

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

ENERO - JUNIO 2020

INGENIERÍA INFORMÁTICA

MATERIA

DATOS MASIVOS

CATEDRÁTICO:

JOSE CHRISTIAN ROMERO HERNANDEZ

UNIDAD 2

# PRACTICE 3

ALUMNO

López Valencia Luis Daniel #15212159

setwd("C:/Users/Daniel/Desktop/Practicas R")

getwd()

**Importing the dataset**

dataset <- read.csv(file.choose())

**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**

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

**Building the optimal model using Backward Elimination, the regression formula is declared using the following formula, returns the data summary of the function**

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

**The regression formula is declared using the following formula, returns the summary of data of the function**

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

summary(regressor)

**The regression formula is declared using the following formula, returns the summary of data of the function**

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

summary(regressor)

**The regression formula is declared using the following formula, returns the summary of data of the function**

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

summary(regressor)

**The method for prediction is performed with 2 parameters, the result of the prediction is printed**

y\_pred = predict(regressor, newdata = test\_set)

y\_pred