COMMON CLASSIFICATION ALGORITHMS

1. *k*-Nearest Neighbour (*k*NN)
2. Support Vector Machine (SVM)
3. Decision tree
4. Naïve Bayes classifier
5. Random Forest Classifier

**K-Nearest Neighbours (KNN) Algorithm**

The K-Nearest Neighbours (KNN) algorithm is one of the simplest and most widely used machine learning algorithms for classification and regression problems. It is based on the concept of similarity (or closeness) between data point

**How KNN Works:**

1. Training Phase:
   * Unlike other machine learning algorithms, KNN does not explicitly learn from the training data.
   * It simply stores the entire dataset and waits for new test data to make predictions.
2. Prediction Phase:
   * When new data is given, KNN finds the K data points (neighbours) that are closest to the new data point.

It then assigns the most common class (for classification) or the average value (for regression) among these neighbours to the new da

**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
* Step-2: Calculate the Euclidean distance of K number of neighbors
* Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
* Step-4: Among these k neighbors, count the number of the data points in each category.
* Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
* Step-6: Our model is ready.

A diagram of a graph

AI-generated content may be incorrect.

* Firstly, we will choose the number of neighbors, so we will choose the k=5.
* Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:

A graph of a line with a point and a point

AI-generated content may be incorrect.

* By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:

A diagram of a network

AI-generated content may be incorrect.

* As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

Numerical Example using KNN

A screenshot of a computer

AI-generated content may be incorrect.



A table with numbers and symbols

AI-generated content may be incorrect.



Let us take K value as K=3 then ,it will consider 3 Nearest Neighbours

A table with numbers and a number

AI-generated content may be incorrect.

**Advantages of KNN:**

* Easy to understand and implement.
* No need for a training phase (lazy learning).
* Works well with small datasets and non-linear patterns.

**Disadvantages of KNN:**

* Slow for large datasets as it needs to compute distances for every query.
* Sensitive to irrelevant features and data scaling.
* Performance depends on choosing the right value of K.

**Choosing the Right Value of K:**

* If K=1, the model is too specific and prone to noise (overfitting).
* If K is too large, the model becomes too generalized (underfitting).
* A good practice is to try different values and use cross-validation to find the best K.

**Practical Applications of KNN:**

* Handwriting recognition (e.g., digit recognition).
* Recommender systems (suggesting products based on similarity).
* Credit risk analysis (approving or denying loan applications).
* Medical diagnosis (predicting diseases based on symptoms).

Reference: <https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4>