K-Nearest Neighbors (KNN) — Similarity Search

1. Goal of KNN

We want to classify a new data point by looking at the most similar known points around it.

2. Distance Calculation (similarity)

To measure similarity, we use Euclidean Distance:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2}$$

Where:

- x = test point (unknown label)
- x_i = a training point (known label)
- n = number of features
- Smaller distance → more similar

3. Selecting Nearest Neighbors (K closest)

We compute distance to every training sample, then sort:

$$\text{nearestK} = \text{argmin}(d_1, d_2, \ldots, d_N)$$

Where N = number of training points.

4. Majority Voting

Take the class that appears most frequently among these K neighbors.

$$\hat{y} = \text{mode}(y_K)$$

Where:

- y_K = labels of nearest neighbors
- $\hat{y}$ = predicted label

KNN predicts the class of a point by looking at its K Nearest Neighbors using distance and choosing the major class.