



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO



**TECNOLÓGICO NACIONAL DE MÉXICO**  
**INSTITUTO TECNOLÓGICO DE TIJUANA**  
**SUBDIRECCIÓN ACADÉMICA**  
**DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN**

SEMESTRE ( SEPTIEMBRE, ENERO 2021)

**CARRERA**

INGENIERÍA EN SISTEMAS COMPUTACIONALES

**MATERIA**

MINERÍA DE DATOS

**TÍTULO**

REGRESIÓN LINEAL MÚLTIPLE

**UNIDAD A EVALUAR**

UNIDAD 3

**NOMBRE DEL ALUMNO**

ARROYO GONZALEZ ERICK MANUEL

RAMOS ROBLES KATHY ABIGAIL

**NOMBRE DEL MAESTRO**

JOSE CHRISTIAN ROMERO HERNANDEZ

**FECHA DE ENTREGA**

26 NOVIEMBRE, 2021



## Introducción

Para esta práctica se realiza un ejemplo de regresión lineal múltiple en donde se eliminan del modelo las variables menos significativas.

### Importando los saltos

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

### Codificando los datos categoricos

```
dataset$State = factor(dataset$State,  
                        levels = c('New York', 'California', 'Florida'),  
                        labels = c(1,2,3))  
  
dataset
```

### Dividiendo los datos en entrenamiento y pruebas

```
# 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)
```

### Modelo de regresión

```
regressor = lm(formula = Profit ~ .,  
               data = training_set )  
  
summary(regressor)
```

### Predicción de los valores de prueba

```
y_pred = predict(regressor, newdata = test_set)  
y_pred
```

### Se eliminan variables menos significativas

```
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
```

## Eliminación de variables automática

```
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  
  
training_set  
backwardElimination(training_set, SL)
```