

# Choosing the Optimal Value of K

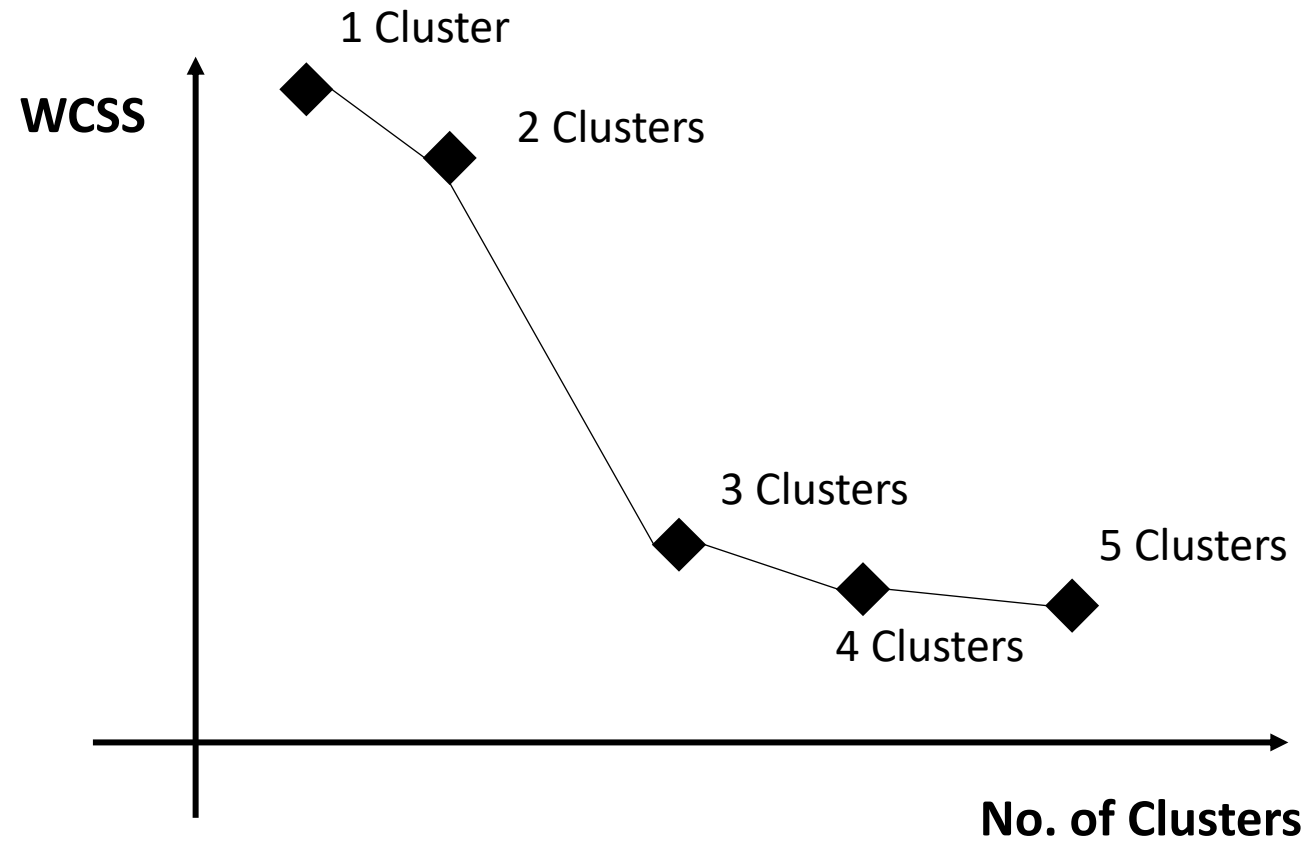
## Elbow Method & Silhouette Score

Dr. Muhammad Wasim

# Elbow Method

- Pick range of candidate values (e.g. 1 to 10)
- Calculate average distance from centroid for each value
- Plot and find the elbow

