# Lab 6 Report: Clustering Analysis

Aditya Jana
Student ID: 202518035

September 26, 2025

## Q1. Choosing Cluster Numbers

From the Elbow method (see Figure 1), the WCSS graph starts to bend around $k = 5$. This suggests that 5 clusters is the good choice for K-Means. In case of hierarchical clustering, the dendrogram (Figure 2) looks like it can be cut into 6 groups, so we used 6 clusters there.

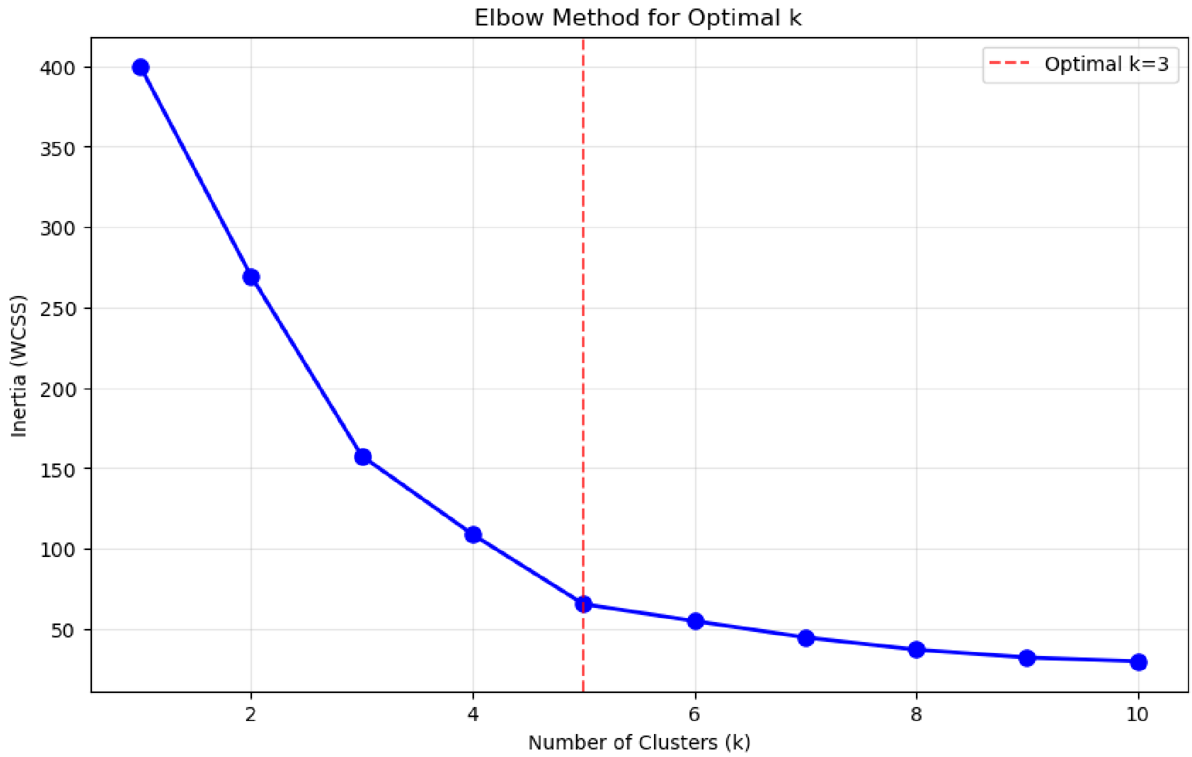


Figure 1: Elbow curve, bend at $k = 5$.

