SAPIENZA
UNIVERSITÀ DI ROMA

Homework 1

Machine Learning

Students:
Flavio Maiorana
2051396

Academic Year 2023/2024

When approaching a classification problem, the steps to be followed usually start from looking at the dataset and eventually doing some actions on it even before "cranking up" the learning algorithm.

1 Data visualization and preprocessing

First of all, it could be useful to gain some insight one how the dataaset is made.

```
[Dataset 1]
    N Examples: 50000
    N Inputs: 100
    N Classes: 10
    Classes: [0 1 2 3 4 5 6 7 8 9]
     - Class 0: 5000 (10.0)
     - Class 1: 5000 (10.0)
     - Class 2: 5000 (10.0)
     - Class 3: 5000 (10.0)
     - Class 4: 5000 (10.0)
     - Class 5: 5000 (10.0)
     - Class 6: 5000 (10.0)
     - Class 7: 5000 (10.0)
     - Class 8: 5000 (10.0)
     - Class 9: 5000 (10.0)

[Dataset 2]
    N Examples: 50000
    N Inputs: 1000
    N Classes: 10
    Classes: [0 1 2 3 4 5 6 7 8 9]
     - Class 0: 5000 (10.0)
     - Class 1: 5000 (10.0)
     - Class 2: 5000 (10.0)
     - Class 3: 5000 (10.0)
     - Class 4: 5000 (10.0)
     - Class 5: 5000 (10.0)
     - Class 6: 5000 (10.0)
     - Class 7: 5000 (10.0)
     - Class 8: 5000 (10.0)
     - Class 9: 5000 (10.0)
```

Some comments:

- Both datasets have examples pertaining 10 different Classes
- The number of examples is pretty high
- Both datasets are balanced

After that, one could want to visualize the data in order to be able to spot some peculiarities of the data by just literally looking at it. Although, the first problem we confront with with this specific dataset is the number of features.