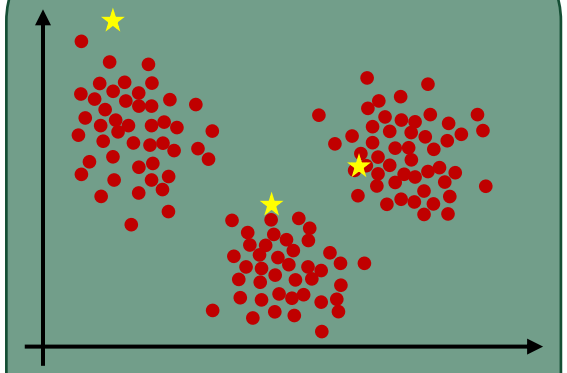
K-means Clustering

Sungyeon Lim

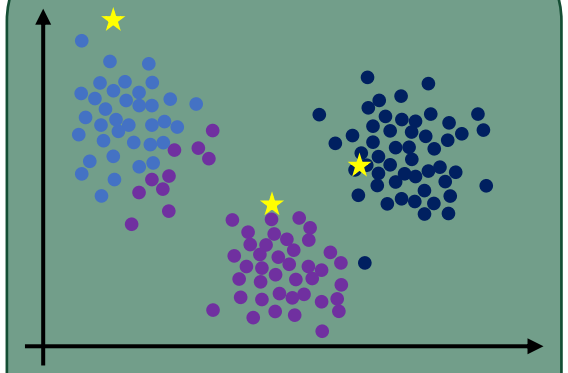
Contents

01 K-means Clustering | Algorithm



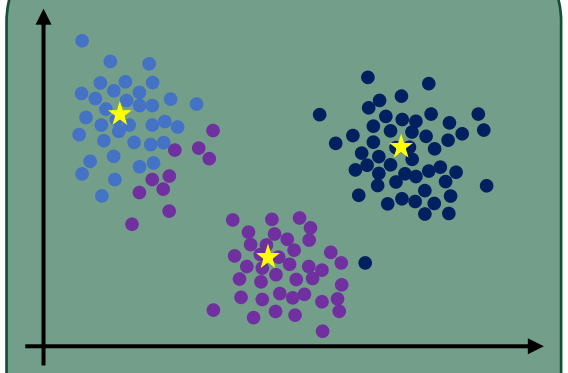
K centroids are randomly selected

01



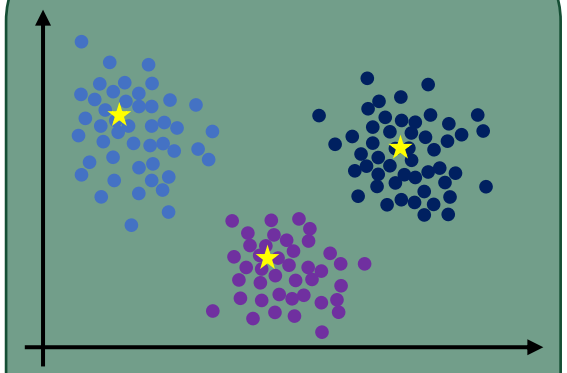
Each data point is assigned to the closest centroid