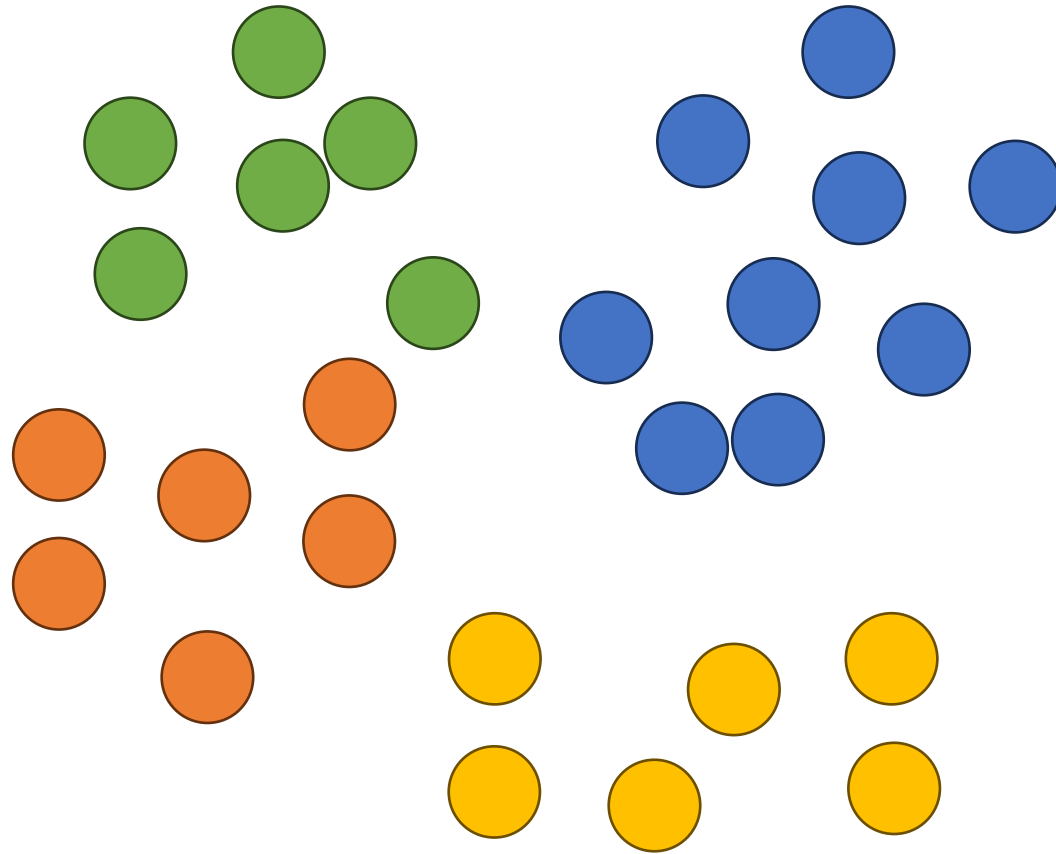
# Elbow Method



# Silhouette Method

- Uses the idea that the clusters should have **high cohesion in every cluster and high separation within different clusters.**
- Pick a range of candidate values of K (e.g. 1 to 10)
- Plot silhouette for each value of K
- Ideal value for silhouette = 1
- Worst value of silhouette = -1

# Silhouette Coefficient

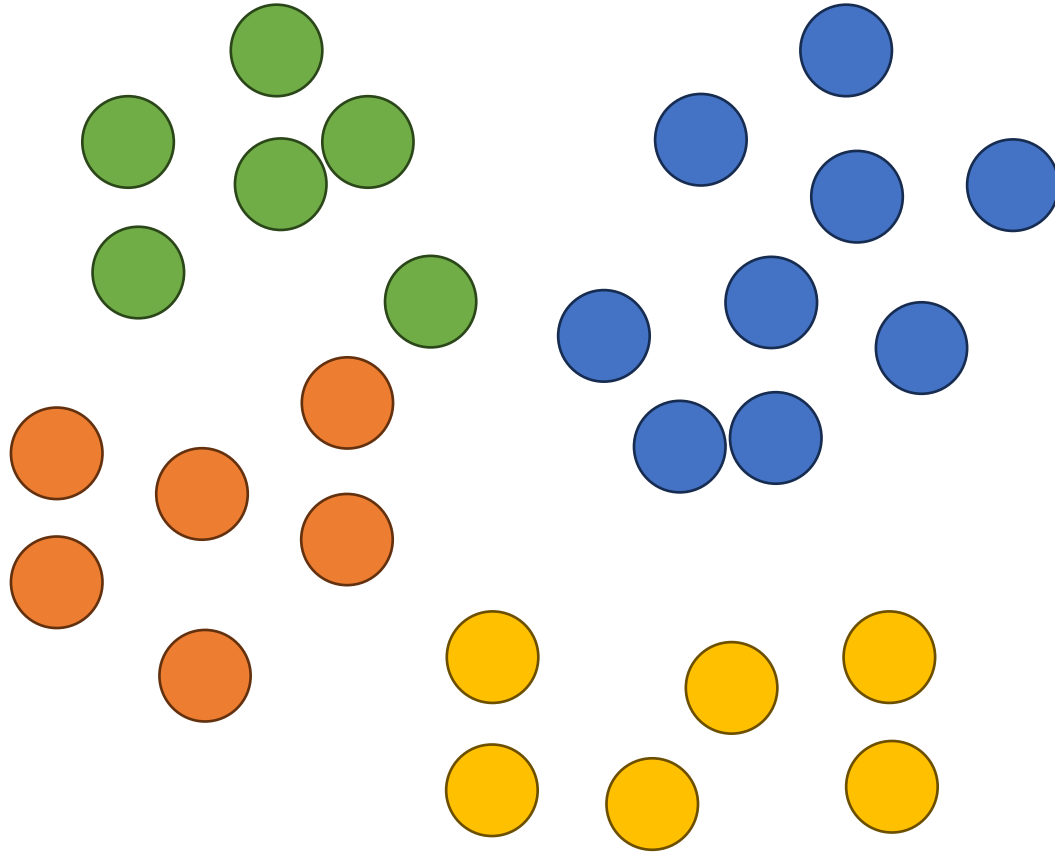


# Silhouette Coefficient

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))}$$

- $a$  = average distance inside the cluster
- $b$  = average distance to other nearest cluster
- Ideally **a** should be close to **zero** and **b** should be close to **infinity**
- The answer will be between -1 and 1

# Silhouette Plot



We calculate  $s(i)$  for each example and average over all the examples

# Silhouette Score

Average Silhouette Score	Cluster Quality
$0.7 < x < 1$	Strong Structure
$0.5 < x < 0.7$	Reasonable Structure
$0.25 < x < 0.5$	Weak Structure
$x < 0.25$	No Substantial Structure