Homework-LogisticRegression

Stefania Sarno

July 2022

1 Introduction

In this homework assignment you will use the database MNIST and logistic regression. Your job will be to solve a simplified classification problem and interpret the results.
To import the data set you will need the following command line:

```
from sklearn.datasets import fetch_openml
dataset = fetch_openml("mnist_784")
```

If you have an old version of sklearn you wil probably need to use this command instead:

```
from sklearn.datasets import fetch_mldata
dataset = fetch_mldata("MNIST original")
```

2 Classification on hand-written digits (MNIST) using Logistic Regression

Generate a notebook that does the following things and try to answer exhaustively to questions below:

- Visualize a sample of of the data corresponding to different digits (try to reproduce a figure as Figure 1).
- Redefine `y_train_5_only` as 1 if the digit is a 5 and 0 otherwise (you will want to split the data into train and test before, as we see in our lessons)
- Train a logistic regression classifier to distinguish the 5 from the rest of the digits.
- Compute the accuracy on the test data. Do you think you get a good score?
- On the same test data, which would be the accuracy of a classifier that would say always "different from 5"?
- Why do this problem arise?

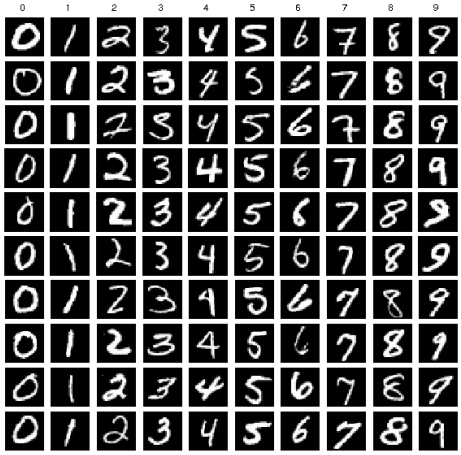

Figure 1: MNIST example data

- Try to find an appropriate metric to evaluate your classifier in this case. You can choose among one of the metrics we see at lesson or look in the web for good alternative.

3 Practical remarks

The homework is quite simple but also quite instructive. You are not very guided because you are supposed to remember and reuse what we see in the class. Feel free to do more data visualization and analysis than what summarized in the list of the previous section if you think that this can be useful. Your main job is to show that you can analyze the problem in a critical way.
You are supposed to work on the assignment alone or in couple. If you work in a group of 2, please declare your partner before submission.