Project STAT 652

Kotomi Oda

```
library(pacman)
  p_load(tidyverse, caTools, C50, doParallel, randomForest, pROC)
  lending_club_data_2012_2014_small <-</pre>
    readRDS("~/Documents/652/652_Project/data_small/lending_club_data_2012_2014_small.Rds")
  data <-lending_club_data_2012_2014_small
  #head(data)
  data <- data %>% select(funded_amnt,emp_length,annual_inc,home_ownership,
                           grade,last_pymnt_amnt, mort_acc, pub_rec, int_rate,
                           open_acc, num_actv_rev_tl, mo_sin_rcnt_rev_tl_op,
                           mo_sin_old_rev_tl_op, bc_util, bc_open_to_buy,
                           avg_cur_bal, acc_open_past_24mths, loan_status)
  data <- data %>%
    filter(loan_status %in% c("Fully Paid", "Charged Off")) %>%
    mutate(loan_status = ifelse(loan_status == "Fully Paid",0,1))
  data <- data %>% mutate(grade = factor(grade, levels = c("G", "F", "E", "D", "C", "B", "A")),
                           home_ownership = factor(home_ownership,
                                                   levels = c("ANY","NONE","OTHER","OWN","REN
                           emp_length = as.numeric(gsub("([0-9]+).*$", "\\1",
                                                        emp_length)))
Warning: There was 1 warning in `mutate()`.
i In argument: `emp_length = as.numeric(gsub("([0-9]+).*$", "\\1",
  emp_length))`.
Caused by warning:
! NAs introduced by coercion
```

```
data$emp_length <- data$emp_length %>% replace(is.na(.), 0)
data$loan_status <- factor(data$loan_status)

set.seed(123)
sample <- sample.split(data$loan_status, SplitRatio = 0.75)
train_data <- subset(data, sample == T)
test_data <- subset(data, sample == F)</pre>
```

C50

```
doParallel::registerDoParallel(cores = 8)
    tree_mod <- C5.0(train_data[,-18], train_data$loan_status)

predictions <- predict(tree_mod, test_data)
    table(test_data$loan_status, predictions)

predictions
    0    1
    0 1879 147
    1 223 180

mean(predictions == test$loan_status)

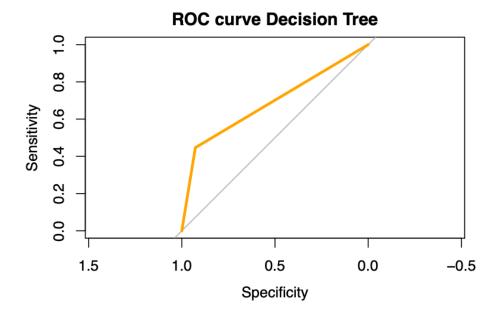
Error in mean(predictions == test$loan_status): object 'test' not found

roc1 <- roc(as.numeric(test_data$loan_status),as.numeric(predictions))

Setting levels: control = 1, case = 2

Setting direction: controls < cases

plot(roc1, col="Orange", lwd=3, main="ROC curve Decision Tree")</pre>
```



Random Forest