Introduction to Machine Learning - Exercise 2

Due Date: November 30th 22:00, 2020

Yosi Shrem, Yael Segal and Joseph Keshet

November 16, 2020

Guidelines

1. You are not allowed to use external packages other than numpy and scipy (KNN should be fully implemented , using an external implementation is forbidden).
2. In order to submit your solution please upload your files to `Submit` and check your inbox for the feedback mail.
3. Technical questions about this exercise should be asked at the course' piazza.
4. Private/Personal issues regarding the deadline should be directed to **Yael Segal**.

1 KNN, Perceptron, and Passive Aggressive (PA)

In this exercise you will implement and compare K-nearest neighbors (KNN), Perceptron, and PA. You are provided with a training set of 355 examples where you need to classify between 3 classes.

To make things easier for you, here are the update rules for all models:

Perceptron

$$w_{t+1}^{y} = w_t^{y} + \eta * x$$

$$w_{t+1}^{\hat{y}} = w_t^{\hat{y}} - \eta * x$$

$$w_{t+1}^{i \neq \hat{y}, y} = w_{t+1}^{i \neq \hat{y}, y}$$

PA

$$w_{t+1}^{y} = w_t^{y} + \tau * x$$

$$w_{t+1}^{\hat{y}} = w_t^{\hat{y}} - \tau * x$$

$$w_{t+1}^{i \neq \hat{y}, y} = w_{t+1}^{i \neq \hat{y}, y}$$

where τ is set to:

$$\tau = \frac{\ell(w, x, y)}{2 \cdot ||x||^2}$$

2 Dataset

Your data set is about Wine Quality classification. In this dataset, you are provided with 12 features per instance (11 of which are numerical and one is the categorial) and three labels. Labels correspond to the wine quality classification. You should explore different ways to convert this categorial attribute to a numerical one. Moreover, try different normalization techniques as well as feature selection.

A summarize of the feature description can be found in the table below:

Name	Data Type
fixed acidity	continuous
volatile acidity	continuous
citric acid	continuous
residual sugar	continuous
chlorides	continuous
free sulfur dioxidet	continuous
total sulfur dioxide	continuous
density	continuous
pH	continuous
sulphates	continuous
alcohol	continuous
wine type	nominal

3 Code

Your main file should be called: `ex_2.py`
Your code should get as input three arguments. The first one will be the training examples (`train_x.txt`), the second one is the training labels (`train_y.txt`), and the third one will be the testing examples (`test_x.txt`), where `train_x.txt` and `test_x.txt` will have the same format. You should train 3 algorithms learned in class: KNN, Perceptron and PA (in that order). Next, you should output to the screen your predictions (only) for `test_x.txt` in the following format:

```
knn:0, perceptron: 0, pa: 1
knn:2, perceptron: 2, pa: 2
knn:1, perceptron: 1, pa: 1
...
you can use the following print line:
print(f"knn: {knn_yhat}, perceptron: {perceptron_yhat}, pa: {pa_yhat}")
```

where each line in your output corresponds to a line in `test_x.txt` and the numbers represent your class predictions for classes 0,1,2. Notice, all your hyper-parameters should be hardcode, no external arguments will be provided during runtime.

4 What to submit?

You should submit the following files:

- A `txt` file, named `details.txt` with your name and ID.

  ```
  name: FIRST_NAME LAST_NAME
  ID: 123456789
  ```

- Python 3.6 code. Your main function should reside in a file called `ex_2.py`. The main function should train all three models and output the predictions as described in 3.
- A `PDF` report named `report.pdf` with all the implementation details and hyper-parameters, (How did you choose the learning rate? What values did you try and how was the performance? How did you choose the k value for KNN? etc.).
- Part of your grade will consist of automatic checks using the `Submit` system. Make sure your output matches the expected output described above.
- **Note**: your code should run up to 5 minutes long. Submissions that will run longer than that, will be graded as 0.

Overall : `ex_2.py, details.txt` and `report.pdf`