

Assignment #10

Data Mining

Due: November 26, 2018

Modify your program for Assignment #8 to do followings. You may use one of two (R or Python) language for this assignment.

1. Prompt the user whether to run regression or classification.
2. If classification is chosen, prompt the user to choose (i) LDA and (ii) QDA, (iii) RDA, (iv) Logistic regression, (v) Naïve Bayes, (vi) 1-level decision tree, or (vii) Bagging Ensemble. However, if the data has more than two classes, do not prompt (iv), (v) and (vi).
3. For Bagging Ensemble method, use LDA as the classifier and 51 bootstraps as the number of re-sampled data.
4. Use a file named "veh.dat" for the training and 'veh.test.dat' as the test data in this assignment. The last column is the class variable.
5. Perform (i)-(vii) methods depending on the choice by the user. Only the output of the **test data** is necessary for this assignment.

The output file for classification generated by the program must look like

```
(1)  LDA - no bagging
ID, Actual class, LDA-nobagging pred
-----
1, 1, 1
2, 2, 2
3, 1, 1
(continue)
```

```
Confusion Matrix (LDA - no bagging)
-----
                Predicted Class
                1      2
Actual    1      239   14
Class     2      12   153
```

```
Model Summary (LDA - no bagging)
-----
Overall accuracy = .793
```

```
(2)  LDA - bagging
ID, Actual class, LDA-bagging pred
-----
1, 1, 1
2, 2, 2
3, 1, 1
(continue)
```

```
Confusion Matrix (LDA - bagging)
-----
                Predicted Class
                1      2
Actual    1      239   14
Class     2      12   153
```

```
Model Summary (LDA - bagging)
-----
Overall accuracy = .793
```