R Notebook

```r
install.packages("dpp")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning: package 'dpp' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
## Warning: Perhaps you meant 'DPP' ?
```

```r
# Installing the package
install.packages("dplyr")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("dplyr"): installation of package 'dplyr' had
## non-zero exit status
```

```r
# For Logistic regression
install.packages("caTools")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("caTools"): installation of package 'caTools' had
## non-zero exit status
```

```r
# For ROC curve to evaluate model
install.packages("ROCR")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("ROCR"): installation of package 'ROCR' had non-zero
## exit status
```

```r
# Installing the package

# For Logistic regression
install.packages("caTools")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("caTools"): installation of package 'caTools' had
## non-zero exit status
```

```r
# For ROC curve to evaluate model
install.packages("ROCR")
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("ROCR"): installation of package 'ROCR' had non-zero
## exit status
    # Loading package
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# Summary of dataset in package
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##       am             gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```r
# Loading package
library(caTools)
library(ROCR)
```

```r
# Loading package
library(caTools)
library(ROCR)
```

```r
# Splitting dataset
split <- sample.split(mtcars, SplitRatio = 0.8)
split
```

```
##  [1] FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
```

```r
train_reg <- subset(mtcars, split == "TRUE")
test_reg <- subset(mtcars, split == "FALSE")

# Training model
logistic_model <- glm(vs ~ wt + disp,
                      data = train_reg,
                      family = "binomial")
logistic_model
```

```
##
## Call:  glm(formula = vs ~ wt + disp, family = "binomial", data = train_reg)
##
## Coefficients:
## (Intercept)           wt          disp
##     4.19333      0.52950      -0.02703
##
## Degrees of Freedom: 22 Total (i.e. Null);  20 Residual
## Null Deviance:       31.49
## Residual Deviance: 15.01     AIC: 21.01
```

```r
# Summary
summary(logistic_model)
```

```
##
## Call:
## glm(formula = vs ~ wt + disp, family = "binomial", data = train_reg)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9507  -0.3341   0.2994   0.4690   1.6555
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.19333    3.95116   1.061   0.2886
## wt           0.52950    1.90004   0.279   0.7805
## disp        -0.02703    0.01548  -1.746   0.0809 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 31.492  on 22  degrees of freedom
## Residual deviance: 15.013  on 20  degrees of freedom
## AIC: 21.013
##
## Number of Fisher Scoring iterations: 6
```

```r
predict_reg <- predict(logistic_model,
                       test_reg, type = "response")
predict_reg
```

```
##          Mazda RX4    Hornet Sportabout          Duster 360          Merc 450SE
##         0.778382747           0.023776512         0.025427735         0.248671051
## Lincoln Continental            Fiat 128          AMC Javelin       Porsche 914-2
##         0.004645554           0.961991355         0.099381608         0.888458175
##       Ford Pantera L
##         0.026218711
```

```r
# Changing probabilities
predict_reg <- ifelse(predict_reg >0.5, 1, 0)

# Evaluating model accuracy
# using confusion matrix
table(test_reg$vs, predict_reg)
```

```
##    predict_reg
##      0 1
##   0 6 2
##   1 0 1
```

```r
missing_classerr <- mean(predict_reg != test_reg$vs)
print(paste('Accuracy =', 1 - missing_classerr))
```

```
## [1] "Accuracy = 0.777777777777778"
```
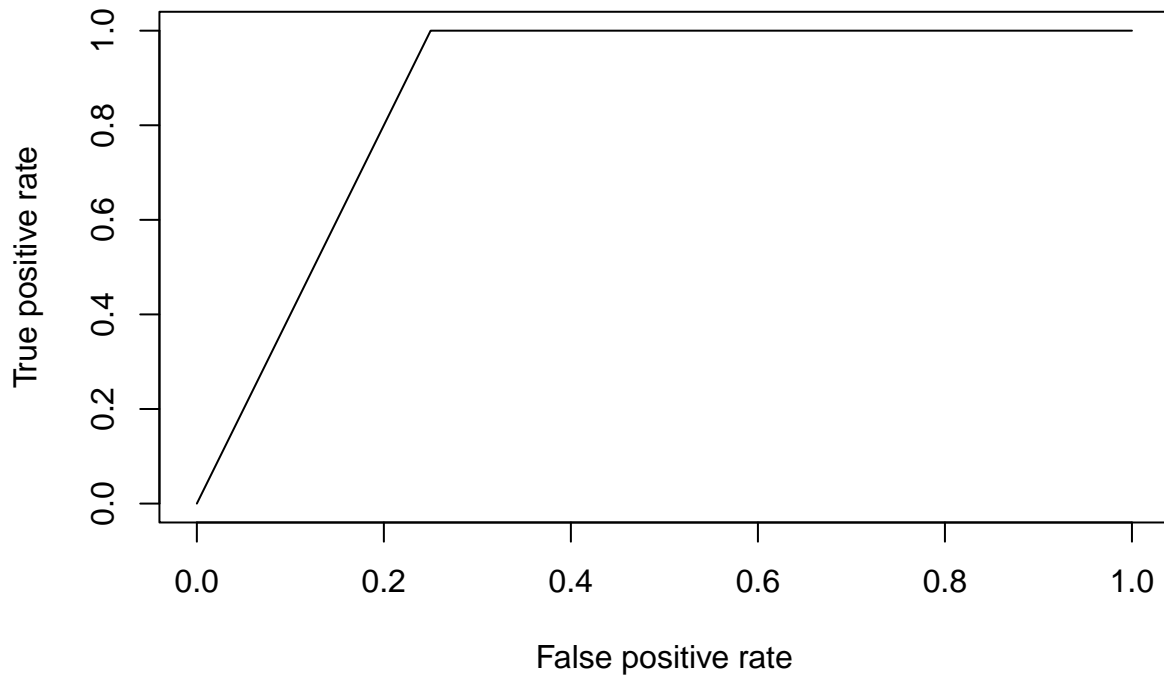
```r
# ROC-AUC Curve
ROCPred <- prediction(predict_reg, test_reg$vs)
ROCPer <- performance(ROCPred, measure = "tpr",
                      x.measure = "fpr")

auc <- performance(ROCPred, measure = "auc")
auc <- auc@y.values[[1]]
auc
```

```
## [1] 0.875
```

```r
# Plotting curve
plot(ROCPer)
```

```
plot(ROCPer, colorize = TRUE,
     print.cutoffs.at = seq(0.1, by = 0.1),
     main = "ROC CURVE")
abline(a = 0, b = 1)

auc <- round(auc, 4)
legend(.6, .4, auc, title = "AUC", cex = 1)
```

ROC CURVE