# **k-Nearest Neighbors (k-NN)**

k-Nearest Neighbors (k-NN) is an algorithm that can be used for both classification and regression tasks. The core idea of k-NN is to predict the output based on the k nearest data points in the feature space. Here's how k-NN works in both contexts:

**k-NN for Classification**

**How it Works:**

* When performing classification, k-NN assigns the label based on the majority class among the k nearest neighbors of the query point.
* The distance between points is usually measured using metrics like Euclidean distance, Manhattan distance, or Minkowski distance.

**Example:**

* Suppose you want to classify a new data point as either 'cat' or 'dog'. You look at the k nearest points in the training data, count how many are labeled 'cat' and how many are labeled 'dog', and assign the label that is more frequent among these k neighbors.

**Algorithm Steps:**

1. Choose the number of neighbors k.
2. Calculate the distance between the new point and all points in the training data.
3. Identify the k nearest neighbors.
4. Assign the label based on the majority vote among these neighbors.

**k-NN for Regression**

**How it Works:**

* When performing regression, k-NN predicts the value of the query point by averaging the values of the k nearest neighbors.
* The distance metric used is the same as in classification (e.g., Euclidean distance).

**Example:**

* Suppose you want to predict the price of a house. You look at the k nearest houses in the training data, and the predicted price is the average price of these k houses.

**Algorithm Steps:**

1. Choose the number of neighbors k.
2. Calculate the distance between the new point and all points in the training data.
3. Identify the k nearest neighbors.
4. Calculate the average of the target values of these k neighbors to make the prediction.

**Summary of k-NN in Both Contexts**

* **Classification:**
  + **Output:** Categorical label.
  + **Prediction:** Based on majority vote among k nearest neighbors.
  + **Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score, ROC-AUC.
* **Regression:**
  + **Output:** Continuous value.
  + **Prediction:** Based on the average of the target values of k nearest neighbors.
  + **Evaluation Metrics:** Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared.

In summary, k-NN is a versatile algorithm that can handle both classification and regression tasks by adjusting how the predictions are made based on the k nearest neighbors.