Logistic Regression Case Study 2

Abhirup Sen

02/06/2021

#import the dataset

```
dataset = read.csv('Social_Network_Ads.csv')
dataset = dataset[3:5]
dataset$Purchased = factor(dataset$Purchased,levels = c(0,1))
```

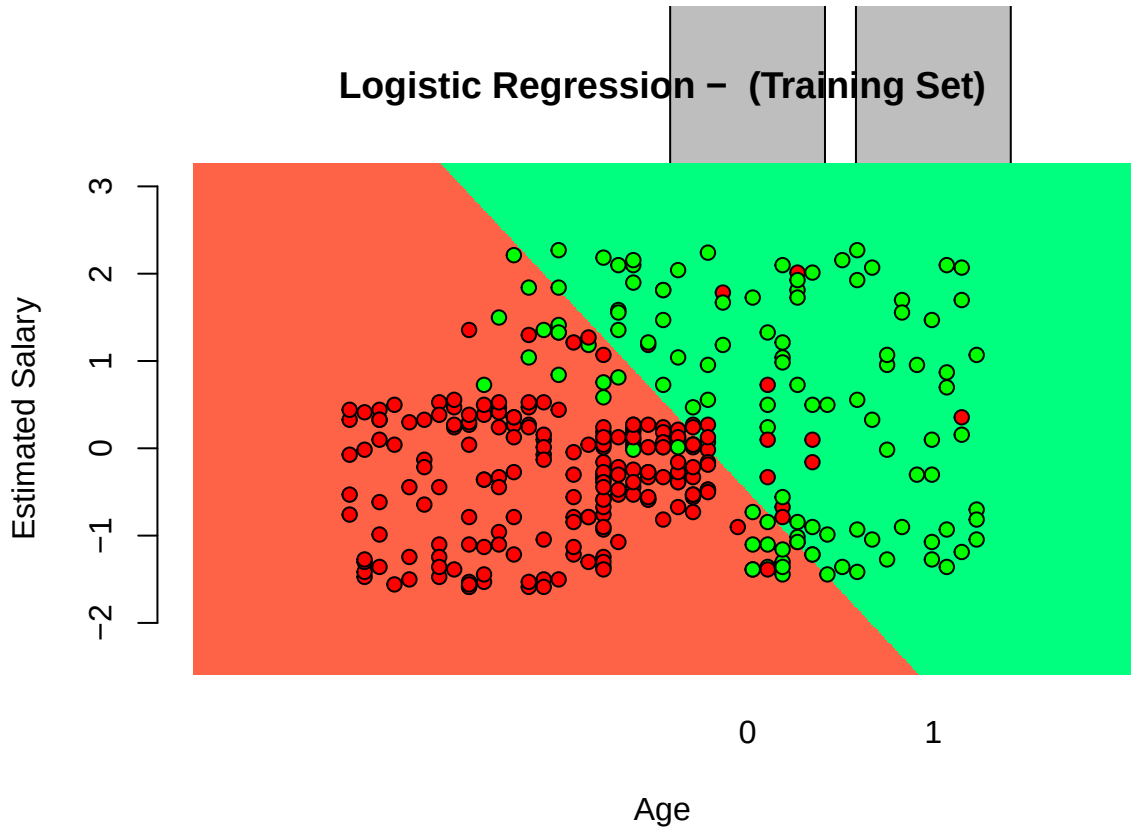
#Split

```
library(caTools)
set.seed(123)


split = sample.split(dataset$Purchased, SplitRatio = 0.8)
train_set = subset(dataset, split == TRUE)
test_set  = subset(dataset, split == FALSE)

#feature scaling
train_set[-3] = scale(train_set[-3])
test_set[-3] = scale(test_set[-3])


#fitting logistic regression to training set
classifier = glm(Purchased ~.,family ="binomial", data = train_set )

#predict the test set
prod_pred = predict(classifier,type = "response", newdata = test_set[-3] )
y_pred = ifelse(prod_pred>0.5,1,0)

#Evaluation - confusion matrxi
cm <- table(test_set[,3],y_pred >0.5)
cm
```

```
##
##     FALSE TRUE
##   0    44    7
##   1     9   20
```

```
#Visualization
set = train_set
X1 = seq(min(set[,1])-1, max(set[,1])+1,0.01)
```

```
X2 = seq(min(set[,2])-1, max(set[,2])+1,0.01)
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,'springgreen','tomato'))
points(set,pch = 21, bg = ifelse(set[,3]==1,'green','red'))
```




```
set = test_set
X1 = seq(min(set[,1])-1, max(set[,1])+1,0.01)
X2 = seq(min(set[,2])-1, max(set[,2])+1,0.01)
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,'springgreen','tomato'))
points(set,pch = 21, bg = ifelse(set[,3]==1,'green','red'))
```