

# Assignment #5

## Data Mining

Due: October 15, 2018

1. (R and Python) Modify your program in Assignment #4 to do followings. For logistic regression, you can use any optimization function in R or Python. However, you should not use the logistic regression function directly.
  - a. Prompt the user whether to run regression or classification.
  - b. If regression is chosen, perform the linear regression as you did in Assignment #3. (You have nothing to work on the regression algorithm in this assignment).
  - c. If classification is chosen, ask the user the filename of the training and test dataset. (Assume the column location of the class variable is the same for both training and test dataset.)
  - d. If classification is chosen, prompt the user to choose (i) LDA and (ii) QDA, (iii) RDA, or (iv) Logistic regression
  - e. Perform (i) LDA and (ii) QDA, (iii) RDA, or (iv) Logistic Regression depending on the choice by the user. However, if the data has more than two classes, do not prompt (iv) Logistic Regression. Use a file named "pid.dat" for the training and 'pidtest.dat' as the test data in this assignment.
  - f. The output file for classification generated by the program must look like below. (The numbers are fictitious).

```
ID, Actual class, Resub pred, Pred Prob (if logistic regression is chosen)
```

```
-----
```

```
1, 1, 1, 0.878
```

```
2, 2, 2, 0.213
```

```
3, 1, 1, 0.675
```

```
(continue)
```

```
Confusion Matrix (Resubstitution)
```

```
-----
```

		Predicted Class	
		1	2
Actual	1	239	14
Class	2	12	153

```
Model Summary (Resubstitution)
```

```
-----
```

```
Overall accuracy = .793
```

```
Sensitivity = .894 #print this line only if there are two classes#
```

```
Specificity = .743 #print this line only if there are two classes#
```

```
ID, Actual class, Test pred, Pred Prob (if logistic regression is chosen)
```

```
-----
```

```
1, 1, 1, 0.878
```

```
2, 2, 2, 0.213
```

```
3, 1, 1, 0.675
```

```
(continue)
```

```
Confusion Matrix (Test)
```

```
-----
```

		Predicted Class	
		1	2
Actual	1	239	14
Class	2	12	153

```
Model Summary (Test)
```

```
-----
```

```
Overall accuracy = .793
```

```
Sensitivity = .894 #print this line only if there are two classes#
```

```
Specificity = .743 #print this line only if there are two classes#
```