I. Types of Clustering

Partitional clustering: splits data into distinct, non-overlapping groups where each point belongs to exactly one cluster.

Hierarchical clustering: builds a tree of nested clusters that show relationships at multiple levels of granularity.

Density-based clustering: forms clusters as dense regions of points separated by areas of low density, handling noise well.

Soft (probabilistic) clustering: assigns each point to clusters with certain probabilities instead of a single label.

II. K-means clustering - Lloyd's Algorithm

K-means clustering is an algorithm that groups data into k clusters by minimizing the distance between points and their cluster centers (means).

Algorithm Steps:

1. Choose k: Decide how many clusters you want.
2. Initialize centers: Randomly pick *k* points as the initial cluster centers (called centroids).
3. Assign points: For each data point, find the closest centroid (using Euclidean distance) and assign the point to that cluster.
4. Update centers: Recalculate each centroid as the mean of all points in that cluster. Now we forget about each datapoints previous cluster, and assign it to whichever centroid is closest.
5. Repeat: Keep reassigning and updating until the centroids stop changing (the algorithm converges).

III. K-means Cost Function

In K-means, the cost function measures how well the data points fit into their assigned clusters -- it's aiming to minimize intra-cluster distance (distance between points within a cluster).

Cost function: $\sum_{i=1}^{k} \sum_{x \in Ci} d(x, \mu i)^2$

k = number of clusters
Ci = set of points in cluster i
x = a datapoint
μi = the mean (center) of cluster i
$d(x, \mu i)^2$ = squared euclidean distance between the datapoint and its cluster center

Basically: it adds up all the squared distances between the point and its cluster center (this is the result of the cost function). Smaller cost = points are close to their centers = better clustering. K-means algorithm keeps updating assignments and centers to reduce this cost until it can't get any lower. **The cost function is evaluated after each full iteration to check if it's still decreasing.**

IV. K-means and Convergence

K-means *always converges* because each iteration reduces the cost function and there are only a finite number of possible cluster assignments. However, it *does not always converge to the optimal solution* -- it can get stuck in a local minimum depending on how the initial cluster centers are chosen. In other words, it will always stop, but not always at the best clustering.