Silhouette Method vs Elbow Method

Dr. Kaushik Deb
Professor,
Dept. of CSE, CUET

Contents

❖ **Silhouette Method vs Elbow Method**

❖ **How Optimal K in Elbow Method calculated?**

❖ **How Optimal K in Silhouette Method calculated?**

❖ **Analytical Result on Silhouette Method and Elbow Method**

Silhouette Method vs Elbow Method

Issues	Silhouette Method	Elbow Method
Objective Evaluation	Clear quantitative measurement on optimal K (Max K) value.	Often vague representation on the optimal K (distortion point)
Considers Cluster Density and Separation	Takes into account both the compactness (intra-cluster similarity) and separation (inter-cluster dissimilarity) of cluster	Focuses on the sum of squared distances, which may not capture the complexity of cluster structures
Better Handling of Overlapping Clusters	Can handle overlapping clusters more effectively as it considers the cohesion and separation of data points with respect to all clusters	May struggle to identify the optimal number of clusters when clusters overlap, as it relies on the compactness of clusters

Department of CSE, CUET

How Optimal K in Elbow Method calculated?

Elbow method

❑ **Steps:**

1. Select a range of values for K to test, for example from 2 to 10

2. For each value of K:

   a) Initialize the K centroids randomly

   b) Run the K-means algorithm with the current K value

   c) Calculate the WCSS for each cluster.

3. Plot the WCSS against the number of clusters (K) on a graph.

4. Look for an "elbow" point on the graph which is the point where the WCSS begins to decrease at a slower rate

5. The value of K at the "elbow" point is considered the optimal number of clusters.
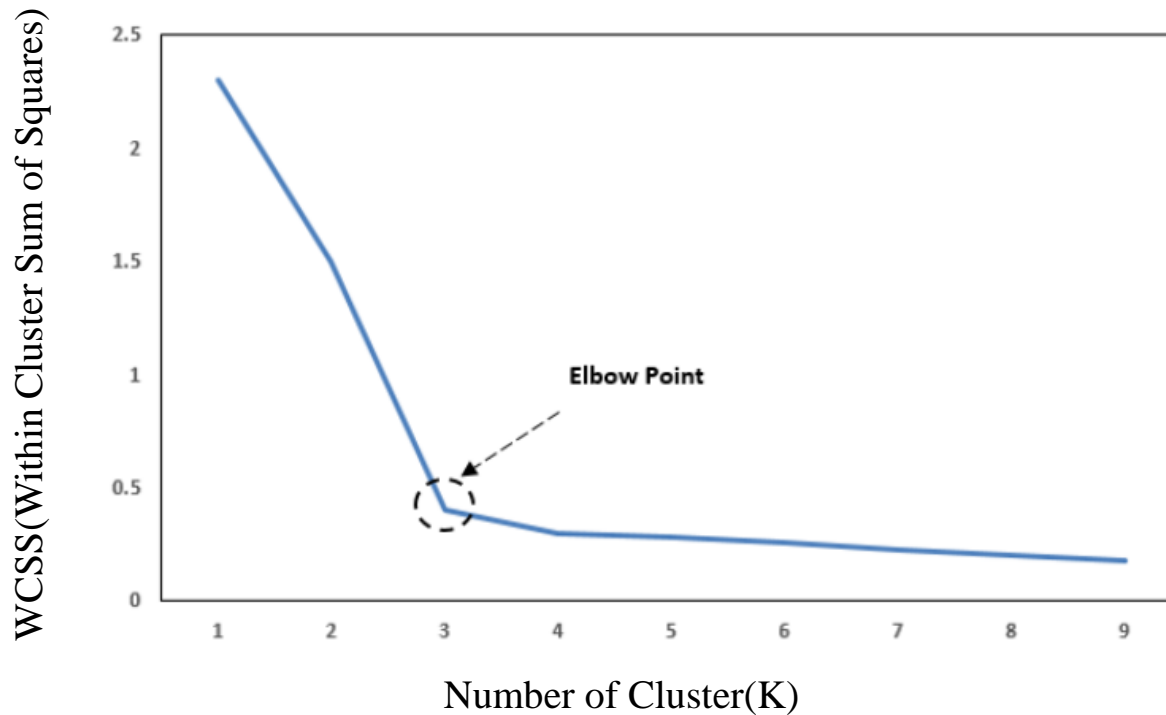
Department of CSE, CUET

Elbow method

$$\text{WCSS} = \sum_{C_k}^{C_n} \left(\sum_{d_i \text{ in } C_i}^{d_m} distance(d_i, C_k)^2 \right)$$

Where,

C is the cluster centroids and d is the data point in each Cluster.

