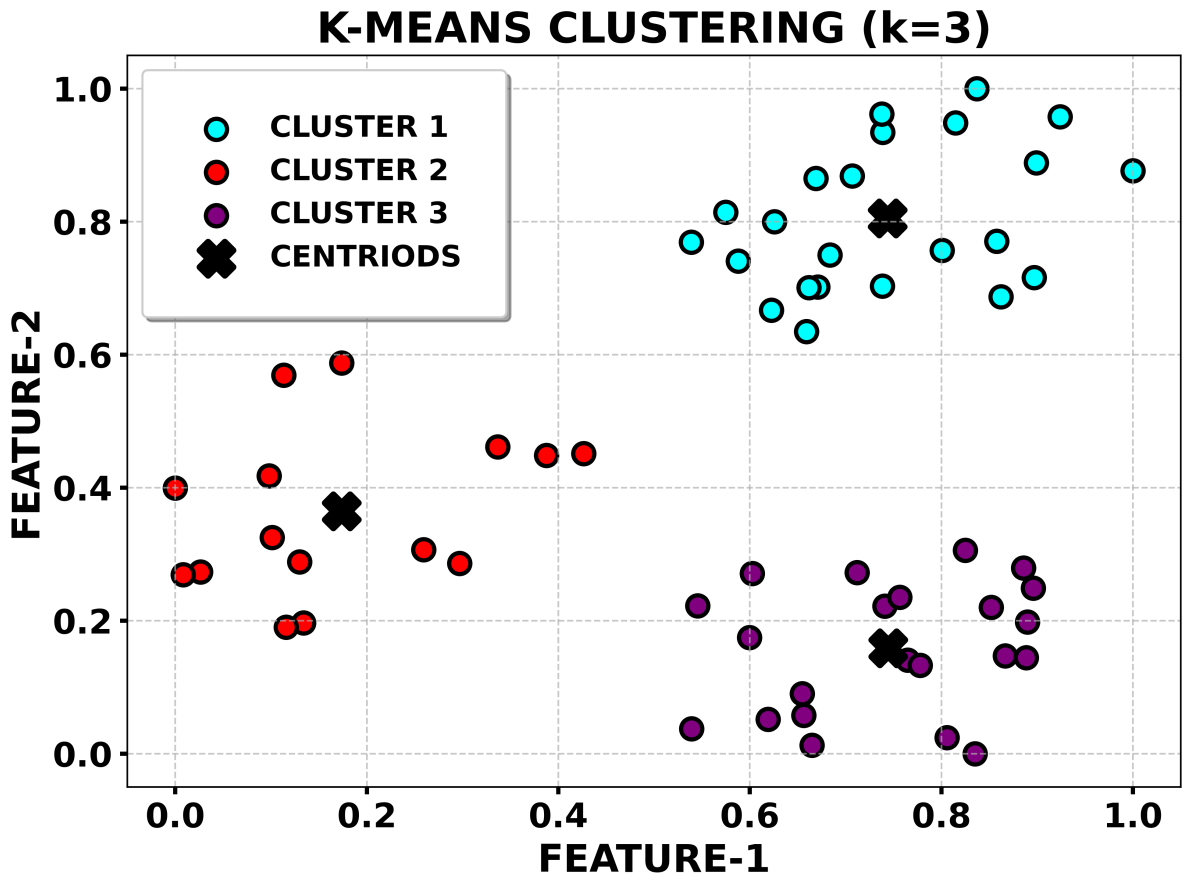

Figure 2: Clustering for k = 3

Conclusion

The implementation of **K-Means clustering** successfully groups data points into meaningful clusters, demonstrating the algorithm's ability to partition datasets based on similarity. Through **normalization**, the clustering process ensures fair distance-based calculations, leading to well-defined cluster formations. The results, visualized for **k=2 and k=3**, illustrate how the number of clusters impacts data segmentation. This project provides a solid foundation for exploring **unsupervised learning techniques** and can be extended further for applications in **data analysis, anomaly detection, and pattern recognition**.