Instituto Tecnológico de Tijuana

Nombre de Facultad

Ingeniería Informática

Proyecto / Tarea / Practica:

Practica 2 Unidad 3

Materia:

Minería de datos

Facilitador:

JOSE CHRISTIAN ROMERO HERNANDEZ

Alumnos:

Erik Saul Rivera Reyes

Fecha:

Tijuana Baja California a 26 de 11 2021

Código

```r
# Importing the dataset
dataset <- read.csv('50_Startups.csv')

# Encoding categorical data
dataset$State = factor(dataset$State,
                       levels = c('New York', 'California', 'Florida'),
                       labels = c(1,2,3))

dataset
# Splitting the dataset into the Training set and Test set
# Install.packages('caTools')
library(caTools)
set.seed(123)
split <- sample.split(dataset$Profit, SplitRatio = 0.8)
training_set <- subset(dataset, split == TRUE)
test_set <- subset(dataset, split == FALSE)

# Fitting Multiple Linear Regression to the Training set
#regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend + State)
regressor = lm(formula = Profit ~ .,
               data = training_set )

summary(regressor)

# Prediction the Test set results
y_pred = predict(regressor, newdata = test_set)
y_pred

# Assigment: visualize the siple liner regression model with R.D.Spend

# Building the optimal model using Backward Elimination
regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend + State,
               data = dataset )
summary(regressor)

regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend,
               data = dataset )
summary(regressor)

regressor = lm(formula = Profit ~ R.D.Spend + Marketing.Spend,
               data = dataset )
summary(regressor)

```

```
regressor = lm(formula = Profit ~ R.D.Spend + Marketing.Spend,
               data = dataset )
summary(regressor)

y_pred = predict(regressor, newdata = test_set)
y_pred

# Homework analise the follow atomation backwardElimination function
backwardElimination <- function(x, sl) {
  numVars = length(x)
  for (i in c(1:numVars)){
    regressor = lm(formula = Profit ~ ., data = x)
    maxVar = max(coef(summary(regressor))[c(2:numVars), "Pr(>|t|)"])
    if (maxVar > sl){
      j = which(coef(summary(regressor))[c(2:numVars), "Pr(>|t|)"] == maxVar)
      x = x[, -j]
    }
    numVars = numVars - 1
  }
  return(summary(regressor))
}

SL = 0.05
#dataset = dataset[, c(1,2,3,4,5)]
training_set
backwardElimination(training_set, SL)
```

Explicación del “Automated Backward Elimination”

Código para automatizar la eliminación hacia atrás en p <.05 Las principales contribuciones sobre los métodos ya implementados en R se encuentran en el tratamiento de las interacciones. Primero eliminará todos los términos NS de mayor potencia (por ejemplo, con cuatro interacciones frente a tres). Además, mantiene NS términos de orden inferior de términos significativos de orden superior.