# Team Member: Tong Wu, Ying Li, Shengqin Zhao, Weiye Xu

# You are asked to suggest the best model to predict credit

# balance of credit holders.

# R comes with many built datasets, including "Credit" data set

# and you decide to start with analysis of that data set first.

# - Load "Credit" data set by loading "library(ISLR)"

library(ISLR)

# - Remove categorical columns (Gender, Student, Married,

# Ethnicity) and also ID by running following command

Credit2 = Credit[-c(8:11)]

Credit2 = Credit2[-1]

Income <- Credit2$Income

Limit <- Credit2$Limit

Rating <- Credit2$Rating

Cards <- Credit2$Cards

Education <- Credit2$Education

Balance <- Credit2$Balance;

# 1. Predict credit balance as a function of the rest of the

# parameters using "Credit2" data set.

# Use lm() function lm(data$responsevariable ~ data$explanatory)

# All the parameters can be used by using "." sign. In the case, data has to be identified

# m = lm(score~., data = student)

m = lm(Balance~.,data = Credit2)

summary(m)

# Answer 1: Fitted model: Balance = -447.95 + Income\*-7.56 + Limit\*0.12 + Rating\*2.06 + Cards\*11.59 + Age\*-0.89 + Education\*2.00

# 2. Suggest the best model for predicting credit balance.

# The best model gives high accuracy by adding only import

# parameters.

# Linear Regression - Two parameters

# Parametes of linear regression model will be added by "+“ sign

# m = lm(score ~ hours + student$id)

m1 <- lm(Balance~Income+Limit+Rating+Cards+Age,data = Credit2)

summary(m1)

m2 <- lm(Balance~Income+Limit+Rating+Age,data = Credit2)

summary(m2)

m3 <- lm(Balance~Income+Limit+Rating,data = Credit2)

summary(m3)

m4 <- lm(Balance~Income+Rating,data = Credit2)

summary(m4)

Call:

lm(formula = Balance ~ Income + Rating, data = Credit2)

Residuals:

Min 1Q Median 3Q Max

-278.57 -112.69 -36.21 57.92 575.24

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -534.81215 21.60270 -24.76 <2e-16 \*\*\*

Income -7.67212 0.37846 -20.27 <2e-16 \*\*\*

Rating 3.94926 0.08621 45.81 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 162.9 on 397 degrees of freedom

Multiple R-squared: 0.8751, Adjusted R-squared: 0.8745

F-statistic: 1391 on 2 and 397 DF, p-value: < 2.2e-16

# Answer 2: Fitted model which is model m4 : Balance = -534.81 + Income\*-7.67 + Rating\*3.95