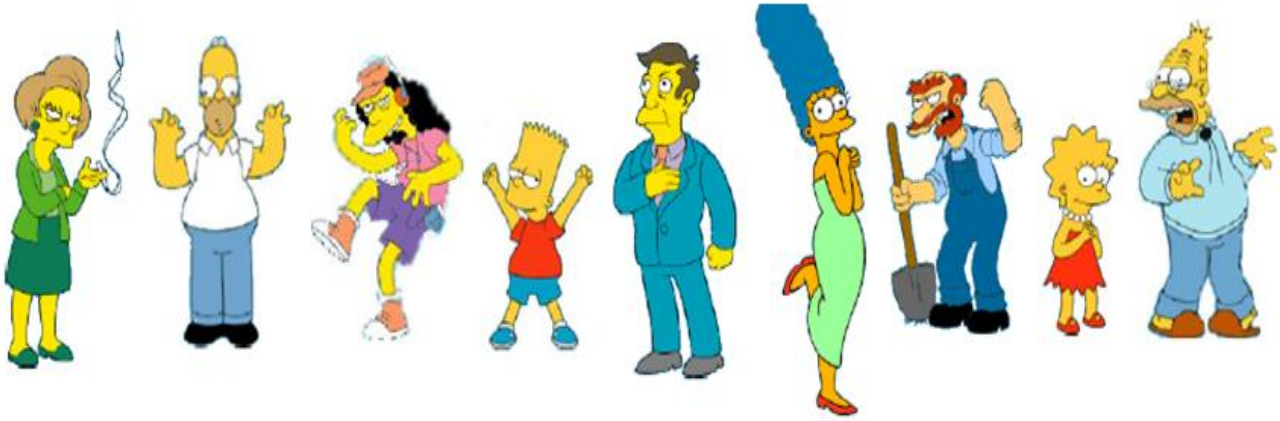
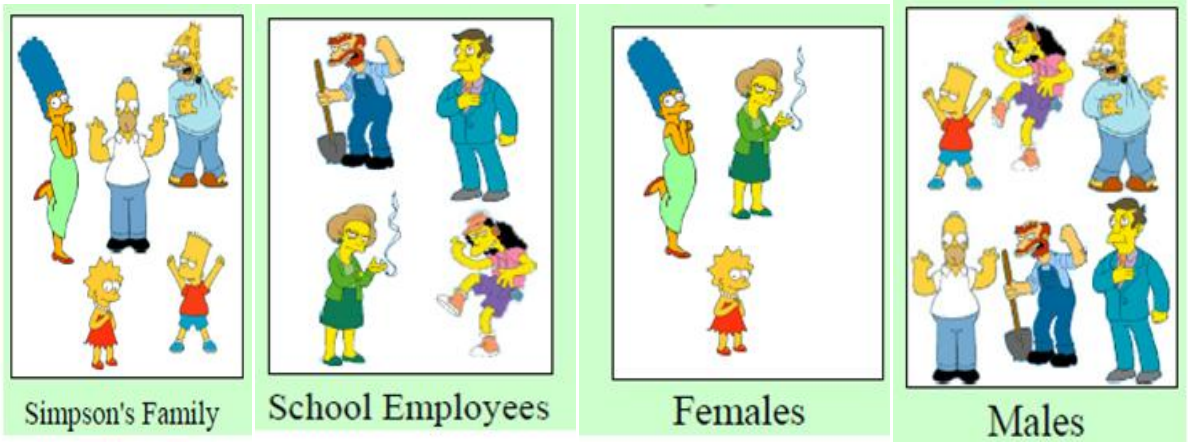
