BI

Q9) classification

In machine learning, classification is a predictive modeling technique where a model learns to assign input data to predefined categories or classes. It's a type of supervised learning where the model is trained on labeled data and then used to predict the class of new, unseen data. Essentially, classification algorithms learn to recognize patterns in data and use those patterns to categorize new data points

**🧩 Types of Classification:**

**1. Binary Classification**

* Only **two classes**
* Example: Disease (Yes/No), Spam (Spam/Not Spam)

**2. Multiclass Classification**

* More than **two classes**
* Example: Iris flowers (3 types), Digit recognition (10 digits)

**3. Multilabel Classification**

* Each instance can belong to **multiple classes**
* Example: Tagging a photo with “beach”, “sunset”, and “vacation”

**Clustering :**

''''🧪 How It Works:

1 Choose K (number of clusters).

2 Randomly initialize K centroids.

3 Assign points to the nearest centroid.

4 Recompute the centroids.

5 Repeat steps 3–4 until centroids stabilize.'''

Clustering is an **unsupervised learning technique** used to group similar data points together based on certain characteristics. Unlike classification, clustering doesn't rely on predefined labels. Instead, the algorithm tries to identify inherent structures in the data.

**🔧 Types of Clustering (Brief Recap)**

**1. Centroid-Based Clustering:**

* **K-Means**: This is the most common centroid-based clustering algorithm where the center (centroid) of each cluster is computed, and data points are assigned to the nearest centroid.

**2. Density-Based Clustering:**

* **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: It clusters data based on the **density of data points**, and it can also detect **outliers**.

**3. Hierarchical Clustering:**

* **Agglomerative**: Builds a tree-like structure (dendrogram) from the bottom up, progressively merging the most similar clusters.
* **Divisive**: Starts with one cluster and recursively divides it.

**4. Model-Based Clustering:**

* **Gaussian Mixture Model (GMM)**: Assumes that the data is a mixture of several Gaussian distributions, and it performs clustering by assigning probabilities to data points for belonging to each cluster.

**5. Grid-Based Clustering:**

* This method divides the data space into a finite number of cells or grids and performs clustering based on grid-based density. It’s **fast** and efficient but doesn’t handle irregular-shaped clusters well.

DL\_code