

Assignment #8

Data Mining

Due: November 12, 2018

(R or Python) Modify your program for Assignment #7 to do followings. For this assignment, use 'titanic.csv' file for categorical variables.

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, then split the current node into two subnodes. (Categorical variables should be 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: Sex in {Male} (21, 25)
  Node 2: Yes (18, 2)
    Node 3: No (3, 23)
```

ID, Actual class, Resub pred

```
-----
1, Yes, Yes
2, No, No
3, Yes, No
(continue)
```

Confusion Matrix (Resubstitution)

```
-----
                        Predicted Class
                        No      Yes
Actual   No      239      14
Class    Yes      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, Yes, Yes
2, No, No
3, Yes, No
(continue)
```

Confusion Matrix (Test)

```
-----
Predicted Class
                        No      Yes
Actual   No      239      14
Class    Yes      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#