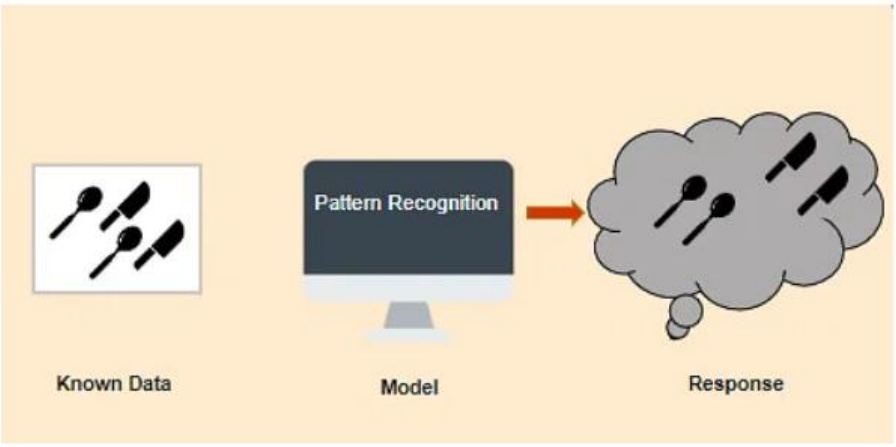
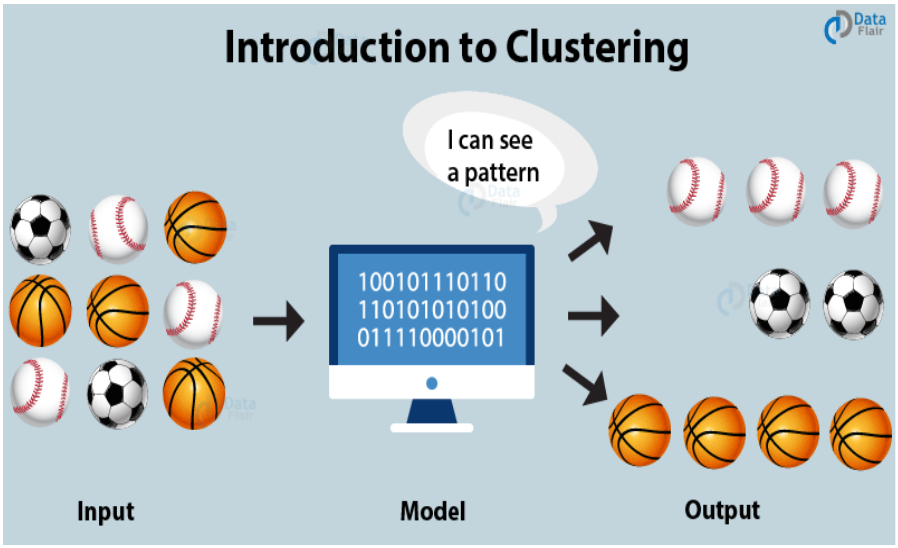
Unsupervised Learning?

