Titanic

Achraf cherkaoui

11/29/2021

Classification:

In this classification setting, we use a Titanic data set,it is a subset of a data set from https://www.kaggle.com/c/titanic. The Titanic data set contains information of 891 passenger with 11 variables (survived,Name,Pclass , Sex , Age , SibSp , Parch , Fare ,Embarked ,ticket, cabin). We will use this data in order to create a model that can predict if passengers are going to survive or not, using two different classifications method logistic regression and linear discriminant analysis.

data description:

survived ==> there are two levels 0 means the passanger did not survived and 1 means that the passenger did survive. 0 = No, 1 = Yes.

pclass ==> Ticket class, there are three levels 1 means the firs class, 2 means the second class , 3 means the third class. 1 = 1st, 2 = 2nd, 3 = 3rd

sex ==> Sex(gender) male or female

Age ==> Age in years.

sibsp ==> number of siblings or spouses aboard the Titanic.

parch ==> number of parents or children aboard the Titanic.

ticket ==> Ticket number.

fare ==> Passenger fare.

cabin ==> Cabin number.

embarked ==> Port of Embarkation. there are tree levels: C = Cherbourg, Q = Queenstown, S = Southampton.

First we remove the missing values from our data in order to make our study easier. since the cabin column has a lot of missing values we are going to remove this column from the data, also we remove the missing values for Age and embarked.

Then we split the data into training and test set. The training set should be used to build our machine learning models. in other words we are going to create a model based on our training data then we test our model on the test data.The test set should be used to see how well our model performs on unseen data.For each passenger in the test set, we use the model that we trained to predict whether or not they survived the sinking of the Titanic.

```
library(readxl)
library(MASS)
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggplot2)
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
t <- read_excel("traintitanic.xlsx")
#cleaning data
tt <- t[,-11] # since the cabin column has many missing values we are going to take it out from our dat
sum(is.na(tt))
```

```
## [1] 179
```

```
sum(is.na(tt$Embarked))
```

```
## [1] 2
```

```
twm <- na.omit(tt) # take out the missing values in age and embarked
sum(is.na(twm))
```

```
## [1] 0
```

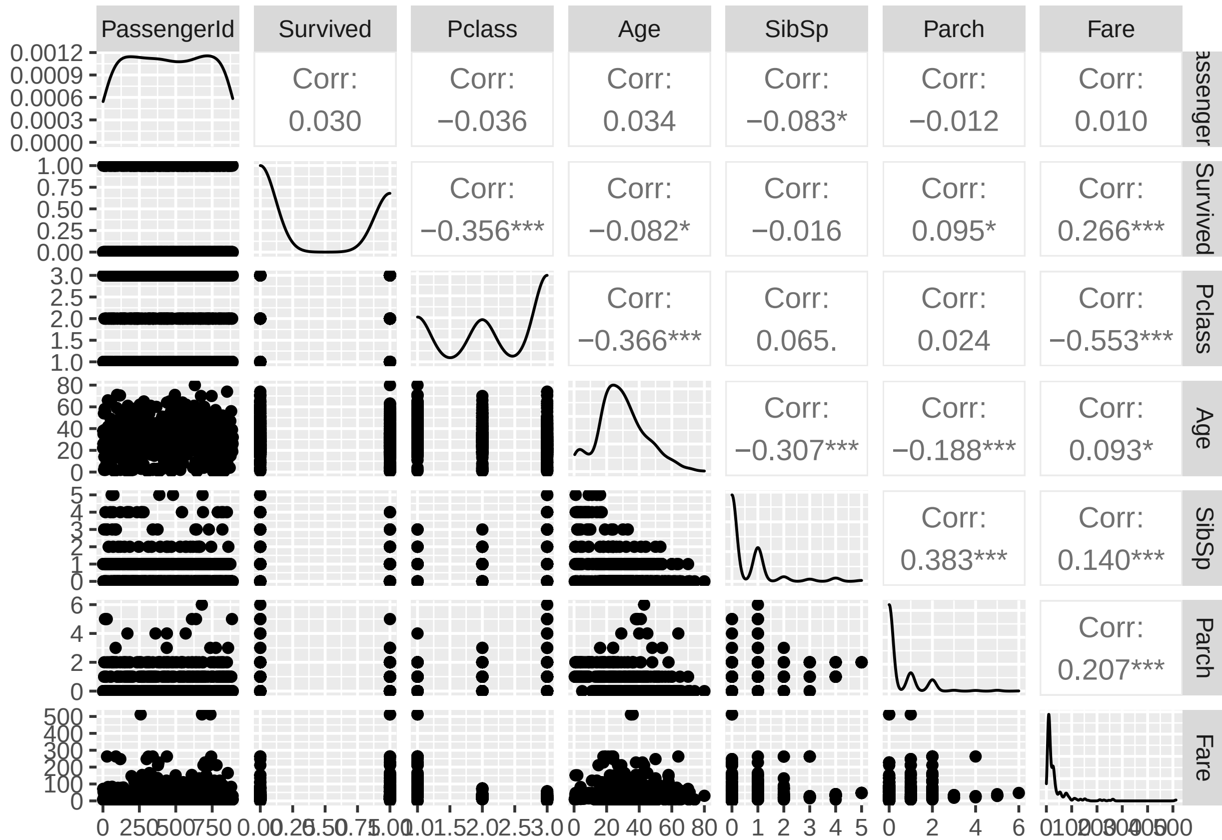
```
twm<- twm[,-4]
twm <-twm [,-8]
View(twm)

attach(twm)

## split the data into train and test data
n<- nrow(twm)
set.seed(123) # change the seed to your favorite number
reorder = sample(1:n) # shuffle
test = twm[reorder[1:(n/2)],]
train = twm[reorder[((n/2)+1):n],]

Survived <- as.factor(Survived)
Pclass <- as.factor(Pclass)
Sex <- as.factor(Sex)
Embarked <- as.factor(Embarked)

tw <-twm[,-4]
tm<- tw[,-8]
ggpairs(tm)
```



1) Logistic regression:

The logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval independent variables.It's used to calculate the likelihood of a given class or event, such as pass/fail, win/lose, alive/dead, or healthy/sick, occurring.

advantages

Logistic regression is easier to implement, interpret, and very efficient to train.also it can interpret model coefficients as indicators of feature importance.

