Ethan Chang

HW 03

CSCI 420

E. In your PDF, show the resulting one-rule threshold classifier code.
What was your one-rule for this assignment?
Which attribute did it use?
It is just an if-else-statement. Don’t forget the Else!!
Copy and paste your code from your output classifier file.
Please use black ink on a white background.
Do not use a screen capture, the fonts do not scale.

None

```
# One-Rule Threshold Classifier

BEST_FEATURE <- "LANE_CHANGES"
BEST_THRESHOLD <- 2L
BEST_FLAG_LEFT <- FALSE

args <- commandArgs(trailingOnly = TRUE)

# Makes sure program has 1 argument.
if (length(args) < 1) stop("Usage: Rscript HW_03_Chang_Ethan_Classifier.r
<test_suite.csv>")

# args[1] = TEST_SUITE_33
fname <- args[1]
fname <- file.path("TEST_SUITE_33", fname)

df <- read.csv(fname, stringsAsFactors = FALSE)

all_speeds <- trunc(df$SPEED)
all_lane_change <- df$LANE_CHANGES
all_brightness <- df$BRIGHTNESS

if (BEST_FEATURE == "SPEED") {
    x <- all_speeds
} else if (BEST_FEATURE == "LANE_CHANGES") {
    x <- all_lane_change
} else {
    x <- all_brightness
}

if (BEST_FLAG_LEFT) {
   pred_aggr <- (x <= BEST_THRESHOLD)
} else {
   pred_aggr <- (x > BEST_THRESHOLD)
}

# Outputs the results
cat('Aggresive: ', sum(pred_aggr), '\n')
cat('Non-Aggresive: ', sum(!pred_aggr), '\n')
```

The one-rule classifier uses an if–else statement based on the left-group flag. The chosen attribute is LANE_CHANGES because it has the minimum badness. The rule is: predict AGGRESSIVE when LANE_CHANGES > 2, and predict NON-AGGRESSIVE otherwise.

F. Run your code to produce your classifier program.
Your code must produce the resulting classifier. (2)

None

```
# Handles the data and then creates the Classifier Rscript file to output results.
Rscript HW_03_Chang_Ethan.r
```

G. Run your resulting classifier on the supplied test suite.
Your resulting classifier must run. (2)

None

```

# Outputs the results by trying out the tested data.

Rscript HW_03_Chang_Ethan_Classifier.r Data_33.csv
```

H. Report how many Aggressive drivers did your classifier routine find in the test suite?
Report these numbers in your write-up. (2)

My classifier routine found 132 aggressive drivers and 168 non-aggressive drivers. ‘
There are a total of 300 drivers in the data set.

I. Conclusion: Write up what you learned here using at least three paragraphs.
(2) What did you discover?
Were the results what you expected? What was surprising?
Was there anything particularly challenging? Did anything go wrong?
Provide strong evidence of learning.
Write a conclusion that describes what you learned in this homework.
Points are taken off for writing with bullet points or checkmarks.

At the beginning of this assignment, I found it more challenging than the previous assignments for this course. I understood how to handle the data, but I initially assumed we were supposed to focus only on the speed attribute to build the classifier. It took me a while to realize we needed to loop over every attribute, including the number of lane changes and brightness. From this process, I learned the importance of interpreting instructions carefully. It's good practice to read the questions closely so you don't waste time. Because I missed that requirement at first, I didn't get the results I expected. The moment I realized we needed to test every attribute, I understood that the assignment demonstrates how a single attribute can have a significant impact on identifying aggressive driving behavior.

The crucial part of this assignment was understanding the math behind the predictive rule. I learned about the "badness" metric and how to compute it. Badness is the sum of two error rates: FAR (false alarm rate) and FNR (false negative rate). We then scan thresholds and pick the minimum badness, which means the lowest combined error, to build the classifier. I computed this minimum for each attribute and chose the overall lowest one to use for predictions.

Another thing I've learned about this project is using logical reasoning to understand why the classifier determines that the best attribute for identifying driver aggressiveness is lane changes. From my classifier script, I see that there are more aggressive drivers with lane changes greater than 2. With more lane changes, it likely means the driver is rushing to reach a destination and is trying to get through other drivers on the road. Brightness doesn't really indicate aggressiveness. Additionally, I've understood the importance of splitting the two groups by using "x > threshold" and "x <= threshold" to determine aggressiveness. It always depends on which group dominates.