K-Means Clustering - Interview Questions & Answers

1. How does K-Means work?

K-Means partitions data into k clusters by iteratively assigning points to the nearest centroid and updating centroids as the mean of assigned points until convergence.

2. What is the Elbow method?

The Elbow method plots inertia vs number of clusters (k). The point where the curve bends (the elbow) is chosen as the optimal k.

3. What are the limitations of K-Means?

- Requires specifying k in advance. - Assumes spherical clusters of similar size. - Sensitive to initialization and outliers.

4. How does initialization affect results?

Poor initialization can lead to bad clustering results. Using k-means++ or multiple runs improves stability.

5. What is inertia in K-Means?

Inertia is the sum of squared distances between data points and their nearest centroid. Lower inertia indicates tighter clusters.

6. What is Silhouette Score?

Silhouette score measures how similar a point is to its own cluster compared to other clusters. Values range from -1 to +1, with higher values indicating better separation.

7. How do you choose the right number of clusters?

Use methods like Elbow method, Silhouette score, domain knowledge, or gap statistic.

8. What's the difference between clustering and classification?

Clustering is unsupervised learning with no labels, grouping data by similarity. Classification is supervised learning that requires labeled training data.