

# CS 165B – Machine Learning

## Spring 2020, Xifeng Yan

### Assignment #1

Due Wednesday, Apr 22 by 5:00pm PST

---

## Programming Assignment

Write a **Python 3** program called **pa1.py** that implements a three-class linear classifier using the following method:

Training (using the training data set):

1. Compute the centroid of each class (e.g. A, B, and C).
2. 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):

1. 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)
2. 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 ([PA1-5data.html](#)), hints in [PA1-5extra.pdf](#). You are also given a starter code file called **pa1.py**. In the **pa1.py** file you will find the function

```
% def run_train_test(training_input, testing_input)
```

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 **pa1.py** module name. We will call this as the entry point to your code.

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 `pal` starter package.*

You can test your program with the training and testing sets provided on the Piazza.

### Evaluation Instructions

```
% python evaluate.py
```

### Submission Instructions

You should use the following command to submit your homework in CSIL:

```
% turnin pal@hongwang600 pal.py
```