disadvantages

The major limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables.Besides that,Logistic Regression requires average or no multicollinearity between independent variables.also it needs that independent variables are linearly related to the log odds (log(p/(1-p)).

```
twml <- glm( data = train ,  Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked , family =
summary(twml)
```

```
##
## Call:
## glm(formula = Survived ~ Pclass + Sex + Age + SibSp + Parch +
##     Fare + Embarked, family = "binomial", data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.9550  -0.6148  -0.3609   0.6084   2.4527
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.461212   0.908055   6.014 1.81e-09 ***
## Pclass      -1.091633   0.240372  -4.541 5.59e-06 ***
## Sexmale     -2.898020   0.318972  -9.086  < 2e-16 ***
## Age         -0.049230   0.012024  -4.094 4.23e-05 ***
## SibSp       -0.405389   0.199273  -2.034   0.0419 *
## Parch       -0.012155   0.193784  -0.063   0.9500
## Fare         0.003776   0.004214   0.896   0.3701
## EmbarkedQ   -0.527354   0.880925  -0.599   0.5494
## EmbarkedS   -0.060602   0.385361  -0.157   0.8750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 484.68  on 355  degrees of freedom
## Residual deviance: 310.47  on 347  degrees of freedom
## AIC: 328.47
##
## Number of Fisher Scoring iterations: 5
```

