

Libraries

Pandas (read the dataset)

NumPy (array related operation)

Math (mathematical operation)

Operator (during sorting for specific key)

Read Train and Test Dataset and split the Train to X_train and Y_train

X_train contain all the data from Train Dataset except the target column

Y_train contain only the Target Column

X_test contain Test Dataset

KNN class:

Distance function:-

Calculate the Euclidean distance

$$d(x1, x2) = \sqrt{\sum_{i=1}^n (x1_i - x2_i)^2}$$

KNN Classification function:-

Iterating through all the rows in the X_train for a specific row of X_test and passing the row to distance function.

Calling the function Distance and calculating the Euclidean distance of X_train and X_test for the row passed using the above formulation.

Once all the distance are calculated for a specific row of X_test.

The list is sorted in ascending order, w.r.t distance.

The closer to a specific column data, the more chance of the test row belonging to that specific target class.

Fit function:-

Iterating through all the rows of the X_test and passing to the KNN Classification function

The sorted distances are returned with the actual target value, and the data are passed to the predicting function.

Prediction function:-

It counts the number of specific target.

According to the data sets, the target are 0 (no disease) and 1 (Heart disease)

If the number of ones is more than the number of zeros, prediction return 1

Else 0

This help us to identify the class of the test rows.