

Lab 6: Clustering Methods – Analysis and Questions

Dhruv Parmar
Student ID: 202518030

September 26, 2025

1. Optimal Clusters

According to the Elbow Method (Figure 1), the WCSS curve noticeably levels off at around $k=5$. This inflection suggests that five clusters provide a good fit for the K-Means model. In the case of hierarchical clustering, inspection of the dendrogram (Figure 2) reveals a prominent cut line corresponding to six groups, which motivated our choice of 6 clusters for the Agglomerative approach.

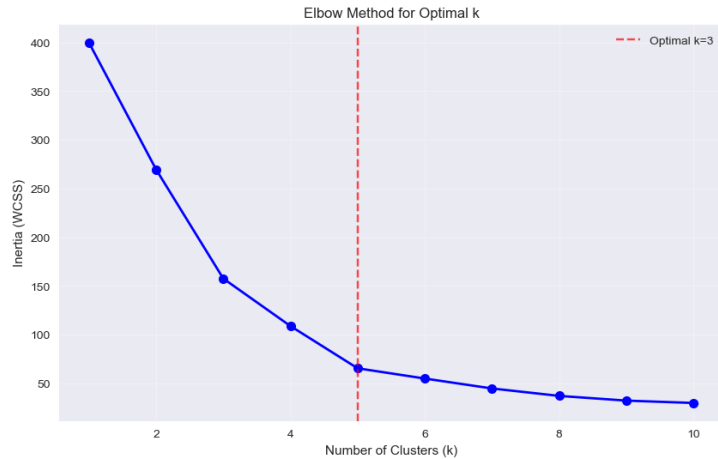


Figure 1: Elbow Method showing optimal $k = 5$.

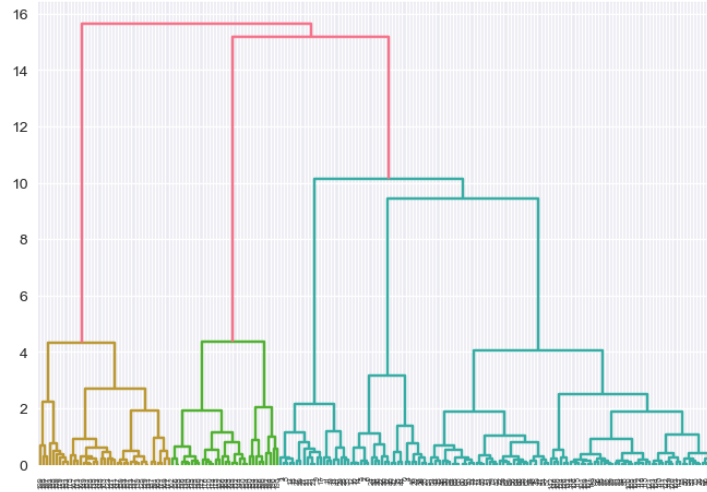


Figure 2: Dendrogram indicating optimal clusters for hierarchical clustering.

2. Cluster Comparison

The K-Means algorithm grouped the customers into five tight, almost circular clusters. Agglomerative hierarchical clustering with six groups produced a broadly similar segmentation, though the borders between groups differ slightly because of its bottom-up merging procedure. DBSCAN, on the other hand, discovered four high-density regions and labelled a handful of observations as noise — points that K-Means and Hierarchical were compelled to assign somewhere.

3. DBSCAN Performance

DBSCAN does not require the number of clusters to be specified in advance. With parameters $\varepsilon = 5$ and `min_samples=5`, the algorithm found four principal clusters and labeled approximately 12 observations as noise (label `-1`). Unlike K-Means and Agglomerative clustering, which assign every point to a cluster, DBSCAN leaves low-density points unassigned — treating them as noise — while accurately identifying high-density regions.

4. Algorithm Suitability

In this dataset, the customer data form fairly compact and well-separated groups in the two-feature space (Annual Income vs. Spending Score). Con-

sequently, K-Means turned out to be the most appropriate method. Hierarchical clustering achieved similar patterns but is slower as the dataset grows. DBSCAN shines for oddly shaped clusters or when identifying outliers is a priority.

5. Real-World Application

These customer segments can directly inform the mall's marketing decisions. For instance, the group with high income but low spending could be approached with tailored premium promotions or loyalty rewards to encourage more purchases. High-spending clusters could be maintained with exclusive benefits, while lower-income yet active spenders might respond well to discounts or bundled offers.