commentAfter fitting the logistic regression in the training set we can see that some variables are not significant in our model.Hence, we will take parch, fare, embarked off our model because there p-value is less than 0.05.

```
# forward stepwise
start <- glm(Survived~1 , data=train)
end <- glm(Survived~Pclass + Sex + Age + SibSp + Parch + Fare + Embarked ,data = train)
step.model <- step(start, direction = "forward" , scope = formula(end))
```

```
## Start:  AIC=511.84
## Survived ~ 1
##
##            Df Deviance    AIC
## + Sex       1   57.437 366.85
## + Pclass    1   79.083 480.71
## + Fare      1   81.845 492.93
## + Embarked  2   84.415 505.93
## + Parch     1   85.365 507.92
## + Age       1   85.911 510.19
## <none>          86.798 511.84
## + SibSp     1   86.797 513.84
##
## Step:  AIC=366.85
## Survived ~ Sex
##
##            Df Deviance    AIC
## + Pclass    1   53.248 341.89
## + Fare      1   55.598 357.27
## + Embarked  2   56.378 364.23
## <none>          57.437 366.85
## + Age       1   57.138 367.00
```

```
## + SibSp     1   57.213 367.46
## + Parch     1   57.411 368.69
##
## Step:  AIC=341.89
## Survived ~ Sex + Pclass
##
##            Df Deviance    AIC
## + Age       1   50.770 326.93
## <none>          53.248 341.89
## + SibSp     1   53.113 342.99
## + Fare      1   53.124 343.06
## + Parch     1   53.199 343.57
## + Embarked  2   53.032 344.45
##
## Step:  AIC=326.93
## Survived ~ Sex + Pclass + Age
##
##            Df Deviance    AIC
## + SibSp     1   50.143 324.51
## <none>          50.770 326.93
## + Fare      1   50.704 328.47
## + Parch     1   50.764 328.89
## + Embarked  2   50.671 330.24
##
## Step:  AIC=324.51
## Survived ~ Sex + Pclass + Age + SibSp
##
##            Df Deviance    AIC
## <none>          50.143 324.51
## + Fare      1   50.009 325.56
## + Parch     1   50.117 326.32
## + Embarked  2   50.067 327.96
```

```
summary(step.model)
```

```
##
## Call:
## glm(formula = Survived ~ Sex + Pclass + Age + SibSp, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.16856  -0.20706  -0.05898   0.20651   1.00779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.417642   0.096306  14.720  < 2e-16 ***
## Sexmale     -0.542398   0.042379 -12.799  < 2e-16 ***
## Pclass      -0.184269   0.027117  -6.795 4.65e-11 ***
## Age         -0.007338   0.001610  -4.559 7.11e-06 ***
## SibSp       -0.050138   0.023936  -2.095   0.0369 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1428587)
##
```

```
##     Null deviance: 86.798  on 355  degrees of freedom
## Residual deviance: 50.143  on 351  degrees of freedom
## AIC: 324.51
##
## Number of Fisher Scoring iterations: 2
```

```r
#backward stepwise
Bstep.model <- step(end, direction = "backward" , scope = formula(start))
```

```
## Start:  AIC=331.12
## Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked
##
##            Df Deviance    AIC
## - Embarked  2   49.999 327.48
## - Parch     1   49.958 329.19
## - Fare      1   50.043 329.80
## <none>          49.948 331.12
## - SibSp     1   50.612 333.82
## - Age       1   52.701 348.22
## - Pclass    1   53.819 355.69
## - Sex       1   71.927 458.94
##
## Step:  AIC=327.48
## Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare
##
##          Df Deviance    AIC
## - Parch   1   50.009 325.56
## - Fare    1   50.117 326.32
## <none>        49.999 327.48
## - SibSp   1   50.686 330.34
## - Age     1   52.858 345.28
## - Pclass  1   54.235 354.43
## - Sex     1   72.241 456.49
##
## Step:  AIC=325.56
## Survived ~ Pclass + Sex + Age + SibSp + Fare
##
##          Df Deviance    AIC
## - Fare    1   50.143 324.51
## <none>        50.009 325.56
## - SibSp   1   50.704 328.47
## - Age     1   52.945 343.87
## - Pclass  1   54.235 352.43
## - Sex     1   72.887 457.66
##
## Step:  AIC=324.51
## Survived ~ Pclass + Sex + Age + SibSp
##
##          Df Deviance    AIC
## <none>        50.143 324.51
## - SibSp   1   50.770 326.93
## - Age     1   53.113 342.99
## - Pclass  1   56.740 366.51
## - Sex     1   73.545 458.86
```

```r
summary(Bstep.model)
```

```
##
## Call:
## glm(formula = Survived ~ Pclass + Sex + Age + SibSp, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.16856  -0.20706  -0.05898   0.20651   1.00779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.417642   0.096306  14.720  < 2e-16 ***
## Pclass      -0.184269   0.027117  -6.795 4.65e-11 ***
## Sexmale     -0.542398   0.042379 -12.799  < 2e-16 ***
## Age         -0.007338   0.001610  -4.559 7.11e-06 ***
## SibSp       -0.050138   0.023936  -2.095   0.0369 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1428587)
##
##     Null deviance: 86.798  on 355  degrees of freedom
## Residual deviance: 50.143  on 351  degrees of freedom
## AIC: 324.51
##
## Number of Fisher Scoring iterations: 2
```

```r
# both stepwise
bothstep.model <- step(start, direction = "both" , scope = formula(end))
```

```
## Start:  AIC=511.84
## Survived ~ 1
##
##            Df Deviance    AIC
## + Sex       1   57.437 366.85
## + Pclass    1   79.083 480.71
## + Fare      1   81.845 492.93
## + Embarked  2   84.415 505.93
## + Parch     1   85.365 507.92
## + Age       1   85.911 510.19
## <none>          86.798 511.84
## + SibSp     1   86.797 513.84
##
## Step:  AIC=366.85
## Survived ~ Sex
##
##            Df Deviance    AIC
## + Pclass    1   53.248 341.89
## + Fare      1   55.598 357.27
## + Embarked  2   56.378 364.23
## <none>          57.437 366.85
## + Age       1   57.138 367.00
## + SibSp     1   57.213 367.46
```

```
## + Parch     1   57.411 368.69
## - Sex       1   86.798 511.84
##
## Step:  AIC=341.89
## Survived ~ Sex + Pclass
##
##            Df Deviance    AIC
## + Age       1   50.770 326.93
## <none>          53.248 341.89
## + SibSp     1   53.113 342.99
## + Fare      1   53.124 343.06
## + Parch     1   53.199 343.57
## + Embarked  2   53.032 344.45
## - Pclass    1   57.437 366.85
## - Sex       1   79.083 480.71
##
## Step:  AIC=326.93
## Survived ~ Sex + Pclass + Age
##
##            Df Deviance    AIC
## + SibSp     1   50.143 324.51
## <none>          50.770 326.93
## + Fare      1   50.704 328.47
## + Parch     1   50.764 328.89
## + Embarked  2   50.671 330.24
## - Age       1   53.248 341.89
## - Pclass    1   57.138 367.00
## - Sex       1   73.833 458.25
##
## Step:  AIC=324.51
## Survived ~ Sex + Pclass + Age + SibSp
##
##            Df Deviance    AIC
## <none>          50.143 324.51
## + Fare      1   50.009 325.56
## + Parch     1   50.117 326.32
## - SibSp     1   50.770 326.93
## + Embarked  2   50.067 327.96
## - Age       1   53.113 342.99
## - Pclass    1   56.740 366.51
## - Sex       1   73.545 458.86
```

```
summary(bothstep.model)
```

```
##
## Call:
## glm(formula = Survived ~ Sex + Pclass + Age + SibSp, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.16856  -0.20706  -0.05898   0.20651   1.00779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.417642   0.096306  14.720  < 2e-16 ***
```

```
## Sexmale     -0.542398   0.042379 -12.799  < 2e-16 ***
## Pclass      -0.184269   0.027117  -6.795 4.65e-11 ***
## Age         -0.007338   0.001610  -4.559 7.11e-06 ***
## SibSp       -0.050138   0.023936  -2.095   0.0369 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1428587)
##
##     Null deviance: 86.798  on 355  degrees of freedom
## Residual deviance: 50.143  on 351  degrees of freedom
## AIC: 324.51
##
## Number of Fisher Scoring iterations: 2
```

```
#prediction
twm.pred<- predict(step.model,newdata = test ,type="response")
twm.pred
```

```
##             1             2             3             4             5
##  0.1609920230  1.1012812996 -0.0518223931  0.1715994355  0.1609920230
##             6             7             8             9            10
##  0.6985129087  0.6924050015  0.0802693134  1.0584812517  0.7620977202
##            11            12            13            14            15
##  0.7590549409  0.1903457356  0.4516508652  0.1169614542 -0.1949217419
##            16            17            18            19            20
##  1.0804965361  0.2498611032  0.0888382625  0.0223698308  0.3379222409
##            21            22            23            24            25
##  1.0132201619  0.4488066055  0.2592379947  0.0961766906  0.1548841158
##            26            27            28            29            30
##  1.0511428236  0.2800227582  0.2865532439  0.1830073074  0.3326222761
##            31            32            33            34            35
##  0.7694361483  0.5358369687  0.0644066336  0.1169614542  0.5578522531
##            36            37            38            39            40
##  0.7474208639  0.9765280212  0.5346064478  1.0205585901  0.0814998344
##            41            42            43            44            45
##  0.3766528450  0.1397846811  0.1402072595  0.6752671034  1.0572507308
##            46            47            48            49            50
##  0.2425226750  0.5001756025  0.2139769050  0.2217379115  0.1683304511
##            51            52            53            54            55
##  0.6887134387  0.0705145408  0.0068850320  0.3171374774  0.9912048775
##            56            57            58            59            60
##  0.7995978033  0.2784068733  0.8236515512  0.1169614542  0.6821829531
##            61            62            63            64            65
##  0.8224210303  0.9769505997  0.7327440076 -0.0138997315  0.1096230260
##            66            67            68            69            70
##  0.1463151667  0.3746143816  0.8729820848  0.7400824357  0.4267913211
##            71            72            73            74            75
##  0.0655924572  0.9977353632  1.0719275871  0.1324462529  0.2050225919
##            76            77            78            79            80
##  0.8729820848  0.9243510818  0.0655924572  0.0081155529  0.8375204650
##            81            82            83            84            85
##  0.9141737307  0.3460686115  0.1316383104  0.5395285314  0.1699463361
##            86            87            88            89            90
##  0.3085685283  0.1923841990  0.2425226750  0.2050225919  0.1683304511
```

```
