

Experiment 6

Objective: Implement K-Means Clustering for Customer Segmentation. Group customers into segments based on purchasing behavior using K-Means.

Explanation:

1. Elbow Method

The Elbow Method is a visual approach to find the optimal number of clusters (k) in K-Means Clustering. It evaluates the within-cluster sum of squares (WCSS), also known as inertia, which measures the sum of squared distances between data points and their nearest cluster center.

Steps:

- Plot the number of clusters (k) on the x-axis and the inertia (WCSS) on the y-axis.
- The WCSS typically decreases as the number of clusters increases.
- The goal is to identify the "elbow point," where the inertia reduction slows down significantly.
- The elbow point suggests the optimal k , balancing both accuracy and simplicity.

2. K-Means Clustering Approach

K-Means Clustering is an unsupervised machine learning algorithm used to group data into clusters based on similarity. It tries to minimize the distance within each cluster while maximizing the separation between clusters. It's commonly used in customer segmentation, image compression, and anomaly detection.

Step 1: Choose the Number of Clusters (k)

Step 2: Initialize Centroids

Step 3: Assign Data Points to Nearest Centroid

Step 4: Update Centroids

Step 5: Repeat Until Convergence

Exercise 1:

Dataset: Mall Customer Segmentation Data (Available on Kaggle)

- Features: Annual Income, Spending Score, Age, Gender

Steps:

1. Perform exploratory data analysis (EDA) and visualize the dataset using scatter plots.
2. Apply the **Elbow Method** and **Silhouette Score** to determine the optimal number of clusters.
3. Implement **K-Means Clustering** using sklearn.
4. Visualize clusters using a scatter plot with color coding.
5. Provide insights on each customer segment.

Exercise 2:

Dataset: **World Happiness Report**

Description: Dataset with various factors like GDP, social support, life expectancy, and happiness score.

Use Case: Cluster countries to analyze common factors influencing happiness.

1. Perform K-Means Clustering using the World Happiness Report dataset.
2. Determine the optimal number of clusters using the Elbow Method and Silhouette Score.
3. Visualize the clustering results.
4. Interpret the findings to identify patterns among countries.

Exercise 3:

Dataset: **World Happiness Report**

Description: Dataset with various factors like GDP, social support, life expectancy, and happiness score.

Use Case: Cluster countries to analyze common factors influencing happiness.

1. Perform K-Means clustering using multiple values of K.
2. Compare the cluster results using metrics like WCSS and Silhouette Score.
3. Visualize the clusters for different values of K to understand the impact of changing the number of clusters.