07/04/25 dat.5: KNN Algorithm

→ Input:
Dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
Test data X_{test}
Number of neighbours - K

→ Output:
Predicted labels Y_{pred} for test data
Classification accuracy.

Algorithm:

1. Load the Dataset
- Load the Iris dataset containing
Input features X
Target labels y

2. Split the Dataset
= Divide the dataset into:
Training set: X_{train}, y_{train}
Test set: X_{test}, y_{test}
- Use a fixed random seed for reproducibility

3. Initialize the KNN classifier
- set the number of neighbours K

4. Train the Classifier.
- Store the training data X_{train}, y_{train}

5. Predict for Each Test Instance

For each test sample $x \in X_{test}$:

a. Compute Euclidean Distance to all training samples:

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

b. Identify the k Nearest Neighbors.
 - Select the k smallest distances

c. Extract Labels of the k nearest neighbors.

d. Determine the Majority class among these labels.

e. Assign the Majority class as the predicted label for x.

6. Evaluate Accuracy.
 - Compare actual and predicted labels
 - Compute accuracy

$$\text{Accuracy} = \frac{\text{No of correct predictions}}{\text{Total test samples}} \times 100$$

7. Display Results

 - Print :
 Predicted labels
 Actual labels
 classification accuracy.