##           91           92           93           94           95 
## 0.1022845979 0.3974376085 0.7849209471 0.6340849042 0.6385749627 
##           96           97           98           99          100 
## 0.6373444418 0.8142746596 0.1463151667 0.0729308853 0.1683304511 
##          101          102          103          104          105 
## 0.9104821680 0.4634834618 0.0655924572 0.7180671513 0.2502836816 
##          106          107          108          109          110 
## 0.1251078248 0.1463151667 0.6092212501 0.0986377325 0.0925298253 
##          111          112          113          114          115 
## 0.1830073074 0.2205073906 0.9789890631 0.7107287232 0.3746143816 
##          116          117          118          119          120 
## 0.1022845979 0.3193553023 0.4561450336 0.4854987462 0.9178205961 
##          121          122          123          124          125 
## 0.8155051805 0.7547592920 0.5786370166 0.0374692655 0.0313613583 
##          126          127          128          129          130 
## 0.1426459526 0.2050225919 0.0142234602 0.0435771727 0.3460686115 
##          131          132          133          134          135 
## 0.2873611864 0.1756688793 0.5835591003 0.9691895931 0.1169614542 
##          136          137          138          139          140 
## 0.5982359566 0.1976841637 0.1072066815 0.1609920230 0.1903457356 
##          141          142          143          144          145 
## 0.7861514680 0.0289003165 0.2812532791 0.3693144168 0.3167148989 
##          146          147          148          149          150 
## 1.1159581559 0.1096230260 0.6740365825 0.0876077416 0.3672759534 
##          151          152          153          154          155 
## 0.4967284835 0.4928371743 1.0438043954 0.2437531960 0.9263895452 
##          156          157          158          159          160 
## 0.0876077416 0.0802693134 1.1159581559 0.3827607522 0.5639601603 
##          161          162          163          164          165 
## 0.4047760366 0.0876077416 0.1316383104 0.1976841637 -0.0298071087 
##          166          167          168          169          170 
## 0.8016362667 0.7590549409 0.0876077416 0.8970358326 0.6616210216 
##          171          172          173          174          175 
## 0.7555672345 0.9324974524 0.2739148510 0.7555672345 0.1389767386 
##          176          177          178          179          180 
## 0.1683304511 0.1316383104 0.2429452535 0.2062531128 0.1039004828 
##          181          182          183          184          185 
## 0.5346064478 0.9842890278 0.1475456876 0.8501588579 0.4121144648 
##          186          187          188          189          190 
## 0.6202512410 0.4781603181 -0.1545603871 0.1022845979 0.1242998823 
##          191          192          193          194          195 
## 0.3525990972 0.3012301002 0.0876077416 0.4844679717 0.0460382146 
##          196          197          198          199          200 
## 0.4697911154 0.4133449857 0.0998682534 0.6104517710 1.1159581559 
##          201          202          203          204          205 
## 0.1255304032 0.1242998823 0.4624526873 0.4573755545 0.3892912379 
##          206          207          208          209          210 
## 0.4011291712 0.2213153331 0.0517235434 0.1609920230 0.1316383104 
##          211          212          213          214          215 
## 0.3892912379 0.9903969350 0.7995978033 0.8803205129 0.4280218420 
##          216          217          218          219          220 
## 0.1609920230 0.0289003165 0.6813750106 0.7327440076 1.0425738745 
##          221          222          223          224          225 
## 0.1475456876 0.3387301834 0.5945443938 0.0802693134 0.1022845979 
```

```
##           226           227           228           229           230 
##  0.6826055315 -0.0004533961  0.6238981064  0.0949461697  0.8228436087 
##           231           232           233           234           235 
##  0.0876077416  0.1316383104  0.4928371743  0.1426683013  0.0876077416 
##           236           237           238           239           240 
##  0.5148524588  0.2731069085  0.5639601603  0.1699463361  0.1609920230 
##           241           242           243           244           245 
##  0.4573755545  0.2792148157  0.7702440908  0.0081155529 -0.1105298182 
##           246           247           248           249           250 
##  0.1022845979  0.1169614542  0.2205073906  0.7033902950  0.9557432577 
##           251           252           253           254           255 
##  0.1830073074 -0.0224686805  0.5148524588  0.1830073074  0.3012301002 
##           256           257           258           259           260 
##  0.8016362667  0.6887134387  0.8436283722  0.3379222409  0.8008283243 
##           261           262           263           264           265 
##  0.0802693134  0.9997738266  0.1756688793  0.1683304511  0.2290763397 
##           266           267           268           269           270 
##  0.5014061234  0.6496049536  0.8228436087  0.6324670554  0.2237763749 
##           271           272           273           274           275 
##  0.8595357494  0.9985433057  0.6960518669  0.9398358805  0.3746143816 
##           276           277           278           279           280 
## -0.0518223931  0.1096230260  0.2180686975  0.9537047943  0.1316383104 
##           281           282           283           284           285 
##  0.1242998823  0.8089746949  0.0301308374  0.4991448280  0.1463151667 
##           286           287           288           289           290 
##  0.1169614542  0.1022845979  0.3607454678  0.7727051326  0.2417147326 
##           291           292           293           294           295 
##  0.6018828220 -0.0493613513  0.3978601869  0.1984921062  0.8281435734 
##           296           297           298           299           300 
##  0.3232453846  0.0435771727  0.8077441740  0.1903457356  0.2812532791 
##           301           302           303           304           305 
##  1.1232965841  0.8069362315  0.6997434296  0.7119592441  0.3085685283 
##           306           307           308           309           310 
##  0.2005305696  0.3012301002  0.1683304511  0.1072066815  0.2678069438 
##           311           312           313           314           315 
##  0.0631761127  0.2885917073  0.1756688793  0.1548841158  0.1830073074 
##           316           317           318           319           320 
##  0.0521461218  0.2359921894 -0.0787150638  0.7327440076 -0.0444839650 
##           321           322           323           324           325 
##  1.0217891110  1.0144506829  0.8289515159  0.6687366177  1.0010043475 
##           326           327           328           329           330 
##  0.0814998344  0.8737900272  0.9410664014  0.1463151667  1.0939428715 
##           331           332           333           334           335 
##  0.4928371743  0.7853435255  0.1756688793  0.1756688793  0.1104309685 
##           336           337           338           339           340 
##  0.1830073074  0.1251078248  0.7547592920  0.3024606211  0.0056992085 
##           341           342           343           344           345 
##  0.6666981543  0.1830073074  0.9691895931  0.1683304511  0.0876077416 
##           346           347           348           349           350 
##  0.9190511170  0.2649605379  0.1536535949  0.3232453846  0.4121144648 
##           351           352           353           354           355 
## -0.1252066745 -0.2206062404  0.2731069085  0.7408903782  0.7482288064 
##           356 
##  0.3746143816 
```

```r
twml.pred.test <- ifelse(twm.pred > 0.5 , 1,0)
twml.pred.test
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20 
##   0   1   0   0   0   1   1   0   1   1   1   0   0   0   0   1   0   0   0   0 
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40 
##   1   0   0   0   0   1   0   0   0   0   1   1   0   0   1   1   1   1   1   0 
##  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60 
##   0   0   0   1   1   0   1   0   0   0   1   0   0   0   1   1   0   1   0   1 
##  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80 
##   1   1   1   0   0   0   0   1   1   0   0   1   1   0   0   1   1   0   0   1 
##  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 
##   1   0   0   1   0   0   0   0   0   0   0   0   1   1   1   1   1   0   0   0 
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 
##   1   0   0   1   0   0   0   1   0   0   0   0   1   1   0   0   0   0   0   1 
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 
##   1   1   1   0   0   0   0   0   0   0   0   0   1   1   0   1   0   0   0   0 
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 
##   1   0   0   0   0   1   0   1   0   0   0   0   1   0   1   0   0   1   0   1 
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 
##   0   0   0   0   0   1   1   0   1   1   1   1   0   1   0   0   0   0   0   0 
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 
##   1   1   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0   1   1 
## 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 
##   0   0   0   0   0   0   0   0   0   0   0   1   1   1   0   0   0   1   1   1 
## 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 
##   0   0   1   0   0   1   0   1   0   1   0   0   0   0   0   1   0   1   0   0 
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 
##   0   0   1   0   0   0   0   0   1   1   0   0   1   0   0   1   1   1   0   1 
## 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 
##   0   1   0   0   0   1   1   1   1   0   1   1   1   1   0   0   0   0   1   0 
## 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 
##   0   1   0   0   0   0   0   0   1   0   1   0   0   0   1   0   0   1   0   0 
## 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 
##   1   1   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0 
## 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 
##   1   1   1   1   1   0   1   1   0   1   0   1   0   0   0   0   0   1   0   0 
## 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 
##   1   0   1   0   0   1   0   0   0   0   0   0   0   1   1   0
```

```r
#confusion matrix
twm.table <- table(test$Survived,twml.pred.test )
twm.table
```

```
##    twml.pred.test
##       0   1
##   0 183  35
##   1  45  93
```

```r
sensitivity <- (93/(93+45))*100
specificity <- (183/(183+35))*100
sensitivity
```

```
## [1] 67.3913
```

```
specificity
```

```
## [1] 83.94495
```

```
testerrorrate <- ((45+35)/(183+35+45+93))*100
testerrorrate
```

```
## [1] 22.47191
```