- In UL, the machine uses unlabeled data and learns on itself without any supervision. The machine tries to find a pattern in the unlabeled data and gives a response.







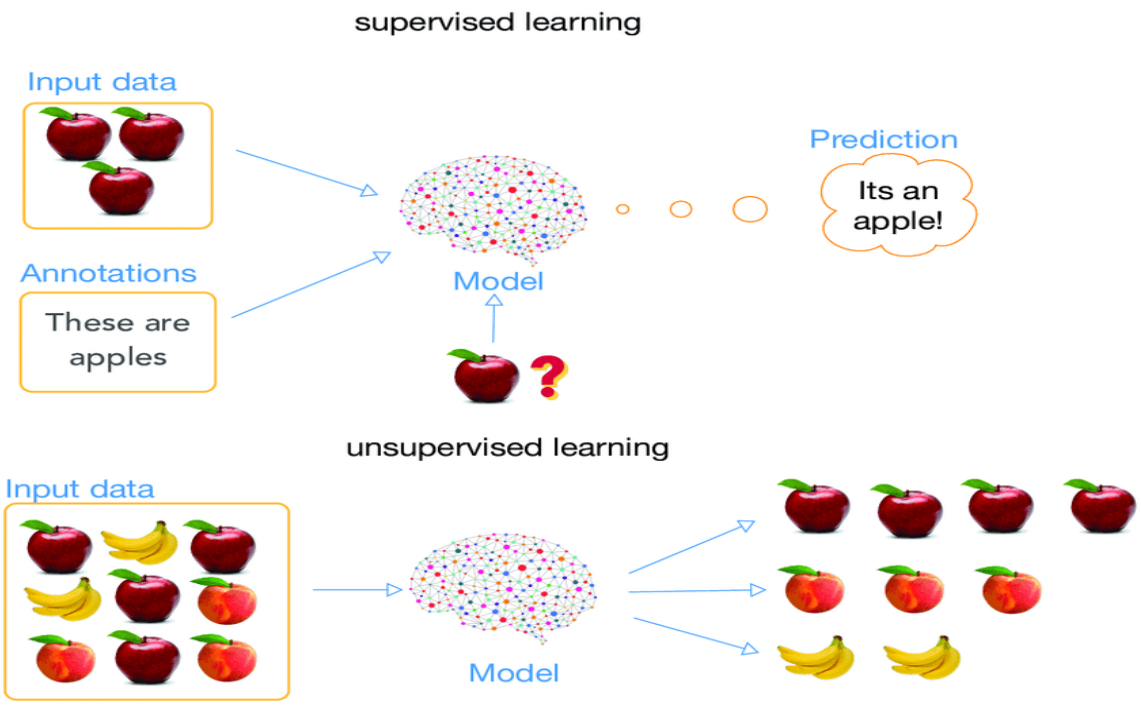


Unsupervised learning can be further grouped into types:

1. Clustering
2. Dimensionality Reduction
3. Association

Applications of Unsupervised Learning

- Customer segmentation
- Recommendation Engines
- Identifying Accident Prone Areas
- Anomaly detection
- Measuring academic results
- Real estate



K-means Clustering:

- K-means is a centroid-based algorithm or a distance-based algorithm, where we calculate the distances to assign a point to a cluster.
- In K-Means, each cluster is associated with a centroid.

Steps:

Step 1: Choose the number of clusters *k*

Step 2: Select k random points from the data as centroids

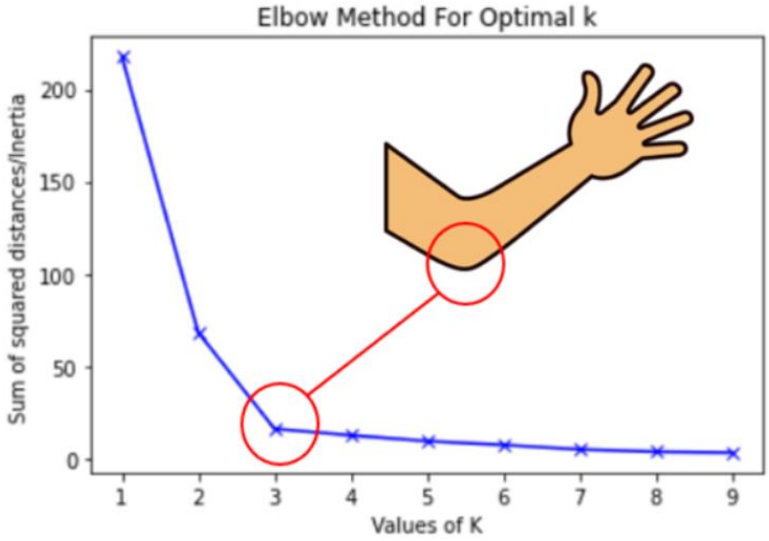
Step 3: Assign all the points to the closest cluster centroid

Step 4: Recompute the centroids of newly formed clusters

Step 5: Repeat steps 3 and 4

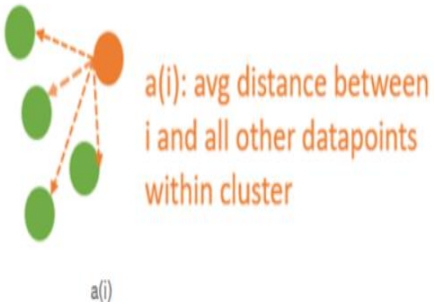
How to choose K value:

Elbow method:



Silhouette Score

$$s = (b - a) / \max(a, b)$$



a(i)

b(i)