Tecnológico de Monterrey

TC3020 Machine Learning (Gpo 2)

Prof. Rafael Pérez Torres
Team #04

Luis Felipe Álvarez Sánchez	**A01194173**
Jesús Omar Cuenca Espino	**A01378844**
Juan José González Andrews	**A01194101**
Rodrigo Montemayor Faudoa	**A00821976**

Description

K-nearest neighbor is a non-parametric, classification algorithm that has an output classified by its neighboring values given by a data set. It assumes that data with similar characteristics is in close proximity and the algorithm captures the idea of them being alike regarding distance by calculating the distance between different points. The calculation between two points in a database is a measure of their dissimilarity.

In order to select the right K for the dataset, the algorithm is run several times with different values of K and choose whichever reduces the number of errors encountered while preserving its ability to make accurate predictions with the dataset. Its main advantages are its simplicity and versatility used for classification and regression, although it is generally more used for classification in the industry. Another important use of the kNN analysis is that it computes values for a continuous target.

The objective is to classify and give a value for any output that has similar characteristics of the local values in a data set, in other words, the results depend solely on the local values in a given data set. In relation to the case, we will be applying a basic version of the kNN algorithm on given datasets where we will be presenting our findings on its implementation.

Datasets

The following datasets include the findings obtained from the implementation of the code:

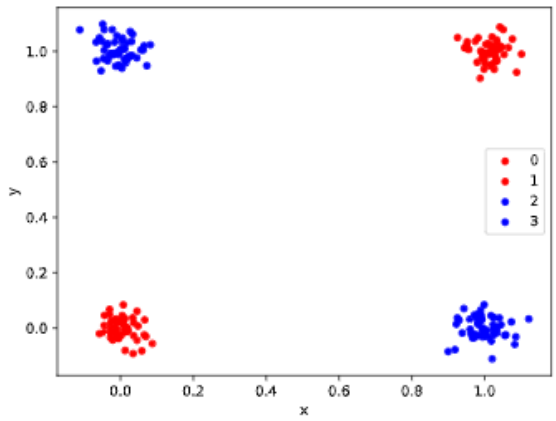
XOR dataset
[[0 0]] predicted as 0 [[0 1]] predicted as 1 [[1 0]] predicted as 1 [[1 1]] predicted as 0

blobs dataset
[[1 -9]] predicted as 0 [[-4. 7.8]] predicted as 1 [[-9. 4.5]] predicted as 0

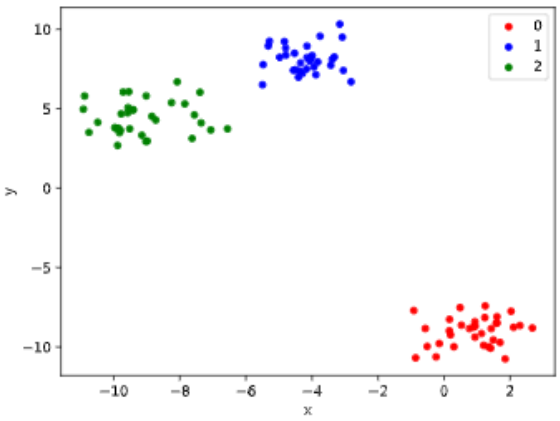
Moons dataset
[[-0.5 0.5]] predicted as 1 [[1. 0.5]] predicted as 0 [[0 0]] predicted as 1 [[1.5 -0.5]] predicted as 0

circles dataset
[[-0.6 -0.85]] predicted as 0 [[0.75 -0.06]] predicted as 1

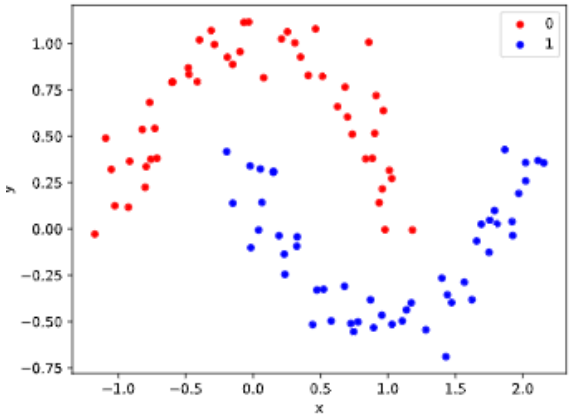
Figures



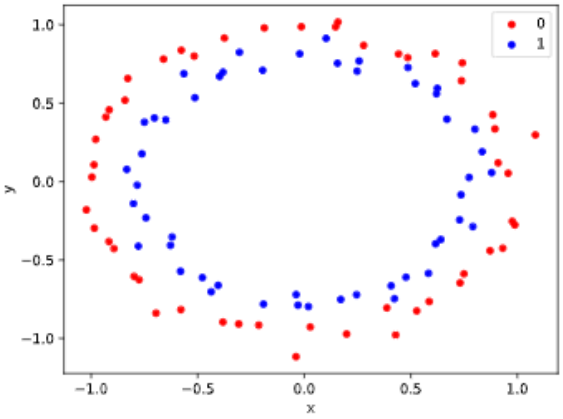
(a) XOR dataset



(b) Blobs dataset



(c) Moons dataset



(d) Circles dataset

Individual Contributions

Jesús Omar Cuenca Espino

My contribution to this homework was working on the missing parts of the code that my teammate Luis Felipe left for me.
Pretty much it was completing the voting mechanism of the KNN algorithm and completing the function used for predictions.

Luis Felipe Álvarez Sánchez

My contribution to this homework was the euclidean distance and the get neighbors function from the code. I also explained the code to my teammates. I really enjoyed working with this algorithm because I was able to apply the theoretical part from the class into functional code.

Juan José González Andrews

My contribution to this homework was working on the report and explaining the main functionalities of the implementation code in the description. I was not able to provide a lot in the code but was able to understand its implementation by studying how Felipe and Omar applied the code. Working on the description helped me better comprehend the concepts and its direct application in the homework.

Rodrigo Montemayor Faudoa

My contribution to this homework was working on the report, specifically on the description of the code. I still struggle with the implementation of the code but my teammate Felipe helped me understand the code as he was working on it and describing his process step by step and occasionally I contributed by suggesting ideas when he got stuck.

Bibliography / References

1. *Machine Learning Basics with the K-Nearest Neighbors Algorithm. (2021). Retrieved 19 October 2021, from https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761*
2. *Overview of kNN. (2021). Retrieved 19 October 2021, from https://www.ibm.com/docs/en/spss-statistics/24.0.0?topic=knn-overview-command*