02



The centroid for each cluster is recomputed

03



The data points are reassigned to the closest centroids

04

This process continues until the centroids no longer change

05

01 K-means Clustering | Advantages and Disadvantages

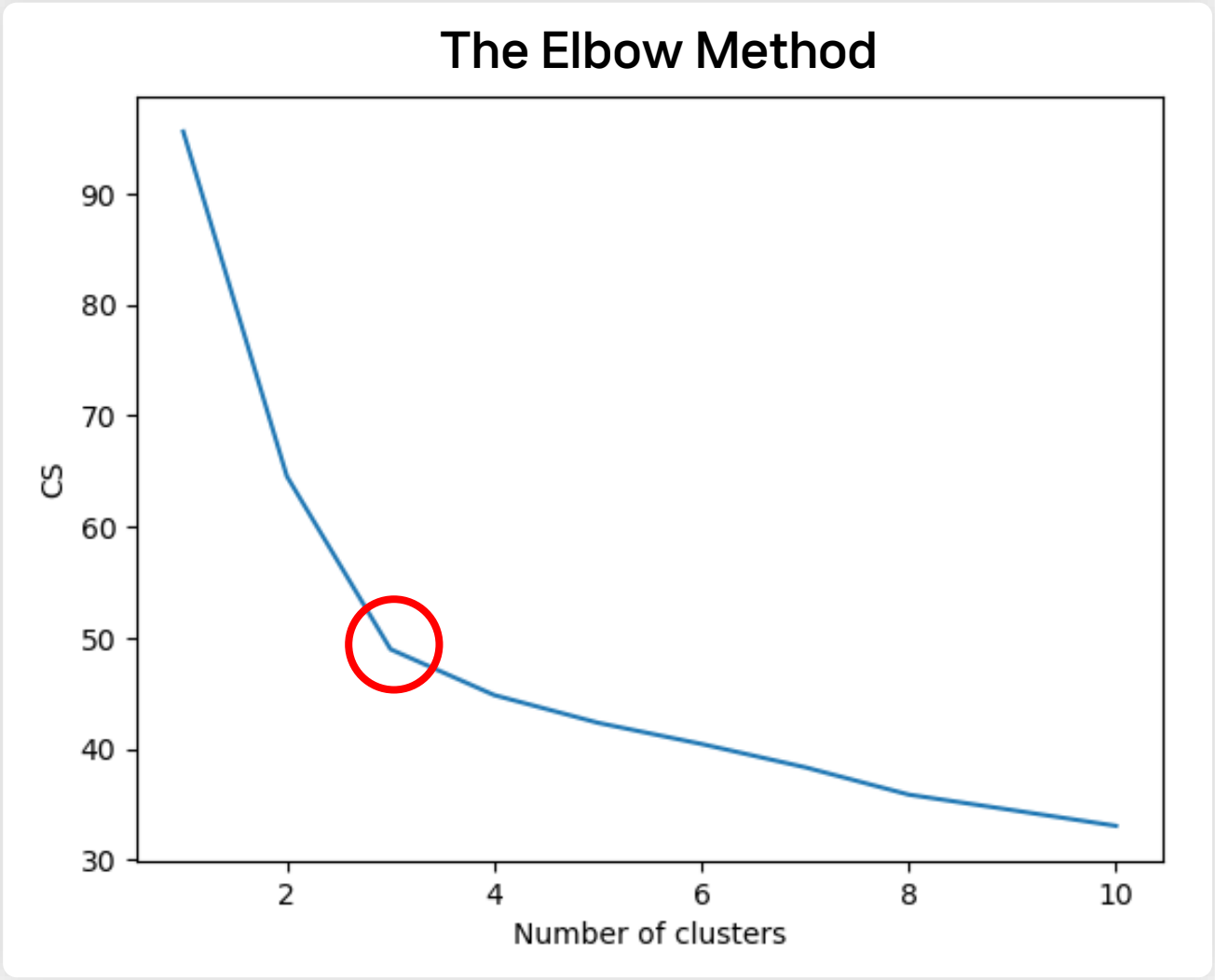
Advantages

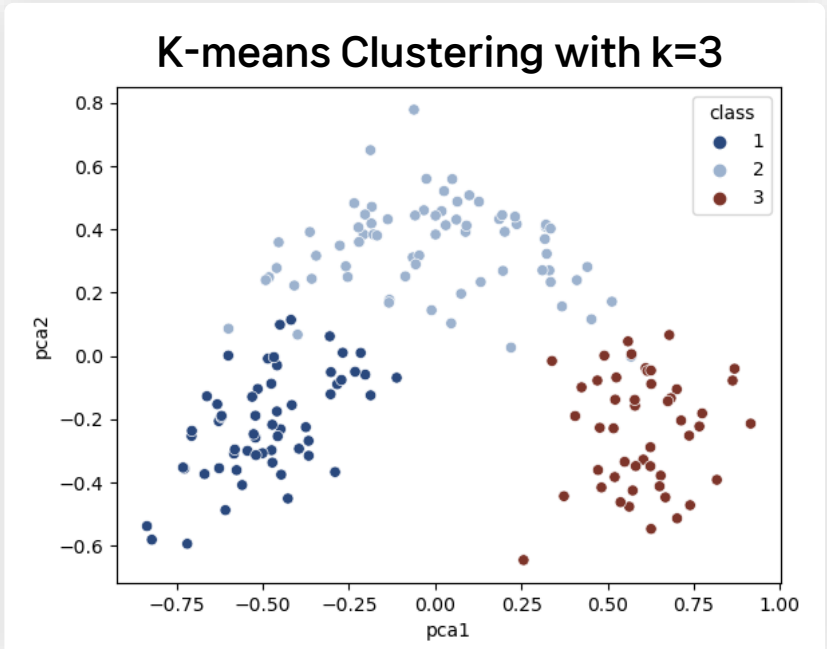
1. Simple and fast to compute, even with large datasets
2. Guarantees convergence
3. Adapts unseen data points well

