K-Means Clustering Machine Learning Project

Objective

The aim of this project is to implement K-Means Clustering on a dataset and determine the optimal number of clusters using the Elbow Method. The final clustering results are visually analyzed to understand how the data is grouped.

1. Importing Required Libraries

The project begins by importing necessary libraries. These include tools for performing K-Means clustering and for visualizing the results through plots.

2. Elbow Method for Optimal Number of Clusters

To find the best number of clusters, the Elbow Method is used. This involves running the K-Means algorithm with varying numbers of clusters and recording the resulting inertia values. Inertia refers to the sum of squared distances from each data point to its nearest cluster center. By observing how inertia decreases with an increasing number of clusters, one can identify the point where adding more clusters no longer significantly reduces inertia. This point, known as the "elbow," suggests an appropriate number of clusters for the data.

3. Plotting the Elbow Curve

The recorded inertia values are plotted against the number of clusters. The resulting graph, known as the Elbow Plot, helps visualize the trade-off between the number of clusters and the model performance. The ideal number of clusters is typically located at the point where the curve bends, forming an elbow-like shape.

4. Final K-Means Model and Visualization

Once the optimal number of clusters is determined, the K-Means model is trained using that specific number. The dataset is then clustered accordingly. To evaluate the clustering results, the data is plotted with colors indicating the assigned clusters. This visual representation makes it easy to interpret the grouping structure within the dataset.

It is important to ensure that the data used for visualization is appropriately selected, typically by choosing two features from the dataset to represent the data in a 2D plot.

Conclusion

This project demonstrates the effectiveness of K-Means clustering in grouping similar data points. By using the Elbow Method, the project identifies the most suitable number of clusters and visualizes the clustering results. This process provides valuable insights into the structure of the dataset and the relationships among the data points.