

BUAN 6357 (Johnston)

Homework 4

Code Due: 2 March 2019 (6PM)

Part B Due: 3 March 2019 (11:59PM)

Points available: 90.

This assignment is about evaluating both the classification accuracy and the risk of miscalculation across 3 modeling strategies and multiple training sample sizes (number of replications per digit). Use the “noisy_segments_multiclassif_v4.txt” file from UTDbox>demo as a starting point. Part A uses 3 sample sizes: 25, 50, and 100. Part B may extend the number of sample sizes.

For this assignment you will need the package “partykit” and may use the packages “data.table”, “tidyverse” and “broom”, in that order. You should not use any additional packages. You should use only the “require()” or “library()” statement in your code. Any use of the install.packages() function in submitted code will result in a score of 0 for that submission.

The first commands of your code submitted for grading to eLearning MUST be:

```
setwd(“c:/data/BUAN6357/HW_4”); source(“prep.txt”, echo=T)
```

and the last command of your code MUST be:

```
source(“validate.txt”, echo=T)
```

Be careful with the quote characters as they must ALL be the same at the beginning and end of a string. (Use the single or double quote character from the key next to “Enter”.) Inclusion of these lines is required BEFORE your code will be tested.

Submit the code to eLearning as an ASCII file which can be copied directly into R.

If you use the Shiny app to evaluate your code before grading, leave out the `setwd()` line and the `source()` line presented above. This will give you much quicker feedback on your code than submission through eLearning. Code submitted through the Shiny app should also be in an ASCII file.

You may submit this assignment as many times as needed until you get full credit.

Deliverables (individual vectors; all names start lower case; BR = Bayes Risk; PC = percent correct):

1. `holdPClogit` PC classification for the logit model approach
2. `holdMedBRlogit` median BR value for the logit model approach
3. `hold75BRlogit` 3rd quartile BR value for the logit model approach
4. `holdPCtree10` PC classification for the multiple trees approach
5. `holdMedBRtree10` median BR value for the multiple trees approach
6. `hold75BRtree10` 3rd quartile BR value for the multiple trees approach
7. `holdPCtree1` PC classification for the single tree approach
8. `holdMedBRtree1` median BR value for the single tree approach
9. `hold75BRtree1` 3rd quartile BR value for the single tree approach

Note: these deliverables are intended to demonstrate your ability to successfully modify the original code and extract collections of summary values.

Part B of HW 4 will direct you to explore both the intermediate results and the deliverables from Part A and answer questions about each of them. You may submit answers to HW 4 part B as many times as you wish but only the score for the last submitted code will be retained.