

# Assignment #4

Data Mining

Due: October 8, 2018

1. (R and Python) Modify your program in Assignment #3 to do followings.
  - a. Prompt the user whether to run regression or classification.
  - b. If regression is chosen, perform the linear regression as you did in Assignment #2. (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, or (iii) RDA.
  - e. Perform (i) LDA and (ii) QDA, or (iii) RDA depending on the choice by the user. Use a file named "veh.dat" for the training and 'veh.test.dat' as the test data in this assignment.
  - f. For RDA, consider the extended model with two parameters. Try values in [0, 1] with increment of 0.05 for both the parameter values. Choose the parameters that minimize the test error rate. The program must be able to choose the optimal parameters by itself after comparing test error rates. Also produce a (3-dimensional) plot that displays both parameter values and their corresponding test accuracy. I.e. X1 is for parameter 1, X2 is for parameter 2, Y is for test accuracy. (Note: You need to estimate  $\sigma^2$ . Use the average of diagonal elements of the pooled covariance matrix  $S_p^2$ .)
  - g. The output file for classification generated by the program must look like below. (The numbers are fictitious).

```
ID, Actual class, Resub pred
```

```
-----
```

```
1, 1, 1
2, 2, 2
3, 1, 1
(continue)
```

```
Confusion Matrix (Resubstitution)
```

```
-----
```

		Predicted Class			
		1	2	3	4
Actual Class	1	239	14	6	8
	2	12	153	5	12
	3	2	4	98	2
	4	3	6	8	123

```
Model Summary (Resubstitution)
```

```
-----
```

```
Overall accuracy = .793
```

```
ID, Actual class, Test pred
```

```
-----
```

```
1, 1, 1
2, 2, 2
3, 1, 1
(continue)
```

```
Confusion Matrix (Test)
```

```
-----
```

		Predicted Class			
		1	2	3	4
Actual Class	1	239	14	6	8
	2	12	153	5	12
	3	2	4	98	2
	4	3	6	8	123

```
Model Summary (Test)
```

```
-----
```

```
Overall accuracy = .793
```