

What is K-NN?

K-Nearest Neighbors (KNN) is a **supervised machine learning algorithm** used for both **classification** and **regression**. It works by finding the **K closest data points** (neighbors) to a new input and predicts the label based on:

- **Majority vote** (for classification)
- **Average value** (for regression)

It's simple, effective, and requires **no training** — the algorithm makes predictions based on raw data during testing time.

Basic Idea

Imagine you're new to a neighborhood and want to know if a local restaurant is good. You ask a few **neighbors closest to it**. If most say it's good — you trust the majority.

That's how KNN works: it trusts nearby data points.

How does K-NN work?

The K-NN working can be explained on the basis of the below algorithm:

Step 1: Select the number **K** of the neighbors (select odd numbers).

Step 2: Calculate the **Euclidean distance** of **K number of neighbors**.

Step 3: Take the **K nearest neighbors** as per the calculated Euclidean distance. **Sort** the values in ascending order.

Step 4: Among these **sorted K neighbors**, **count the number** of the data points in each category.

Step 5: Assign the new data point to the **category for which the number of neighbors is maximum**.

Step 6: Our **model is ready**.

Quick Example

Let's say we have these fruit data:

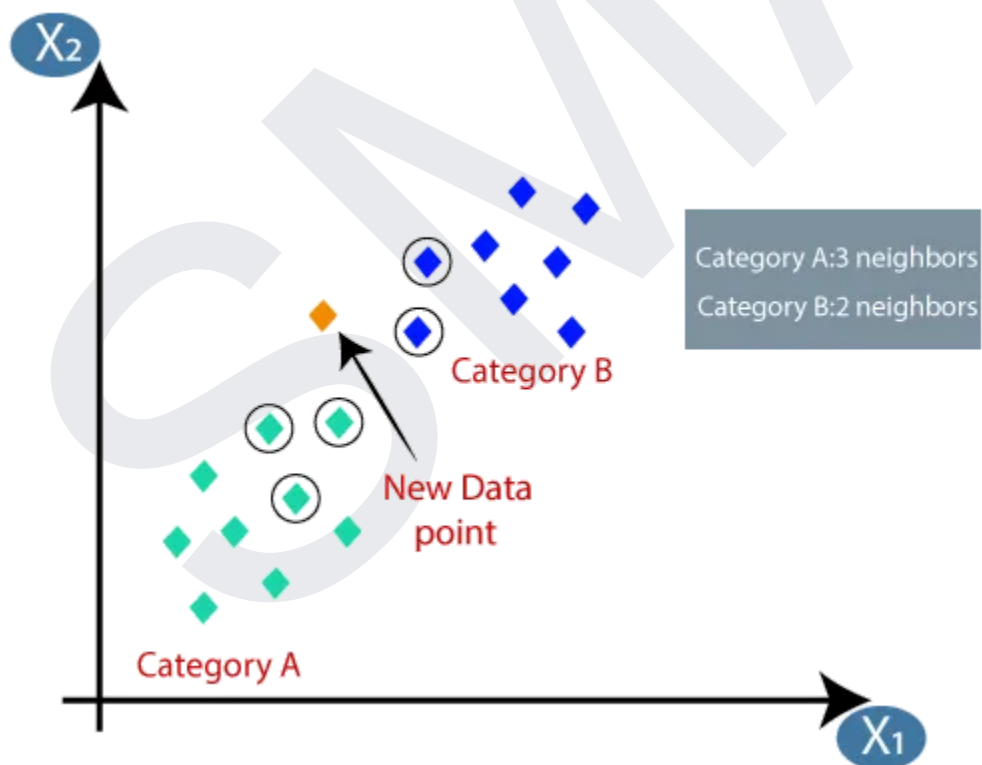
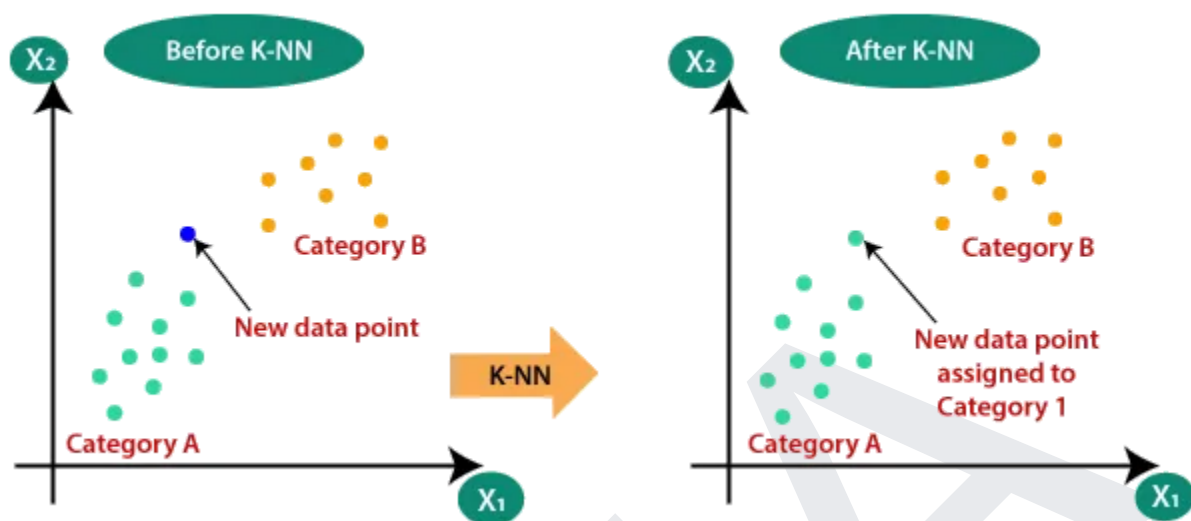
Weight	Color	Label
150g	Red	Apple
170g	Red	Apple
140g	Green	Apple
130g	Yellow	Banana
160g	Yellow	Banana

