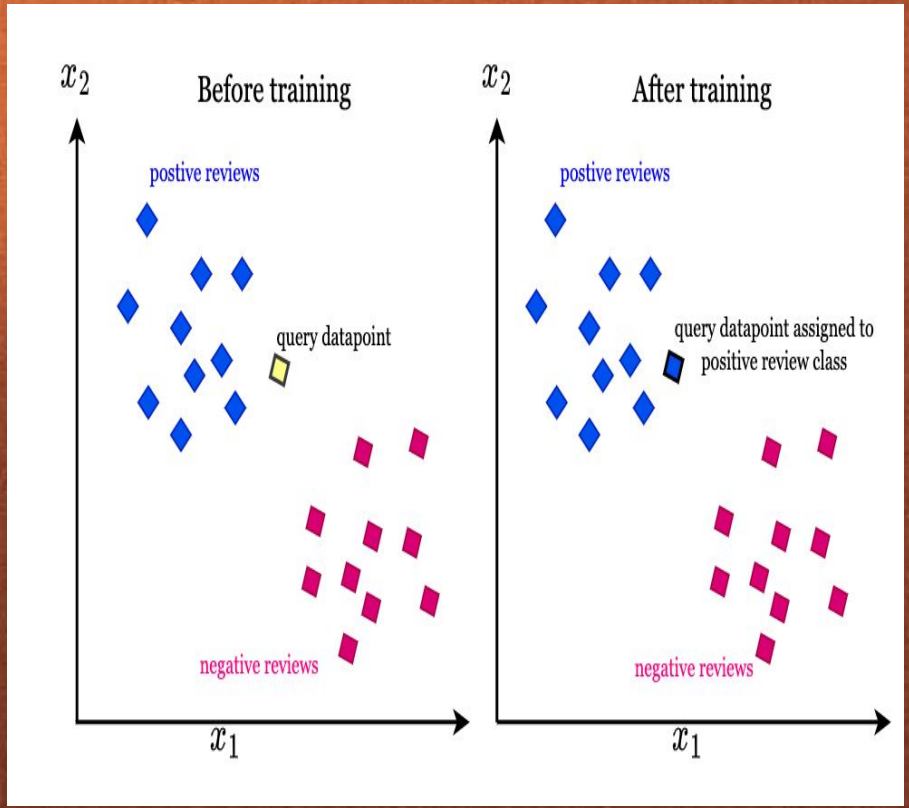
K-Nearest Neighbor Regression & Classification

Machine Learning and Data Science Club

What is the K-Nearest Neighbor Algorithm in ML?

- K-Nearest Neighbors (KNN) is a machine learning algorithm that makes predictions based on the most similar data points in the training set.
- It works by finding the K closest neighbors to a new data point using a distance metric, then combining their outcomes by majority vote for classification or averaging for regression.
- Because KNN relies on similarity rather than learning a model, it is simple to understand but sensitive to feature scaling and the choice of K.



An intuitive Example

Scenario:

- We want to guess what might a new member of the MLDS Clubs major is.
- We know that many members of MLDS come from a tech related major.

How KNN Works:

- Step 1: Look at the 3 members of the MLDS club (K = 3).
- Step 2: Check which major is most common among them.
- Step 3: Predict that the new student will likely have that major.
-

Key Idea:

- “Students with similar interests tend to choose similar majors.”

Classification vs Regression in K-NN

	Classification	Regression
Output	Class label	Continuous value
Aggregation	Majority vote	Mean / weighted mean
Examples	-Handwriting recognition -Spam detection -Disease diagnosis	-House price prediction -Recommendation scores -Demand estimation

When to employ the K-NN Algorithm

- When your dataset is not too large.
- When the relationship between features and the target is nonlinear.
- When you want a simple model that’s easy to understand.
- When all features are on a similar scale.
- When similar data points are expected to have similar outcomes.

Limitations of the K-NN Algorithm

- Can be slow when the dataset is large because it compares every point.
- Uses a lot of memory since it stores all training data.
- Performance depends heavily on the choice of K and distance metric.
- Sensitive to feature scaling — unscaled data can give poor results.
- Not ideal for real-time predictions where fast responses are needed.

Key Parameters in a K-NN Model

- **Choose the value of K:**

A small K can overfit the data, while a large K can oversimplify the model.

- **Select a distance metric:**

Euclidean distance measures the straight-line distance between two points.

Manhattan distance measures distance by moving along horizontal and vertical paths, like navigating city streets.

- **Evaluate model performance:**

Use accuracy for classification and error metrics (MSE, RMSE) for regression.