

Instituto Tecnológico de Tijuana

Nombre de Facultad

Ingeniería Informática



Proyecto / Tarea / Practica:

Practica 3 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

```
# Importing the dataset
dataset <- read.csv('Social_Network_Ads.csv')
dataset <- dataset[, 3:5]

# Splitting the dataset into the Training set and Test set
# Install.packages('caTools')
library(caTools)
set.seed(123)
split <- sample.split(dataset$Purchased, SplitRatio = 0.75)
training_set <- subset(dataset, split == TRUE)
test_set <- subset(dataset, split == FALSE)

# Feature scaling
training_set[, 1:2] <- scale(training_set[, 1:2])
test_set[, 1:2] <- scale(test_set[, 1:2])

# Fitting Logistic Regression to Training set
classifier = glm(formula = Purchased ~ .,
                  family = binomial,
                  data = training_set)

# Predicting the Test set results
prob_pred = predict(classifier, type = 'response',
                    newdata = test_set[-3])
prob_pred
```

```

y_pred = ifelse(prob_pred > 0.5, 1, 0)
y_pred

# Making the Confusion Metrix
cm = table(test_set[, 3], y_pred)
cm

#
library(ggplot2)
ggplot(training_set, aes(x=EstimatedSalary,
y=Purchased)) + geom_point() +
  stat_smooth(method="glm",
method.args=list(family="binomial"), se=FALSE)

ggplot(training_set, aes(x=Age, y=Purchased)) +
geom_point() +
  stat_smooth(method="glm",
method.args=list(family="binomial"), se=FALSE)

ggplot(test_set, aes(x=EstimatedSalary, y=Purchased)) +
geom_point() +
  stat_smooth(method="glm",
method.args=list(family="binomial"), se=FALSE)

ggplot(test_set, aes(x=Age, y=Purchased)) + geom_point()
+
  stat_smooth(method="glm",
method.args=list(family="binomial"), se=FALSE)

# Visualization the Training set result

```

```
# install.packages('ElemStatLearn') No work for me,  
# manual mode. Go to this URL https://cran.r-project.org/src/contrib/Archive/ElemStatLearn/  
# Download with the latest date 2019-08-12 09:20 12M  
# Then follow this page steps  
https://riptutorial.com/r/example/5556/install-package-from-local-source
```

```
library(ElemStatLearn)  
set = training_set  
X1 = seq(min(set[, 1]) - 1, max(set[, 1]) + 1, by = 0.01)  
X2 = seq(min(set[, 2]) - 1, max(set[, 2]) + 1, by = 0.01)  
grid_set = expand.grid(X1, X2)  
colnames(grid_set) = c('Age', 'EstimatedSalary')  
prob_set = predict(classifier, type = 'response',  
newdata = grid_set)  
y_grid = ifelse(prob_set > 0.5, 1, 0)  
plot(set[, -3],  
      main = 'Logistic Regression (Training set)',  
      xlab = 'Age', ylab = 'Estimated Salary',  
      xlim = range(X1), ylim = range(X2))  
contour(X1, X2, matrix(as.numeric(y_grid), length(X1),  
length(X2)), add = TRUE)  
points(grid_set, pch = '.', col = ifelse(y_grid == 1,  
'springgreen3', 'tomato'))  
points(set, pch = 21, bg = ifelse(set[, 3] == 1,  
'green4', 'red3'))
```

```
# Visualising the Test set results
```

```
library(ElemStatLearn)

set = test_set
X1 = seq(min(set[, 1]) - 1, max(set[, 1]) + 1, by = 0.01)
X2 = seq(min(set[, 2]) - 1, max(set[, 2]) + 1, by = 0.01)
grid_set = expand.grid(X1, X2)
colnames(grid_set) = c('Age', 'EstimatedSalary')
prob_set = predict(classifier, type = 'response',
newdata = grid_set)
y_grid = ifelse(prob_set > 0.5, 1, 0)
plot(set[, -3],
      main = 'Logistic Regression (Test set)',
      xlab = 'Age', ylab = 'Estimated Salary',
      xlim = range(X1), ylim = range(X2))
contour(X1, X2, matrix(as.numeric(y_grid), length(X1),
length(X2)), add = TRUE)
points(grid_set, pch = '.', col = ifelse(y_grid == 1,
'springgreen3', 'tomato'))
points(set, pch = 21, bg = ifelse(set[, 3] == 1,
'green4', 'red3'))
```

Grafica

Logistic Regression (Test set)