Department of CSE, CUET

How Optimal K in Silhouette Method calculated?

Silhouette Score

□ **Steps:**

➢ Calculate the silhouette coefficient for each point of the clusters

➢ After this calculate average to get silhouette score for each clusters

➢ Calculate silhouette scores for different values of K

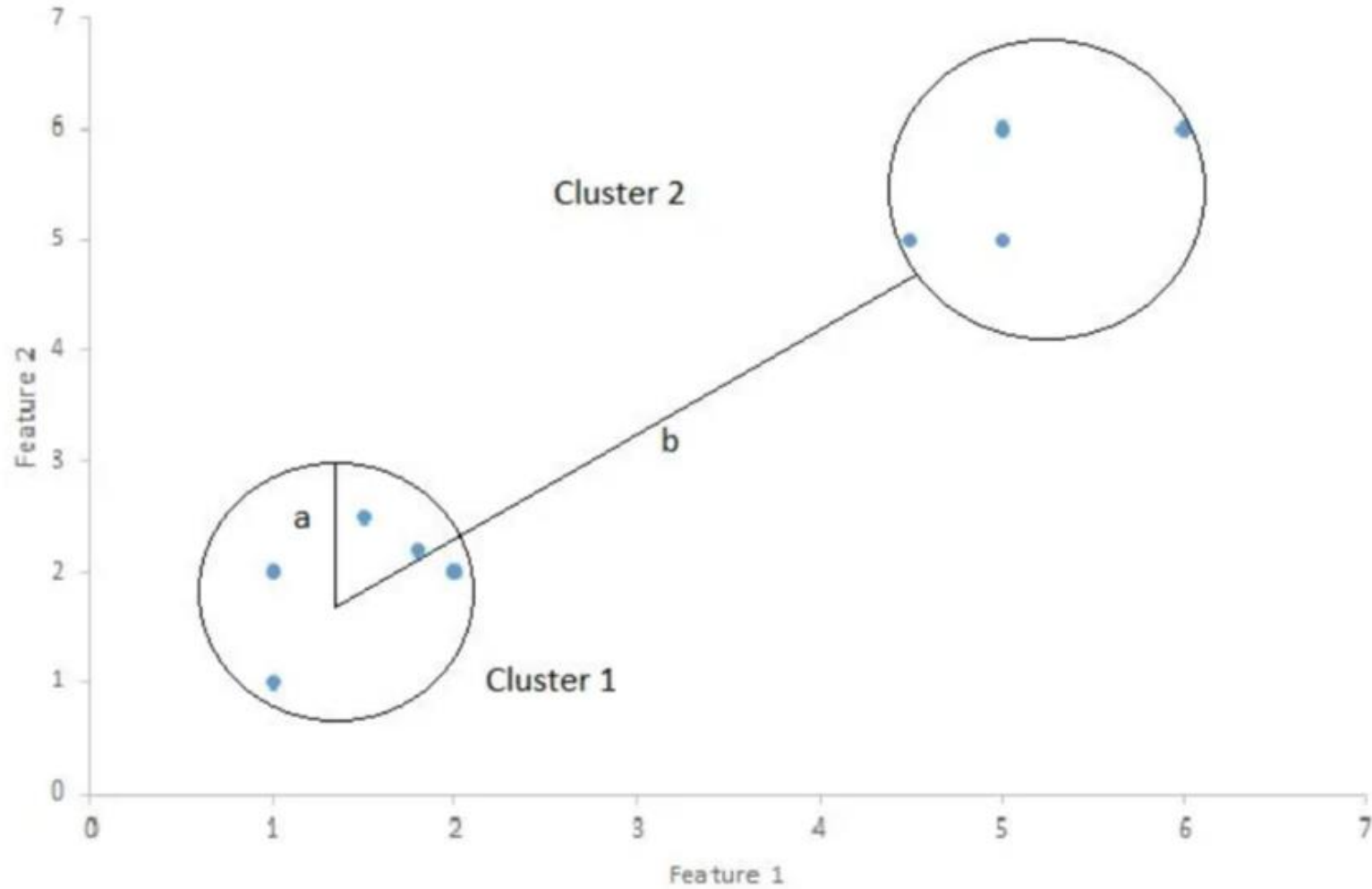➢ Plot the silhouette score vs number of clusters and get the optimal K

