

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

ENERO - JUNIO 2020

INGENIERÍA INFORMÁTICA

MATERIA

DATOS MASIVOS

CATEDRÁTICO:

JOSE CHRISTIAN ROMERO HERNANDEZ

UNIDAD 2

# PRACTICE 4

ALUMNO

López Valencia Luis Daniel #15212159

**1. First, we need to select a level of significance to stay in the model. (SL = 0.05)**

**2. Fit the complete model with all possible predictors / independent variables.**

**3. Choose the predictor that has the highest P value, such that, if the P value> SL, go to step 4.**

**4. Remove that predictor.**

**5. Reconstruct and fit the model with the remaining variables.**

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 #nivel de significancia

#dataset = dataset[, c(1,2,3,4,5)]

training\_set

backwardElimination(training\_set, SL)