Assignment #3

Data Mining Due: October 1, 2018

1. (R and Python) For classification, assume that there may be more than two classes. You can assume that values of the class variable are integers starting with 1. Assume that a training dataset and a test dataset are available. Modify your program in Assignment #2 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.)
    a. Make the program to implement (i) LDA and (ii) QDA that can handle more than two classes.
    d. Perform (i) LDA or (ii) QDA depending on the choice by the user. Use a data file named 'veh.dat' for the training and 'vehtest.dat' as the test data.
    e. 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      1        239      14       6       8
Class       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      1        239      14       6       8
Class       2         12     153       5      12
            3          2       4      98       2
            4          3       6       8     123

Model Summary (Test)
----------------------------
Overall accuracy = .793
```