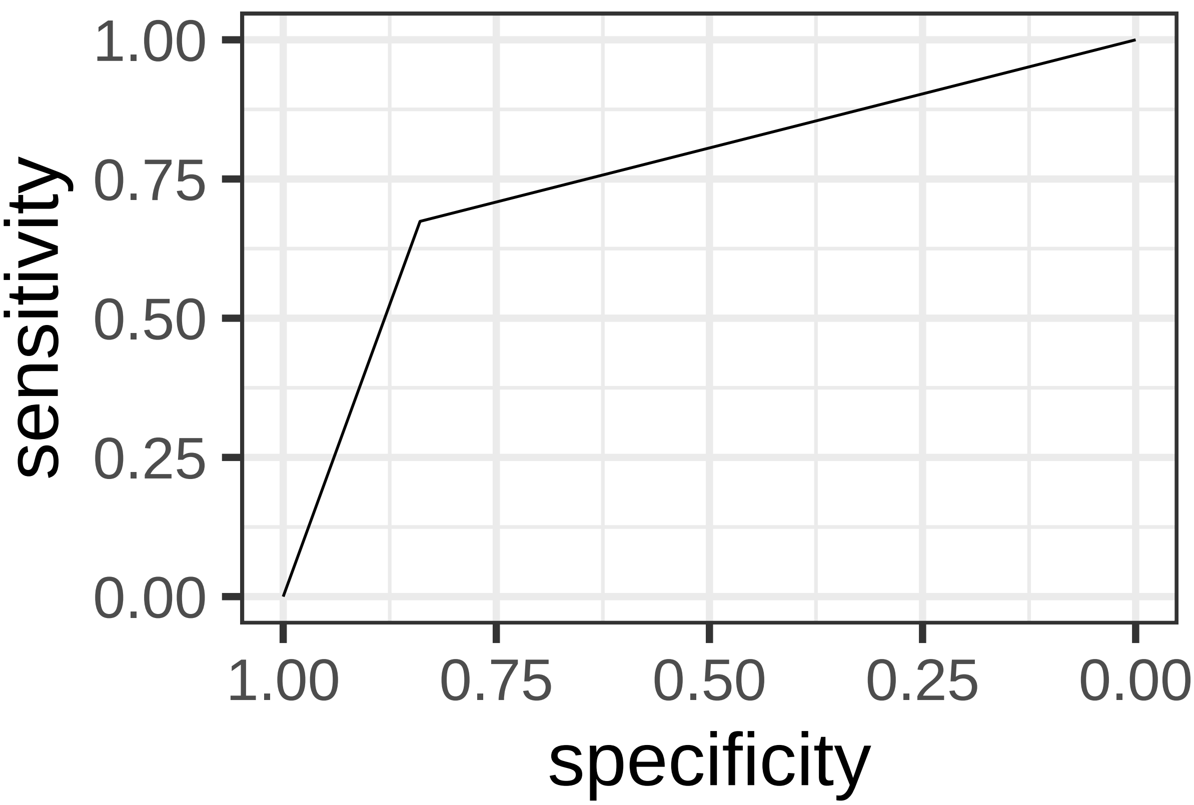
```
#area under the curve

twm.roc <- roc(test$Survived, twml.pred.test,legacy.axes=TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
ggroc(twm.roc)+theme_bw(28)
```




```
auc(twm.roc)
```

```
## Area under the curve: 0.7567
```