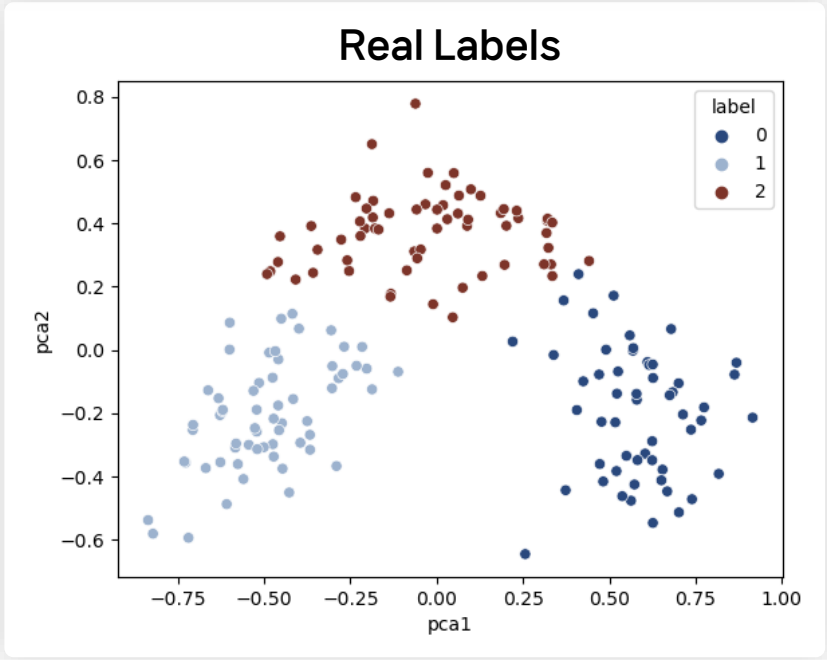
Disadvantages

1. Requires the number of clusters k to be specified manually
2. Sensitive to the initial centroids
3. Makes a restrictive assumption that clusters are spherical and equally sized

02 Evaluation | The Elbow Method

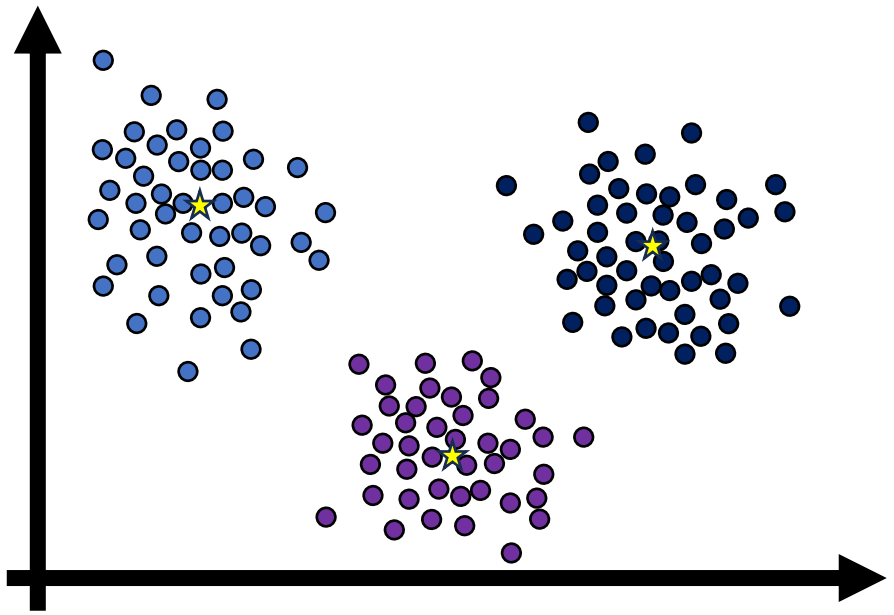






03 Conclusion

1. The most commonly used clustering algorithm is k-means clustering due to its simplicity and speed
2. It requires k to be assigned manually, where the optimal k can be found by the elbow method
3. It is sensitive to initial clusters
4. It assumes that the clusters are spherical and equally sized



Thank you for listening