National Technology of Mexico
Technological Institute of Tijuana

ACADEMIC SUBDIRECTION
Department of Systems and Computing

SEMESTER
February - June 2021

RACE
Ing. Technologies of the Information and Communication

SUBJECT AND KEY :
Data Mining BDD-1703TI9A

STUDENT'S NAME AND REGISTRATION:
Camacho Manabe Juan Daniel 17210534
Velázquez Farrera César Alejandro 17212937

NAME OF THE JOB:
Euclidean Distances

UNIT TO BE EVALUATED
Unit III

TEACHER'S NAME:
Mc . Jose Christian Romero Hernandez

Euclidean Distance and K-NN (K Nearest Neighbors)

KNN or K-Nearest Neighbor is a method that simply searches the observations closest to the one you are trying to predict and classifies the point of interest based on most of the data around you. KNN is an algorithm that:

- It is supervised: it means that the training data sets are labeled, with the case or expected result given a “data row”.
- It is instance-based: This means that our algorithm does not explicitly learn a model (such as logistic regression or decision trees). Instead, memorize the training instances that are used as the "knowledge base" for the prediction phase.

KNN works as follows:

1. Calculate the distance between the item to be classified and the rest of the items in the training dataset.
2. Select the closest “k” elements (with less distance, depending on the function used)
3. Carry out a “majority vote” among the k points: those of a class / label that <<dominen>> will decide their final classification.

Taking into account point 3, we will see that to decide the class of a point the value of k is very important, since this will almost end up defining to which group the points will belong, especially at the "boundaries" between groups.

The most popular ways to "measure closeness" between points are the Euclidean distance.

Euclidean distance formula (1 - D)

$$d(A, B) = \sqrt{(X_B - X_A)^2}$$

Donde:

d = distance between points

A, B: Points A and B

$d\ (A, B)$: Distance between A and B

X_B: Value of X subscript B

X_A: Value of X subscript A

Euclidean distance is a positive number that measures the distance between two points on a plane (a line) .

The formula can be applied in Euclidean spaces, it means that it can be n dimensions (one dimension (1 -D), two dimensions (2 - D), three dimensions (3 - D) up to n).

Due date: May 01, 2021

Euclidean distance formula (N dimensions)

$$d(P,Q) = \sqrt{(Q_1 - P_1)^2 + (Q_2 - P_2)^2 + \cdots + (Q_n - P_n)^2}$$

Donde:

d = distance between points

X, Y: Points A and B

$d\ (X, Y)$: Distance between points A and B

Q_n: Q point value in n axis

P_n : Point value P on axis n

In third dimension (3 - D), according to the formula, a point Q has coordinates (X_Q, $Y_Q$and Z_Q).

In the fourth dimension (4 - D), according to the formula, a point Qhas the coordinates (X_Q, Y_Q, $Z_Q$and W_Q) .

Therefore, the distance between points $P$and $Q$is the length of the line that joins these points.

Due date: May 01, 2021

References

1. (2018). Clasificar con K-Nearest-Neighbor ejemplo en Python. 30 de Mayo del 2021. De Aprende Machine Learning. Sitio web: https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/
2. Pérez, Ricardo. (4 de December de 2019). Distancia euclidiana: concepto, fórmula, cálculo, ejemplo. Lifeder. Recuperado de https://www.lifeder.com/distancia-euclidiana/.

Due date: May 01, 2021