

TC3020 Machine Learning

Assignment 5: k-Nearest Neighbours

Dr. Rafael Pérez Torres

October 12th, 2021

Abstract

In this assignment, you will implement a very basic version of the k-NN algorithm studied in class.

1 Description

Overall goal: Implement the k-NN classifier.

You must not employ specialized libraries (sklearn or similar) to produce your solution; it is ok to employ them to compare though.

You have access to a partially implemented source code for an k-NN. Take your time to study the code, understand its structure, and analyze the shapes of defined elements. There are three main elements:

- A launcher (`launcher.py`) that creates some instances of the k-NN and runs them.
- A kNN class (`knn.py`) with some empty methods. **You will be working in this file.**
- A utils script (`utils.py`) that defines a function to read the dataset.

2 Implementation

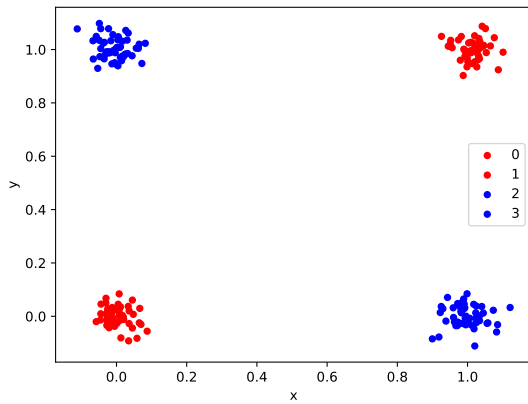
- Locate the TODO comments in the `knn.py` file and implement the required functionality.
- Pay attention to the dimensions of the parameters employed in the functions, so that you can be sure your implementation is returning arrays/values with the proper dimensions..
- **Debug as much as you can** to understand the code flow and identify issues.

As this assignment is simple (not even vectorization is mandatory) you are not given guideline results.

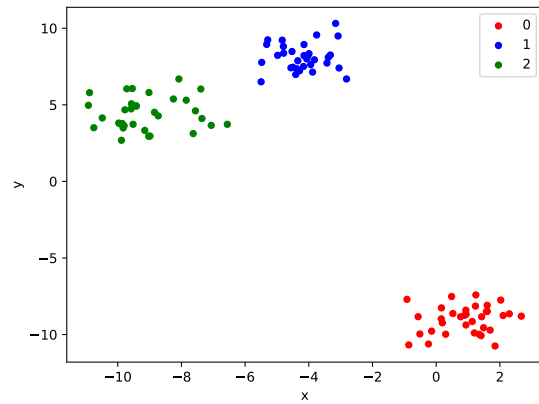
3 Datasets

You are provided with several datasets.

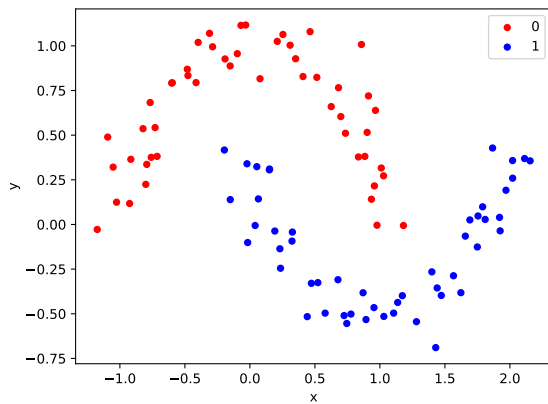
- XOR problem. Shown in Fig 1a.
- Blobs: Three separable clouds of data, shown in Fig. 1b.
- Moons: Two intertwined moon-shaped sets of data, shown in Fig. 1c.
- Circles: Two data classes nested in circles, shown in Fig. 1d.



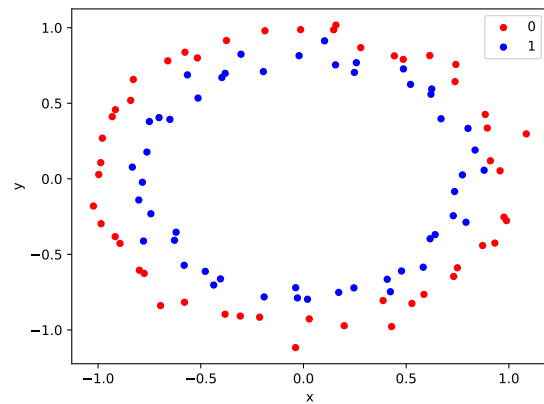
(a) XOR dataset



(b) Blobs dataset



(c) Moons dataset



(d) Circles dataset

Figure 1: Datasets

4 Further work

Prepare a short report (3 text pages top, could be more with cover, figures), where you present your findings during your implementation. After you implemented correctly your code, for this assignment, you need to play around with the k hyperparameter value that is set in the `launcher.py` file. Modify also the examples that are predicted. Which datasets are more likely to be affected by changing the k value? **The report MUST include a section with the individual contributions from each team member.**

5 Notes

- Your code should be ready to be run by just executing the `launcher.py` script.
- Include everything (even datasets) in your submitted code so that it can be executed straightforwardly.
- It will be penalized if your code can't be run directly.

6 Deliverables

- Your implementation as a zip file with the source code.

- Name your file as A1_A2_A3-HW-02 . zip where A1, A2, A3 are the reg number (matrícula) of students.
 - Include the names of team members and student numbers as comments.
- Your report as a PDF.

A single submission per team is enough, there is no need for submitting more than once.