

K-Means Clustering - Notes

K-Means is an unsupervised machine learning algorithm used for clustering data into k groups. The algorithm aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centroid).

Steps in K-Means Algorithm:

- 1 Choose the number of clusters (k).
- 2 Initialize cluster centroids (randomly or using K-Means++).
- 3 Assign each data point to the nearest centroid.
- 4 Update centroids by calculating the mean of assigned points.
- 5 Repeat assignment and update until convergence or max iterations.

Features of K-Means:

- Unsupervised learning algorithm.
- Centroid-based clustering method.
- Distance-based (commonly Euclidean distance).
- Requires user to predefine number of clusters (k).
- Iterative process until convergence.
- Efficient and scalable for large datasets.
- Assumes spherical, evenly sized clusters.
- Sensitive to feature scaling, so normalization is required.

Key Terms:

- Cluster: A group of similar data points.
- Centroid: The mean point of a cluster.
- Inertia (WCSS): Sum of squared distances of points to their nearest cluster center.
- Iteration: One cycle of assignment + centroid update.
- Convergence: When cluster assignments stop changing.
- Initialization: Method of choosing initial centroids (Random / K-Means++).

Important Parameters (Scikit-Learn KMeans):

- n_clusters: Number of clusters (k).
- init: Method for initialization ('random' or 'k-means++').
- n_init: Number of times algorithm will run with different centroid seeds.
- max_iter: Maximum number of iterations per run.
- tol: Convergence threshold.
- random_state: Ensures reproducibility.

Outputs from K-Means:

- cluster_centers_: Final centroid coordinates.
- labels_: Cluster assignment for each data point.
- inertia_: Final WCSS value.

Applications:

- Market segmentation (customer grouping).
- Image compression (color quantization).
- Document clustering (topic grouping).
- Anomaly detection (points far from clusters).