**K-Nearest Neighbors Algorithm:**

The k-nearest neighbors (KNN) algorithm is a data classification method for estimating the likelihood that a data point will become a member of one group or another based on what group the data points nearest to it belong to.

The k-nearest neighbor algorithm is a type of **supervised machine learning algorithm** used to solve classification and regression problems. However, it's mainly used for classification problems.

It's considered a non-parametric method because it doesn’t make any assumptions about the underlying data distribution. Simply put, KNN tries to determine what group a data point belongs to by looking at the data points around it.

Consider there are two groups, A and B.

To determine whether a data point is in group A or group B, the algorithm looks at the states of the data points near it. If the majority of data points are in group A, it's very likely that the data point in question is in group A and vice versa.

In short, KNN involves classifying a data point by looking at the nearest annotated data point, also known as the nearest neighbor.

**How does KNN work:**

As mentioned above, the KNN algorithm is predominantly used as a classifier.

Unlike classification using artificial neural networks, k-nearest neighbors classification is easy to understand and simple to implement. It's ideal in situations where the data points are well defined or non-linear.

In essence, KNN performs a voting mechanism to determine the class of an unseen observation. This means that the class with the majority vote will become the class of the data point in question.

If the value of K is equal to one, then we'll use only the nearest neighbor to determine the class of a data point. If the value of K is equal to ten, then we'll use the ten nearest neighbors, and so on.

To put that into perspective, consider an unclassified data point X. There are several data points with known categories, A and B, in a scatter plot.

Suppose the data point X is placed near group A.

We classify a data point by looking at the nearest annotated points. If the value of K is equal to one, then we'll use only one nearest neighbor to determine the group of the data point.

In this case, the data point X belongs to group A as its nearest neighbor is in the same group. If group A has more than ten data points and the value of K is equal to 10, then the data point X will still belong to group A as all its nearest neighbors are in the same group.

Suppose another unclassified data point Y is placed between group A and group B. If K is equal to 10, we pick the group that gets the most votes, meaning that we classify Y to the group in which it has the most number of neighbors. For example, if Y has seven neighbors in group B and three neighbors in group A, it belongs to group B.

**What does the KNN algorithm do:**

The KNN algorithm calculates the probability of the test data belonging to the classes of ‘K’ training data and class holds the highest probability will be selected. In the case of regression, the value is the mean of the ‘K’ selected training points.

Let see the below example to make it a better understanding.

Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

Graphical user interface, application

Description automatically generated

**In which situations KNN will be most useful:**

Classification is a critical problem in data science and machine learning. The KNN is one of the oldest yet accurate algorithms used for pattern classification and regression models.

Here are some of the areas where the k-nearest neighbor algorithm can be used:

* Credit rating: The KNN algorithm helps determine an individual's credit rating by comparing them with the ones with similar characteristics.
* Loan approval: Similar to credit rating, the k-nearest neighbor algorithm is beneficial in identifying individuals who are more likely to default on loans by comparing their traits with similar individuals.
* Data preprocessing: Datasets can have many missing values. The KNN algorithm is used for a process called missing data imputation that estimates the missing values.
* Pattern recognition: The ability of the KNN algorithm to identify patterns creates a wide range of applications. For example, it helps detect patterns in credit card usage and spot unusual patterns. Pattern detection is also useful in identifying patterns in customer purchase behavior.
* Stock price prediction: Since the KNN algorithm has a flair for predicting the values of unknown entities, it's useful in predicting the future value of stocks based on historical data.