

KNNIMPUTER

KNN is a lazy learning and non-parametric algorithm.

Imputation for completing missing values using k-Nearest Neighbors. Each sample's missing values are imputed using the mean value from $n_neighbors$ nearest neighbors found in the training set. Two samples are close if the features that neither is missing are close

Why using KNN ?

KNN is an algorithm that is useful for matching a point with its closest k neighbors in a multi-dimensional space. It can be used for data that are continuous, discrete, ordinal and categorical which makes it particularly useful for dealing with all kind of missing data.

The assumption behind using KNN for missing values is that a point value can be approximated by the values of the points that are closest to it, based on other variables.

Place's 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.

Recommendation systems: Since KNN can help find users of similar characteristics, it can be used in recommendation systems. For example, it can be used in an online video streaming platform to suggest content a user is more likely to watch by analyzing what similar users watch.

Computer vision: The KNN algorithm is used for image classification. Since it's capable of grouping similar data points, for example, grouping cats together and dogs in a different class, it's useful in several computer vision applications.

advantages

It's easy to understand and simple to implement

It can be used for both classification and regression problems

It's ideal for non-linear data since there's no assumption about underlying data

It can naturally handle multi-class cases

It can perform well with enough representative data

Disadvantages

Accuracy depends on the quality of the data.

With large data, the prediction stage might be slow.

Sensitive to the scale of the data and irrelevant features.

Require high memory – need to store all of the training data.

Given that it stores all of the training, it can be computationally expensive.