Introduction to Machine Learning - Exercise 2

Due Date: December 9th 22:00, 2019

Yosi Shrem and Joseph Keshet

November 26, 2019

1 Perceptron, SVM, and Passive Aggressive (PA)

In this exercise you will implement and compare Perceptron, SVM, and PA. You are provided with a training set of 3,341 examples where you need to classify between 3 classes.

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

Perceptron

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

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

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

SVM

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

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

$$w_{t+1}^{i \neq \hat{y}, y} = (1 - \eta\lambda) w_{t+1}^{i \neq \hat{y}, y}$$

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 Abalone age classification. In this dataset, you are provided with eight features per instance (seven of which are numerical and one is the categorial) and three labels. Labels correspond to the Abalones age. 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	Meas.	Description
Sex	nominal	-	M, F, and I (infant)
Length	continuous	mm	longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried

3 Code

Your main file should be called: `ex2.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: Perceptron, SVM and PA (in that order). Next, you should output to the screen your predictions (only) for `test_x.txt` in the following format:

```
perceptron: 0, svm: 0, pa: 1
perceptron: 2, svm: 2, pa: 2
perceptron: 1, svm: 1, pa: 1
...
you can use the following print line:
print(f"perceptron: {perceptron_yhat}, svm: {svm_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 `ex2.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 lambda value for SVM? 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 : `ex2.py, details.txt` and `report.pdf`