Manual for Homework 3 – Machine Learning

Authors: Manuel di Pietro, Lorenzo Pichilli

Steps:

1. To execute the script, simply run this command:
   ```
   python3 main.py
   ```

2. After this, the program will prompt to you this query:
   ```
   What is the column (0,1,2,...) of the label?
   ```
   So you must answer with the number of the column that contains the labels of your data set.

3. Then it will prompt to you another query:
   ```
   Path of the dataSet file:
   ```
   In this case you must write the path of you data set file.

4. The last query is:
   ```
   # of training set rows (MAX is 20000):
   ```
   In this case you must write the number of the rows that training set will use from the data set (the remaining ones are used for the test set).

5. After that the script will create various files containing all the sets encoded and also the original version, then it will start to train the MLP.

6. At the end, the script will prompt to you the accuracy obtained from the prediction and the number of rows decoded that are equal to the original file.

Furthermore, the script will save also the mapping of the label and features that was used to encode the data set in the `cvLabel.pkl` and `cvFeatures.pkl` file respectively and the Neural Net already trained in the `mlp.pkl` file to a possible reuse of it.

If you want use the same script in the same folder with another data set, then you must delete the `dataSetEncoded.csv` file and the `mlp.pkl` file.

Here an example of the script execution using the **letter-recognition.data** data set:

```
$ python3 main.py
What is the column (0,1,2,...) of the label? 0
Path of the dataSet file: letter-recognition.data
# of training set rows (MAX is 20000): 18000
dataSetEncoded ready.
trainingSetEncoded ready.
testSetEncoded ready.
Training...
Iteration 1, loss = 7.09013583
Iteration 2, loss = 2.27238026
Iteration 3, loss = 1.70552360
Iteration 4, loss = 1.36432007
Iteration 5, loss = 1.11410135
Iteration 6, loss = 0.92477065
Iteration 7, loss = 0.77407101
Iteration 8, loss = 0.65827060
Iteration 9, loss = 0.56373545
Iteration 10, loss = 0.48779990
```

```
Iteration 11, loss = 0.42566857
Iteration 12, loss = 0.38246014
Iteration 13, loss = 0.34112004
Iteration 14, loss = 0.30365107
Iteration 15, loss = 0.27925815
Iteration 16, loss = 0.25800109
Iteration 17, loss = 0.23687010
Iteration 18, loss = 0.22865789
Iteration 19, loss = 0.21976704
Iteration 20, loss = 0.20151360
Iteration 21, loss = 0.18902333
Iteration 22, loss = 0.18380827
Iteration 23, loss = 0.18289679
Iteration 24, loss = 0.17267556
Iteration 25, loss = 0.16847859
Iteration 26, loss = 0.16141209
Iteration 27, loss = 0.16013984
Iteration 28, loss = 0.15501855
Iteration 29, loss = 0.14932605
Iteration 30, loss = 0.14969354
Iteration 31, loss = 0.14911787
Iteration 32, loss = 0.14388611
Iteration 33, loss = 0.13834015
Done.
Accuracy for Encoded Test: 28.35%
# of right answers: 719 on 2000, 35.949999999999996%
```