Department of CSE, CUET

Silhouette Score

❑ **Steps to find the silhouette coefficient of an i'th point:**

1. Compute a(i): The average distance of that point with all other points in the same clusters.

2. Compute b(i): The average distance of that point with all the points in the closest cluster to its cluster.

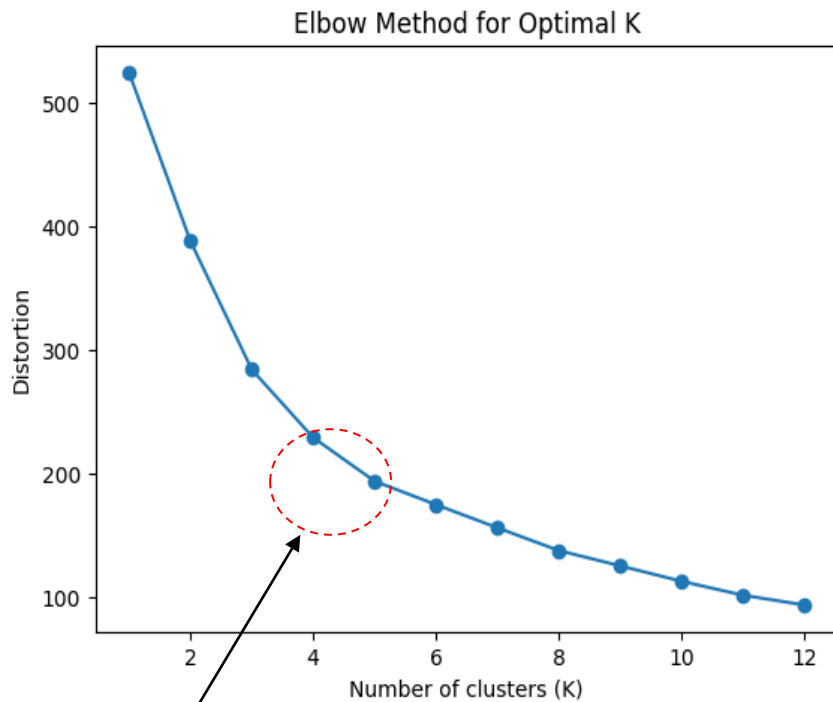3. Compute s(i): Silhouette coefficient or i'th point using below mentioned formula:
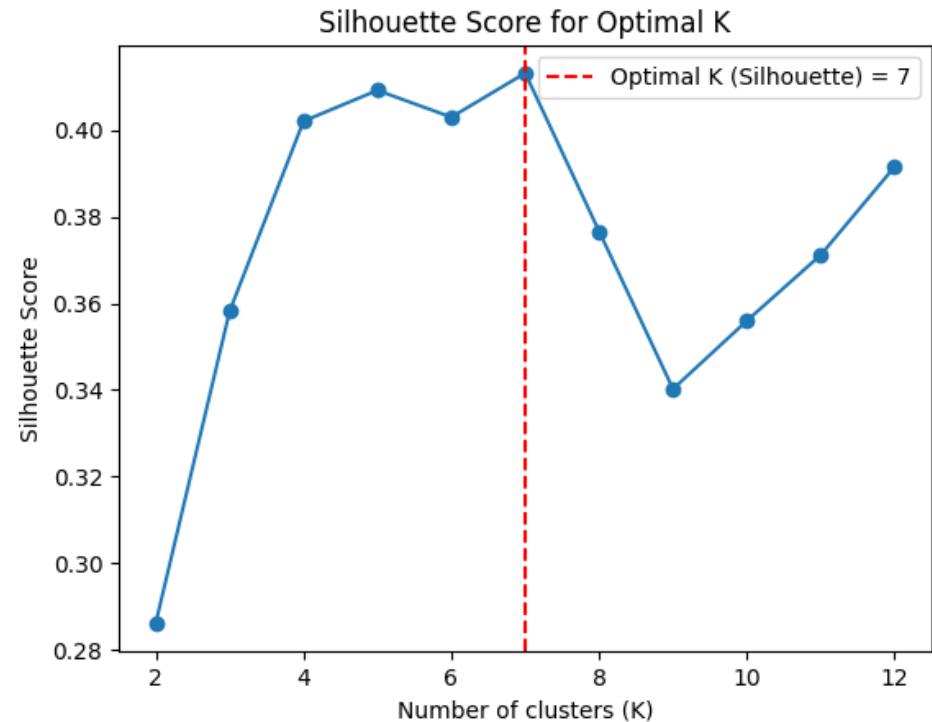
$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}$$