comment The area under the curve is between 0.8 and 0.7 witch means that the model is a good fit for our data.

2) LDA (linear discriminant analysis):

linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

advantages

It is unbias, simple, fast and easy to implement.

disadvantages

It requires normal distribution assumption on features/predictors and Sometimes not good for few categories variables.

```
LDA.twm <- lda(Survived~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked , data = train)
LDA.twm
```

```
## Call:
## lda(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked,
##     data = train)
##
## Prior probabilities of groups:
##         0         1
## 0.5786517 0.4213483
##
## Group means:
##     Pclass  Sexmale      Age     SibSp     Parch     Fare  EmbarkedQ EmbarkedS
## 0 2.412621 0.868932 31.88835 0.4854369 0.3009709 22.87020 0.03883495 0.8203883
## 1 1.913333 0.300000 28.96773 0.4800000 0.5000000 48.69775 0.02666667 0.7000000
##
## Coefficients of linear discriminants:
##                   LD1
## Pclass    -0.677996990
## Sexmale   -2.195385635
## Age       -0.029221005
## SibSp     -0.222918179
## Parch      0.030854981
## Fare       0.001546634
## EmbarkedQ -0.277145471
## EmbarkedS -0.091950225
```

```
#Prediction
la.pre <- predict(LDA.twm ,test)
la.pre
```

```
## $class
##   [1] 0 1 0 0 0 1 1 0 1 1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0 0 1 1 1
##  [38] 1 1 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 1 1 1 0 0 0 0 1 1 1 0 1 1 0
##  [75] 0 1 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 0 0 1 0 0 0 1 0 0 0
## [112] 0 1 1 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 1 0 0 0 0 1 0 1
## [149] 0 0 1 1 1 0 1 0 0 1 0 1 0 0 0 0 0 1 1 0 1 1 1 1 0 1 0 0 0 0 0 0 1 1 0 1 0
## [186] 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0
## [223] 1 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1 0 0 1 1 1 0
## [260] 1 0 1 0 0 0 0 1 1 1 0 1 1 1 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 1 0 1 0 0 0 1 0
## [297] 0 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 0 1 1 0 1 1 1 0
## [334] 0 0 0 0 1 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 1 0
## Levels: 0 1
##
## $posterior
##             0           1
## 1   0.915982161 0.084017839
```

```
## 2   0.013739535 0.986260465
## 3   0.979336194 0.020663806
## 4   0.910285510 0.089714490
## 5   0.916141170 0.083858830
## 6   0.184995086 0.815004914
## 7   0.212301687 0.787698313
## 8   0.941858429 0.058141571
## 9   0.016814118 0.983185882
## 10  0.138883788 0.861116212
## 11  0.122769788 0.877230212
## 12  0.898906517 0.101093483
## 13  0.609966775 0.390033225
## 14  0.936633752 0.063366248
## 15  0.994341639 0.005658361
## 16  0.014803672 0.985196328
## 17  0.865861001 0.134138999
## 18  0.949214977 0.050785023
## 19  0.968815714 0.031184286
## 20  0.777280964 0.222719036
## 21  0.022103207 0.977896793
## 22  0.646149800 0.353850200
## 23  0.845286998 0.154713002
## 24  0.945577169 0.054422831
## 25  0.915665203 0.084334797
## 26  0.012917857 0.987082143
## 27  0.846856391 0.153143609
## 28  0.832674708 0.167325292
## 29  0.902981627 0.097018373
## 30  0.788756268 0.211243732
## 31  0.115771013 0.884228987
## 32  0.365363967 0.634636033
## 33  0.954620031 0.045379969
## 34  0.928422566 0.071577434
## 35  0.386925863 0.613074137
## 36  0.197547586 0.802452414
## 37  0.030616961 0.969383039
## 38  0.381306289 0.618693711
## 39  0.018663746 0.981336254
## 40  0.944660435 0.055339565
## 41  0.719659376 0.280340624
## 42  0.932039791 0.067960209
## 43  0.925496694 0.074503306
## 44  0.234248537 0.765751463
## 45  0.018088829 0.981911171
## 46  0.870887449 0.129112551
## 47  0.505193640 0.494806360
## 48  0.865095280 0.134904720
## 49  0.886741528 0.113258472
## 50  0.911991873 0.088008127
## 51  0.190233686 0.809766314
## 52  0.967268606 0.032731394
## 53  0.969320225 0.030679775
## 54  0.788612363 0.211387637
## 55  0.031533471 0.968466529
```

```
## 56  0.119127546 0.880872454
## 57  0.819381220 0.180618780
## 58  0.086252853 0.913747147
## 59  0.936361853 0.063638147
## 60  0.217870947 0.782129053
## 61  0.078499734 0.921500266
## 62  0.028510255 0.971489745
## 63  0.214563271 0.785436729
## 64  0.964668276 0.035331724
## 65  0.939537620 0.060462380
## 66  0.923537845 0.076462155
## 67  0.697230045 0.302769955
## 68  0.075327592 0.924672408
## 69  0.135902683 0.864097317
## 70  0.305789296 0.694210704
## 71  0.954790050 0.045209950
## 72  0.028976330 0.971023670
## 73  0.018868004 0.981131996
## 74  0.935258162 0.064741838
## 75  0.889072213 0.110927787
## 76  0.075327592 0.924672408
## 77  0.054387506 0.945612494
## 78  0.954674102 0.045325898
## 79  0.969667610 0.030332390
## 80  0.071391759 0.928608241
## 81  0.033827638 0.966172362
## 82  0.765969258 0.234030742
## 83  0.917055626 0.082944374
## 84  0.442000992 0.557999008
## 85  0.907980804 0.092019196
## 86  0.810404524 0.189595476
## 87  0.895565224 0.104434776
## 88  0.851910469 0.148089531
## 89  0.889453097 0.110546903
## 90  0.912024167 0.087975833
## 91  0.942353666 0.057646334
## 92  0.725196092 0.274803908
## 93  0.130182838 0.869817162
## 94  0.242792518 0.757207482
## 95  0.278384561 0.721615439
## 96  0.275852478 0.724147522
## 97  0.108893550 0.891106450
## 98  0.923157284 0.076842716
## 99  0.952618092 0.047381908
## 100 0.912137113 0.087862887
## 101 0.053712205 0.946287795
## 102 0.593565597 0.406434403
## 103 0.954674102 0.045325898
## 104 0.179073753 0.820926247
## 105 0.850380584 0.149619416
## 106 0.938186566 0.061813434
## 107 0.923462030 0.076537970
## 108 0.310160159 0.689839841
## 109 0.941587909 0.058412091
```

```
## 110 0.962096783 0.037903217
## 111 0.901959355 0.098040645
## 112 0.883501430 0.116498570
## 113 0.018022614 0.981977386
## 114 0.187059426 0.812940574
## 115 0.731152474 0.268847526
## 116 0.933190754 0.066809246
## 117 0.746578164 0.253421836
## 118 0.580191325 0.419808675
## 119 0.572217737 0.427782263
## 120 0.035358110 0.964641890
## 121 0.102536282 0.897463718
## 122 0.189520385 0.810479615
## 123 0.363354189 0.636645811
## 124 0.963072864 0.036927136
## 125 0.966528608 0.033471392
## 126 0.925205587 0.074794413
## 127 0.889380533 0.110619467
## 128 0.967809909 0.032190091
## 129 0.960833059 0.039166941
## 130 0.778431607 0.221568393
## 131 0.838738317 0.161261683
## 132 0.894590360 0.105409640
## 133 0.354386223 0.645613777
## 134 0.034288283 0.965711717
## 135 0.936617828 0.063382172
## 136 0.331548642 0.668451358
## 137 0.894114295 0.105885705
## 138 0.941756943 0.058243057
## 139 0.916059554 0.083940446
## 140 0.899029374 0.100970626
## 141 0.129842408 0.870157592
## 142 0.958657509 0.041342491
## 143 0.817726653 0.182273347
## 144 0.733115147 0.266884853
## 145 0.759557619 0.240442381
## 146 0.011857618 0.988142382
## 147 0.939579517 0.060420483
## 148 0.228580825 0.771419175
## 149 0.947636655 0.052363345
## 150 0.740225489 0.259774511
## 151 0.499595905 0.500404095
## 152 0.486527328 0.513472672
## 153 0.019466645 0.980533355
## 154 0.850903210 0.149096790
## 155 0.050762829 0.949237171
## 156 0.947640542 0.052359458
## 157 0.950114601 0.049885399
## 158 0.014578635 0.985421365
## 159 0.731301203 0.268698797
## 160 0.355207019 0.644792981
## 161 0.714980326 0.285019674
## 162 0.947578382 0.052421618
## 163 0.930328083 0.069671917
```

```
## 164 0.894504177 0.105495823
## 165 0.976036173 0.023963827
## 166 0.098648424 0.901351576
## 167 0.122769788 0.877230212
## 168 0.940780395 0.059219605
## 169 0.065880291 0.934119709
## 170 0.234868660 0.765131340
## 171 0.154910743 0.845089257
## 172 0.035316851 0.964683149
## 173 0.847350307 0.152649693
## 174 0.153947349 0.846052651
## 175 0.928389744 0.071610256
## 176 0.934622257 0.065377743
## 177 0.930326640 0.069673360
## 178 0.858278515 0.141721485
## 179 0.885107694 0.114892306
## 180 0.948699864 0.051300136
## 181 0.437085203 0.562914797
## 182 0.034742305 0.965257695
## 183 0.925839392 0.074160608
## 184 0.076324890 0.923675110
## 185 0.704539517 0.295460483
## 186 0.298751211 0.701248789
## 187 0.582327888 0.417672112
## 188 0.992535953 0.007464047
## 189 0.942353666 0.057646334
## 190 0.933557987 0.066442013
## 191 0.728337606 0.271662394
## 192 0.801572217 0.198427783
## 193 0.961486740 0.038513260
## 194 0.466433816 0.533566184
## 195 0.961583504 0.038416496
## 196 0.554024588 0.445975412
## 197 0.678732946 0.321267054
## 198 0.942813762 0.057186238
## 199 0.294294931 0.705705069
## 200 0.013568383 0.986431617
## 201 0.935486604 0.064513396
## 202 0.933541344 0.066458656
## 203 0.513433265 0.486566735
## 204 0.548076550 0.451923450
## 205 0.711316278 0.288683722
## 206 0.563002678 0.436997322
## 207 0.892557059 0.107442941
## 208 0.962072381 0.037927619
## 209 0.915950319 0.084049681
## 210 0.930347644 0.069652356
## 211 0.702702771 0.297297229
## 212 0.030437071 0.969562929
## 213 0.110950551 0.889049449
## 214 0.061809050 0.938190950
## 215 0.650406127 0.349593873
## 216 0.916115402 0.083884598
## 217 0.964474841 0.035525159
```

```
## 218 0.219802608 0.780197392
## 219 0.165083482 0.834916518
## 220 0.022777233 0.977222767
## 221 0.909782091 0.090217909
## 222 0.786617384 0.213382616
## 223 0.326479062 0.673520938
## 224 0.950002685 0.049997315
## 225 0.933190051 0.066809949
## 226 0.220656386 0.779343614
## 227 0.970792264 0.029207736
## 228 0.299058627 0.700941373
## 229 0.945084529 0.054915471
## 230 0.103791377 0.896208623
## 231 0.947620012 0.052379988
## 232 0.930299820 0.069700180
## 233 0.557154238 0.442845762
## 234 0.922436376 0.077563624
## 235 0.947646080 0.052353920
## 236 0.447050540 0.552949460
## 237 0.801346180 0.198653820
## 238 0.308094645 0.691905355
## 239 0.904597559 0.095402441
## 240 0.916002806 0.083997194
## 241 0.571631695 0.428368305
## 242 0.839618299 0.160381701
## 243 0.142098582 0.857901418
## 244 0.967331914 0.032668086
## 245 0.986158388 0.013841612
## 246 0.942470148 0.057529852
## 247 0.936624453 0.063375547
## 248 0.887043448 0.112956552
## 249 0.193970718 0.806029282
## 250 0.033224080 0.966775920
## 251 0.903478940 0.096521060
## 252 0.974746008 0.025253992
## 253 0.327901201 0.672098799
## 254 0.903572482 0.096427518
## 255 0.819066449 0.180933551
## 256 0.104590901 0.895409099
## 257 0.210753651 0.789246349
## 258 0.090722654 0.909277346
## 259 0.747420731 0.252579269
## 260 0.120223679 0.879776321
## 261 0.949894796 0.050105204
## 262 0.023544278 0.976455722
## 263 0.907483549 0.092516451
## 264 0.911832937 0.088167063
## 265 0.853791091 0.146208909
## 266 0.537471648 0.462528352
## 267 0.258074306 0.741925694
## 268 0.085687584 0.914312416
## 269 0.294069131 0.705930869
## 270 0.868989993 0.131010007
## 271 0.082471733 0.917528267
```

