CS 165B – Machine Learning, Spring 2023

Machine Problem #1
Due Thursday, April 27 by 11:00pm

TA: Danqing Wang
Email: danqingwang@ucsb.edu

Change the **Python 3** program ***mp1.py*** that implements a three-class linear classifier.

Code Instructions

Specially, change the function ***run_train_test(training_input, testing_input)*** defined in ***mp1.py***. The function should include the following process:

- Training (using the training data set):
  - Compute the centroid of each class (e.g. A, B, and C).
  - Construct a discriminant function between each pair of classes (e.g. A/B, B/C, and A/C), halfway between the two centroids and orthogonal to the line connecting the two centroids. This is the "basic linear classifier" that we have discussed.
- Testing (using the testing data set):
  - For each instance, use the discriminant function to decide "A or B" and then (depending on that answer) to decide "A or C" or "B or C." (Ties should give priority to class A, then B, then C)
    - First, use A/B discriminant function:
      - If classified as A, then decide A or C
      - If classified as B, then decide B or C
  - Keep track of true positives, true negatives, false positives, and false negatives.

The training and testing data sets are available in this starter package, along with a description of their formats (***MP1_data.md***)

This function is where you implement the assignment. Feel free to define additional functions, but **DO NOT** change this function signature and **DO NOT** change the mp1.py module name. We will call this as the entry point to your code.

Output Instructions

As output, the program should return a dictionary of averages over all three classes of the true positive rate, the false positive rate, the error rate, the accuracy, and the precision:

```
>> print(run_train_test(training_input, testing_input))
```

```
 {
   “tpr”: 0.80 # true positive rate
   “fpr”: 0.27 # false positive rate
   “error_rate”: 0.44
   “accuracy”: 0.60
   “precision”: 0.90
}
```

(Note: These numbers are made up, for purposes of illustration only.) The ***run_train_test*** function should **return results using the same keys as shown.**

Information on computing these averages ***is included in the MP1_extra.pdf.***

Evaluation Instructions

You can test your program with the training1.txt and testing1.txt provided in this starter package. You can use the provided ***evaluation.py*** for checking:

```
$ python evaluate.py
```

Submission Instructions

You should upload ***mp1.py*** to Gradescope for grading.