Department of CSE, CUET

Silhouette Score

Department of CSE, CUET
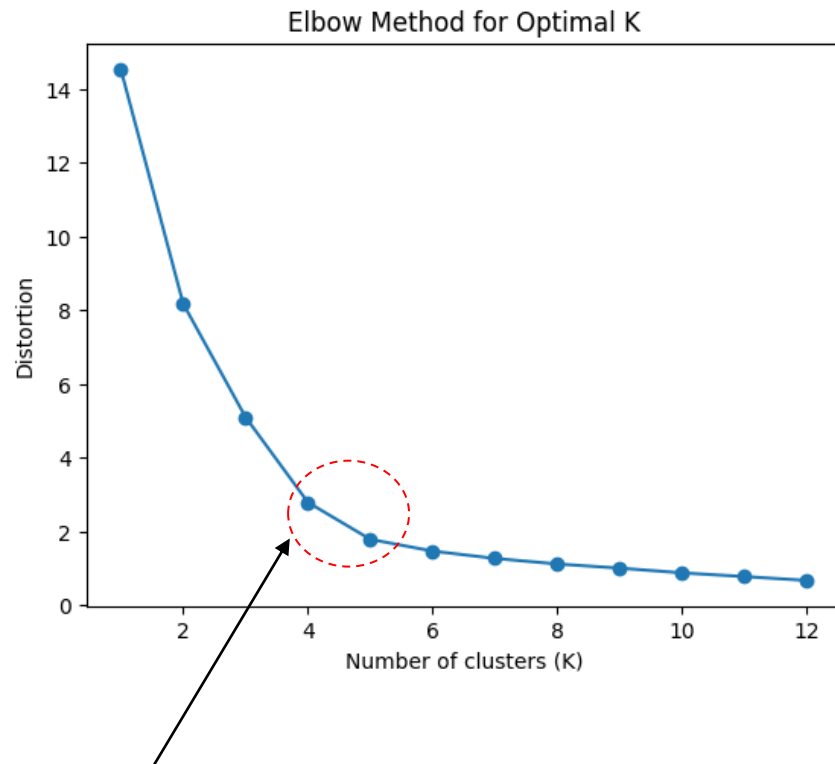
Analytical Result on Silhouette Method and Elbow Method

No clear distortion point hence no precise optimal value

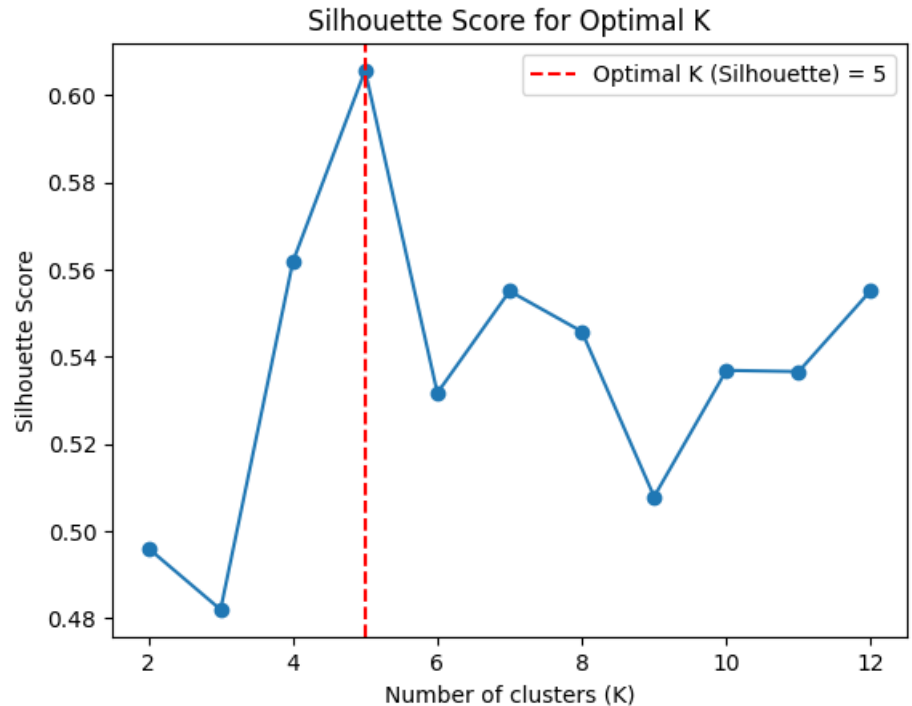Source: Analysis on UCF11/biking/v_biking_05_04.mpg video

Department of CSE, CUET

Analytical Result on Silhouette Method and Elbow Method

No clear distortion point hence no precise optimal value

Source: Analysis on UCF11/trampoline_jumping/v_jumping_02_06.mpg video

Department of CSE, CUET