```
## 272 0.026404438 0.973595562
## 273 0.202847028 0.797152972
## 274 0.037344823 0.962655177
## 275 0.731152474 0.268847526
## 276 0.979390189 0.020609811
## 277 0.939564285 0.060435715
## 278 0.871333865 0.128666135
## 279 0.041641859 0.958358141
## 280 0.930326640 0.069673360
## 281 0.933694445 0.066305555
## 282 0.108835248 0.891164752
## 283 0.962522108 0.037477892
## 284 0.402788572 0.597211428
## 285 0.923480991 0.076519009
## 286 0.934247582 0.065752418
## 287 0.942141979 0.057858021
## 288 0.759984149 0.240015851
## 289 0.128955488 0.871044512
## 290 0.837900031 0.162099969
## 291 0.323415491 0.676584509
## 292 0.980593707 0.019406293
## 293 0.681140687 0.318859313
## 294 0.901406696 0.098593304
## 295 0.091689684 0.908310316
## 296 0.794336497 0.205663503
## 297 0.960755577 0.039244423
## 298 0.119556010 0.880443990
## 299 0.899040527 0.100959473
## 300 0.804506288 0.195493712
## 301 0.008788543 0.991211457
## 302 0.100351414 0.899648586
## 303 0.192534264 0.807465736
## 304 0.167075150 0.832924850
## 305 0.804987644 0.195012356
## 306 0.874934611 0.125065389
## 307 0.817870626 0.182129374
## 308 0.912007117 0.087992883
## 309 0.958221779 0.041778221
## 310 0.883529109 0.116470891
## 311 0.968835917 0.031164083
## 312 0.801741278 0.198258722
## 313 0.907809634 0.092190366
## 314 0.903777292 0.096222708
## 315 0.903453600 0.096546400
## 316 0.959267228 0.040732772
## 317 0.882288250 0.117711750
## 318 0.983836785 0.016163215
## 319 0.164217849 0.835782151
## 320 0.978285129 0.021714871
## 321 0.028136193 0.971863807
## 322 0.022396669 0.977603331
## 323 0.132333763 0.867666237
## 324 0.241303651 0.758696349
## 325 0.015519704 0.984480296
```

```
## 326 0.951593729 0.048406271
## 327 0.079353523 0.920646477
## 328 0.061448955 0.938551045
## 329 0.923604902 0.076395098
## 330 0.016629926 0.983370074
## 331 0.372264413 0.627735587
## 332 0.116908114 0.883091886
## 333 0.907853588 0.092146412
## 334 0.893773801 0.106226199
## 335 0.941932435 0.058067565
## 336 0.903515010 0.096484990
## 337 0.938108729 0.061891271
## 338 0.126560811 0.873439189
## 339 0.823351249 0.176648751
## 340 0.969283064 0.030716936
## 341 0.217295769 0.782704231
## 342 0.928323918 0.071676082
## 343 0.027547287 0.972452713
## 344 0.911964953 0.088035047
## 345 0.947636655 0.052363345
## 346 0.043820627 0.956179373
## 347 0.827805006 0.172194994
## 348 0.919538213 0.080461787
## 349 0.794336497 0.205663503
## 350 0.599593963 0.400406037
## 351 0.987509637 0.012490363
## 352 0.993471586 0.006528414
## 353 0.845935077 0.154064923
## 354 0.168651003 0.831348997
## 355 0.152275298 0.847724702
## 356 0.730360699 0.269639301
##
## $x
##             LD1
## 1   -1.05790566
## 2    2.78304544
## 3   -1.90507630
## 4   -1.01648937
## 5   -1.05909780
## 6    1.17418028
## 7    1.07516291
## 8   -1.28620068
## 9    2.66482758
## 10   1.37118367
## 11   1.45296880
## 12  -0.94039649
## 13   0.06152644
## 14  -1.23338635
## 15  -2.66054512
## 16   2.73941797
## 17  -0.75575398
## 18  -1.36867368
## 19  -1.66160512
## 20  -0.40123534
```

```
## 21   2.50404478
## 22  -0.02782147
## 23  -0.65962751
## 24  -1.32657695
## 25  -1.05553540
## 26   2.81907606
## 27  -0.66657461
## 28  -0.60578238
## 29  -0.96672405
## 30  -0.44017974
## 31   1.49138826
## 32   0.63763370
## 33  -1.43682027
## 34  -1.15806526
## 35   0.58465115
## 36   1.12738526
## 37   2.31116409
## 38   0.59834552
## 39   2.60357682
## 40  -1.31638781
## 41  -0.22418373
## 42  -1.19020247
## 43  -1.13314918
## 44   1.00216047
## 45   2.62195201
## 46  -0.78110840
## 47   0.30734164
## 48  -0.75196243
## 49  -0.86703632
## 50  -1.02863949
## 51   1.15436481
## 52  -1.63276976
## 53  -1.67130826
## 54  -0.43968196
## 55   2.29361485
## 56   1.47271924
## 57  -0.55243226
## 58   1.67999897
## 59  -1.23075058
## 60   1.05614440
## 61   1.73916857
## 62   2.35351396
## 63   1.06739661
## 64  -1.58714694
## 65  -1.26221420
## 66  -1.11696632
## 67  -0.16155895
## 68   1.76492929
## 69   1.38568501
## 70   0.79197343
## 71  -1.43908686
## 72   2.34388923
## 73   2.59718190
## 74  -1.22015812
```

```
## 75  -0.88053668
## 76   1.76492929
## 77   1.96560973
## 78  -1.43754023
## 79  -1.67807965
## 80   1.79831487
## 81   2.25176127
## 82  -0.36422373
## 83  -1.06599398
## 84   0.45366741
## 85  -1.00040522
## 86  -0.51811936
## 87  -0.91950376
## 88  -0.68935151
## 89  -0.88276647
## 90  -1.02887149
## 91  -1.29143520
## 92  -0.24010168
## 93   1.41427723
## 94   0.97503938
## 95   0.86842203
## 96   0.87570891
## 97   1.53116125
## 98  -1.11386656
## 99  -1.41072300
## 100 -1.02968347
## 101  1.97322410
## 102  0.10098590
## 103 -1.43754023
## 104  1.19710779
## 105 -0.68239020
## 106 -1.24864447
## 107 -1.11634767
## 108  0.78015037
## 109 -1.28335900
## 110 -1.54510675
## 111 -0.96002836
## 112 -0.84866517
## 113  2.62410503
## 114  1.16632082
## 115 -0.25745027
## 116 -1.20076095
## 117 -0.30355087
## 118  0.13278893
## 119  0.15161352
## 120  2.22533774
## 121  1.56993787
## 122  1.15703808
## 123  0.64263637
## 124 -1.56073165
## 125 -1.61944023
## 126 -1.13071968
## 127 -0.88234114
## 128 -1.64270591
```

```
## 129 -1.52544173
## 130 -0.40507420
## 131 -0.63124440
## 132 -0.91351947
## 133  0.66510734
## 134  2.24368897
## 135 -1.23323169
## 136  0.72354935
## 137 -0.91061483
## 138 -1.28513317
## 139 -1.05848565
## 140 -0.94117630
## 141  1.41601233
## 142 -1.49297100
## 143 -0.54601002
## 144 -0.26321973
## 145 -0.34379635
## 146  2.86906589
## 147 -1.26263952
## 148  1.02053162
## 149 -1.35007054
## 150 -0.28435133
## 151  0.32025030
## 152  0.35039362
## 153  2.57882331
## 154 -0.68476163
## 155  2.00757610
## 156 -1.35011570
## 157 -1.37952354
## 158  2.74838044
## 159 -0.25788654
## 160  0.66304028
## 161 -0.21088067
## 162 -1.34939389
## 163 -1.17480251
## 164 -0.91299278
## 165 -1.81771612
## 166  1.59471391
## 167  1.45296880
## 168 -1.27494928
## 169  1.84804364
## 170  1.00016930
## 171  1.29739321
## 172  2.22603548
## 173 -0.66877305
## 174  1.30164646
## 175 -1.15778059
## 176 -1.21413123
## 177 -1.17478968
## 178 -0.71898355
## 179 -0.85771623
## 180 -1.36254283
## 181  0.46517137
## 182  2.23583443
```

```
## 183 -1.13602046
## 184  1.75672479
## 185 -0.18165967
## 186  0.81121237
## 187  0.12772840
## 188 -2.49982960
## 189 -1.29143520
## 190 -1.20416535
## 191 -0.24922196
## 192 -0.48555270
## 193 -1.53553645
## 194  0.39683780
## 195 -1.53704472
## 196  0.19425097
## 197 -0.11187266
## 198 -1.29633626
## 199  0.82352822
## 200  2.79037193
## 201 -1.22233667
## 202 -1.20401068
## 203  0.28833426
## 204  0.20811165
## 205 -0.20055490
## 206  0.17325965
## 207 -0.90119329
## 208 -1.54472111
## 209 -1.05766717
## 210 -1.17497651
## 211 -0.17658207
## 212  2.31466822
## 213  1.51904056
## 214  1.88732499
## 215 -0.03858296
## 216 -1.05890448
## 217 -1.58388374
## 218  1.04963013
## 219  1.25374534
## 220  2.48633013
## 221 -1.01294460
## 222 -0.43280657
## 223  0.73678804
## 224 -1.37816374
## 225 -1.20075446
## 226  1.04676399
## 227 -1.70052927
## 228  0.81036668
## 229 -1.32108153
## 230  1.56211735
## 231 -1.34987721
## 232 -1.17455119
## 233  0.18694373
## 234 -1.10803297
## 235 -1.35018004
## 236  0.44187805
```

```
## 237 -0.48473377
## 238  0.78572593
## 239 -0.97743770
## 240 -1.05806032
## 241  0.15299347
## 242 -0.63500338
## 243  1.35583522
## 244 -1.63392362
## 245 -2.14008164
## 246 -1.29267251
## 247 -1.23329603
## 248 -0.86877142
## 249  1.14048308
## 250  2.26250004
## 251 -0.97000415
## 252 -1.78672298
## 253  0.73306375
## 254 -0.97062281
## 255 -0.55120695
## 256  1.55717901
## 257  1.08051376
## 258  1.64804527
## 259 -0.30612101
## 260  1.46672117
## 261 -1.37685560
## 262  2.46678323
## 263 -0.99698254
## 264 -1.02749885
## 265 -0.69799063
## 266  0.23274752
## 267  0.92809638
## 268  1.68414605
## 269  0.82415514
## 270 -0.77144041
## 271  1.70822238
## 272  2.39899758
## 273  1.10830413
## 274  2.19263432
## 275 -0.25745027
## 276 -1.90661644
## 277 -1.26248486
## 278 -0.78340056
## 279  2.12726844
## 280 -1.17478968
## 281 -1.20543482
## 282  1.53150771
## 283 -1.55186715
## 284  0.54637582
## 285 -1.11650233
## 286 -1.21060566
## 287 -1.28919258
## 288 -0.34514356
## 289  1.42055102
## 290 -0.62767891
```

```
## 291  0.74483949
## 292 -1.94201192
## 293 -0.11825091
## 294 -0.95643443
## 295  1.64131940
## 296 -0.45967735
## 297 -1.52425593
## 298  1.47036902
## 299 -0.94124714
## 300 -0.49624706
## 301  3.04352587
## 302  1.58375645
## 303  1.14579467
## 304  1.24545494
## 305 -0.49801312
## 306 -0.80214131
## 307 -0.54656705
## 308 -1.02874899
## 309 -1.48666474
## 310 -0.84882022
## 311 -1.66199077
## 312 -0.48616566
## 313 -0.99922516
## 314 -0.97197923
## 315 -0.96983665
## 316 -1.50190298
## 317 -0.84190064
## 318 -2.04933264
## 319  1.25737359
## 320 -1.87585529
## 321  2.36134970
## 322  2.49626808
## 323  1.40340270
## 324  0.97971790
## 325  2.71176805
## 326 -1.39777220
## 327  1.73239791
## 328  1.89091465
## 329 -1.11751399
## 330  2.67128567
## 331  0.62054471
## 332  1.48501174
## 333 -0.99952799
## 334 -0.90854441
## 335 -1.28698024
## 336 -0.97024265
## 337 -1.24787116
## 338  1.43293975
## 339 -0.56803146
## 340 -1.67058830
## 341  1.05809214
## 342 -1.15721002
## 343  2.37389331
## 344 -1.02844616
```

```
## 345 -1.35007054
## 346  2.09655587
## 347 -0.58586270
## 348 -1.08507088
## 349 -0.45967735
## 350  0.08654567
## 351 -2.20008962
## 352 -2.57758473
## 353 -0.66248929
## 354  1.23895121
## 355  1.30908031
## 356 -0.25513032
```

```
#confusion matrix
la.class <- la.pre$class
table(test$Survived ,la.class )
```

```
##    la.class
##       0   1
##   0 180  38
##   1  44  94
```

```
sensitivity <- (94/(94+44))*100
specificity <- (180/(180+38))*100
sensitivity
```

```
## [1] 68.11594
```

```
specificity
```

```
## [1] 82.56881
```

```
testerrorrate <- ((44+38)/(180+38+44+94))*100
testerrorrate
```

```
## [1] 23.03371
```