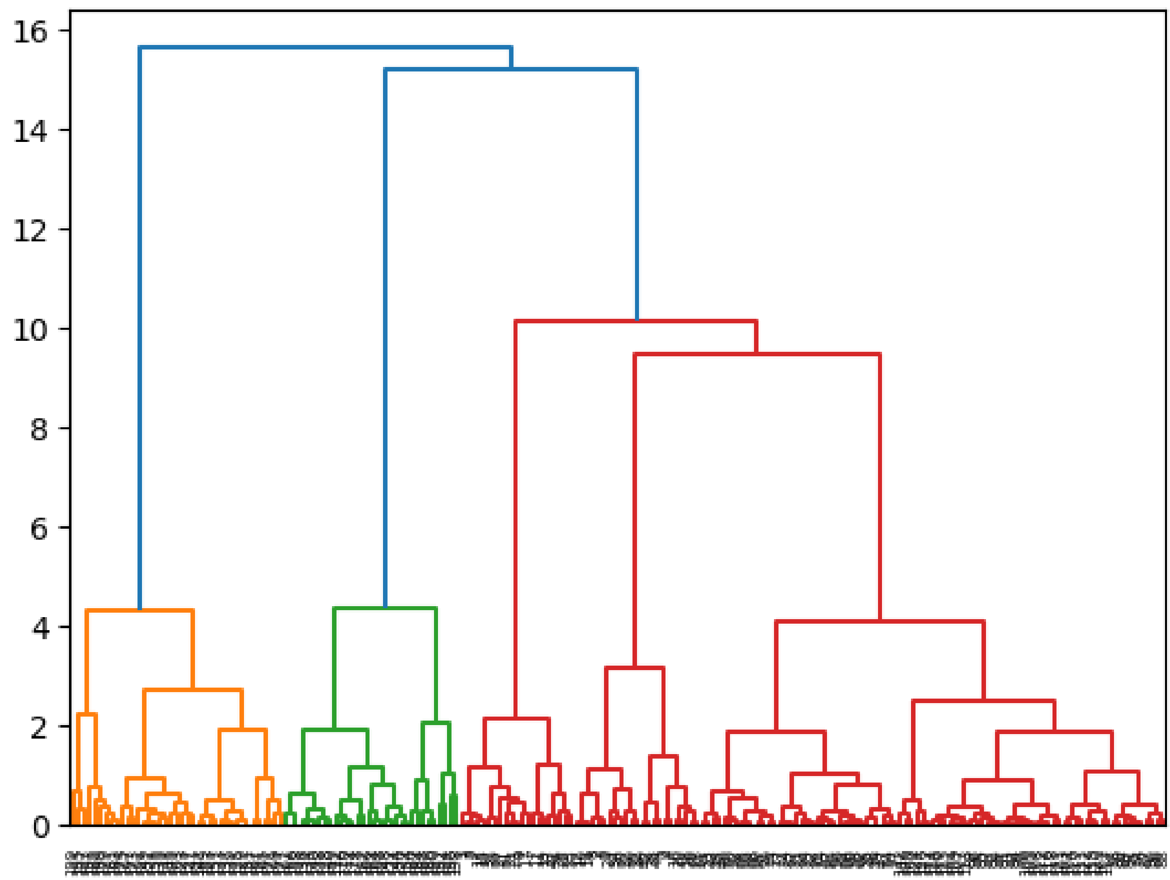


Figure 2: Dendrogram cut for 6 groups.

## Q2. Comparing the Algorithms

**K-Means** gave 5 clear and small round clusters. **Agglomerative clustering** with 6 clusters gave nearly same type of result, but not exactly same edges because it merges step by step. **DBSCAN** was different: it found 4 clusters and also put some points as noise, which the other two methods did not.

## Q3. DBSCAN Observations

The nice thing about DBSCAN is that we don't have to say the number of clusters before. With $\varepsilon = 5$ and `min_samples` $= 5$, it found 4 groups and around 12 points marked as noise ($-1$). Compared to K-Means and hierarchical, DBSCAN did better with dense areas but in sparse places it just pushed them as noise instead of forcing them to belong in some cluster.

## Q4. Which Algorithm Fits Best?

Looking at the dataset, the customers seem kind of **compact and well separated** on the income vs. spending plot. So **K-Means** looks like the best match here. Hierarchical also works but is slower when dataset is big. DBSCAN is more helpful if clusters are odd shaped or when we really need to catch outliers.

## Q5. Real-World Use

These customer groups can help the mall decide marketing. For example, people with **high income but low spend** can be pushed with special premium deals or loyalty cards. **High spenders** can get extra benefits so they keep shopping more. And the group with **lower income but high spend** may be given discounts or combo offers.