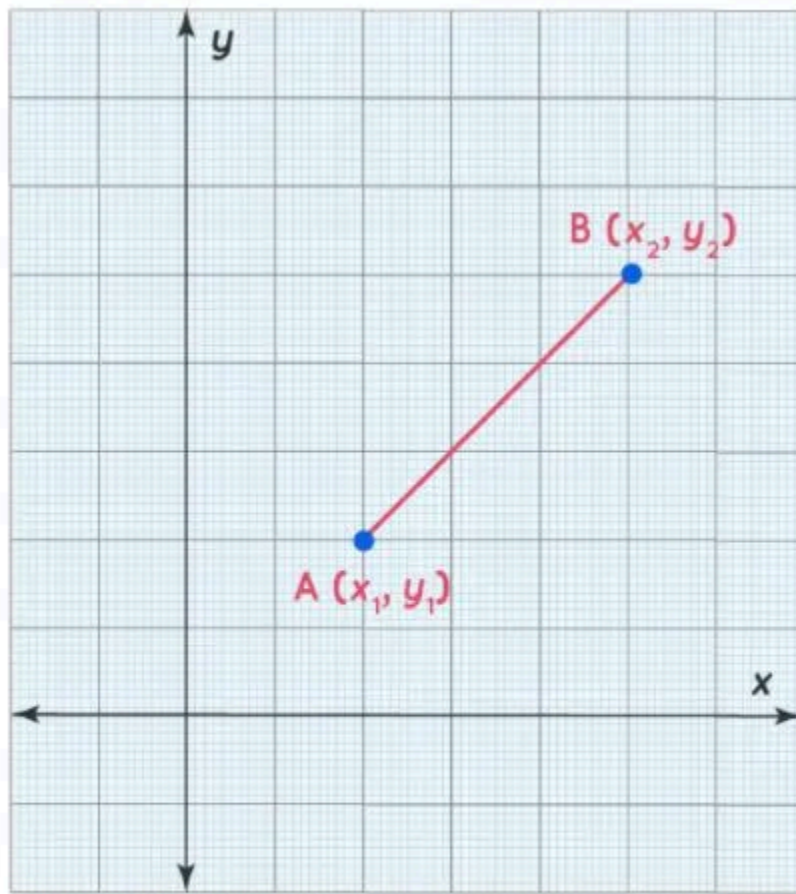
Now we have a new fruit: **145g, Red**. Choose **K = 3**.

- The 3 closest fruits are: Apple, Apple, Banana
- Apple wins by majority vote

Things to Keep in Mind

- **K matters:** Too small = noise, too big = diluted decision. So, Optimal value of K is the **square root** of the **total datapoint**
- **Lazy learner:** No training, all work happens during prediction
- **Distance-based:** Euclidean is common, but other distances like Manhattan or cosine can be used
- **Not great with high dimensions:** It suffers from the “curse of dimensionality”





$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$