

Assignment #7

Data Mining

Due: November 5, 2018

(R and Python) Modify your program for Assignment #6 to do followings.

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, or (vi) 1-level decision tree. However, if the data has more than two classes, do not prompt (iv), (v) and (vi).
3. Make your program to implement (vi) 1-level decision tree only for two classes:
 - a. Find CART splitting rule assuming all variables are continuous, then split the current node into two subnodes. (Categorical variables are not considered in this assignment)
 - b. Print out the 1-level tree information and number of observations from each class (see below).
4. Perform (i)-(vi) methods depending on the choice by the user.

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

Tree Structure

```
Node 1: x3 <= 1.740 (21, 25)
Node 2: 1 (18, 2)
Node 3: 2 (3, 23)
```

ID, Actual class, Resub pred

```
-----
1, 1, 1
2, 2, 2
3, 1, 1
(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

```
-----
1, 1, 1
2, 2, 2
3, 1, 1
(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#