MC886A/MO444A - Machine Learninng
Institute of Computing - Unicamp
First Semester 2021

Prof. Esther Colombini
esther@ic.unicamp.br
PEDs
Alana Correia (alana.correia@ic.unicamp.br
Patrick Ferreira (patrickctrf@gmail.com)

Project 2 - Deadline: 23/05/2021

1 Goal

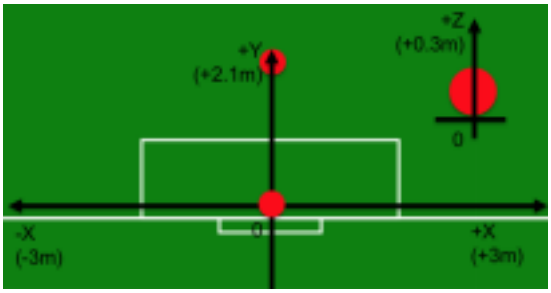
This assignment aims to apply learning methods to solve classification and regression tasks. In each task, you are requested to use the appropriate metrics and learning methods. For the first task, you are required to implement a Gradient Descent-based optimizer from scratch.

2 PART I - Regression Task

2.1 RoboCup soccer context

Consider that your are a goalkeeper in a robotic soccer match at RoboCup Teen Size League. Your opponent will kick a ball against you. This is your view:

Your goal is to predict the position of the ball to intercept it, that is, to defend your goal! As a reference, you, the goalkeeper, is at $x = 0$, $y = 0$. The approximate dimensions of your field are:



The work consists of implementing a linear and a polynomial regression to predict the ball trajectory. You are given two datasets: data1 and data2. They represent two diferent trajectories for the ball. You should learn the best model to fit the data. For this part of the project, you can use all data to learn your model. For this part, you are requested to:

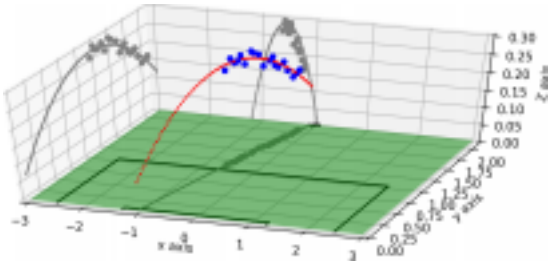


Figura 1: Sample ball learned trajectory

- Define linear and polynomial regressions to learn your models, derive the equations, and implement a gradient descent method from scratch
- Train a linear and polynomial regression in each dataset
- Evaluate the impact of the model complexity in the results (degree of the model)
- Evaluate the impact of the learning rate (test at least 3 values) in the process
- Plot the models learned to predict the trajectory (like Figure 2.1) along with the metrics used to evaluate them, discussing the results.

Consider that:

- Your data is captured with a rate of 60 frames/second, and you have waited 1/3 of a second to make a decision; • The data captured refers to the center of the ball;
- The ball is a size 5 FIFA ball, as depicted next.

Discuss if, given the learned models, you will be able to defend the kicks.

3 PART II - Classification Task

In the second part of the work, you should solve a classification task using supervised learning methods. For this task, you can use libraries that implement the methods you want to use. For this task, we will employ the Pen-Based Recognition of Handwritten Digits Data Set. It is a simple and well-known image bank for image recognition. It consists of 8-by-8-pixel gray-scale images divided into 10 classes of digits. A complete description can be found at the following link: archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits. A sample of the dataset is presented next.

- Load your dataset (loaddigits) from the sklearn.datasets module
- Split your data into training/test sets (9010)
- Apply pre-processing steps over your data that you think necessary (feature scaling, normalization)
- Plot your data using any visualization method that suits you (t-SNE, UMAP, etc.)
- Train with cross-validation two different SL methods in the classification task for the 10 digit-classification • Evaluate different hyperparameters of your model using appropriate metrics (accuracy, recall, f1, etc.) • Choose your best models (the best configuration in the validation set for each method) and test them with the test set. Plot the confusion matrix and corresponding metrics that you will use to assess the model quality

Discuss your results.

4 Programming language

You should use Pyhton as programming language.

5 Evaluation and Discussion

The system should be evaluated according to the quality of the solutions found and a critical evaluation is expected on the relationship between adopted parameters x solution quality. Graphs, tables and images representing the results are expected. Further comparisons with the literature are welcome, although not mandatory.

Please, discuss in the report:

- How/if normalization affected your results
- The quality of the results in each task
- The advantages and disadvantages of each method s

6 Groups

The groups must be composed of 2 members.

7 Report

The definition of the problem, the solution, and the results obtained must be presented in a report created as a Jupyter notebook. Please, make sure you put the graphs, tables, comparisons, and critical analysis in the notebook. The report should clearly indicate what the contribution of each team member was.

8 Grading

This work will be evaluated according to the following criteria:

- Submission within deadline
- Quality of the solution employed
- Report and discussions
- Code analysis
- Individual student participation in the project

8.1 Penalty policy for late submission

You are not encouraged to submit your assignment after due date. However, in case you did, your grade will be penalized as follows:

- late submission one day after the deadline: grade * 0.75
- late submission two days after the deadline: grade * 0.5
- late submission three days after the deadline: grade * 0.25