conclusion comparing the logistic regression and LDA we find that the test error rate for the logistic regression 22.47 is less than the LDA test error rate 23.03 meaning that the logistic regression is slightly more accurate than the LDA by 0.56.Moreover the AUC confirm that the logistic regression model is a good model to our data.Therefore,the best model that will be a a good fit to our data is the logistic regression.Forward selection, Backward selection and Stepwise selection give the same result estimating that our model will include 4 variables (Survived = Sex + Pclass + Age + SibSp), in other words the prediction of a passenger survival is essentially explained based on these 4 variables which are :the gender of the person male or female, the ticket class if it is first , second or third class, the age of the person , and finally if you have any family member aboard the ship).

Final Model

Now we fit our logistic regression model to the whole data:

```
final.model<- glm( data = twm ,  Survived ~ Pclass + Sex + Age + SibSp  , family = "binomial" )
summary(final.model)
```

```
##
## Call:
## glm(formula = Survived ~ Pclass + Sex + Age + SibSp, family = "binomial",
##     data = twm)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7694  -0.6496  -0.3839   0.6298   2.4585
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.59083    0.54342  10.288  < 2e-16 ***
## Pclass      -1.31392    0.14091  -9.324  < 2e-16 ***
## Sexmale     -2.61477    0.21473 -12.177  < 2e-16 ***
## Age         -0.04459    0.00817  -5.457 4.83e-08 ***
## SibSp       -0.37465    0.12093  -3.098  0.00195 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 960.90  on 711  degrees of freedom
## Residual deviance: 636.18  on 707  degrees of freedom
## AIC: 646.18
##
## Number of Fisher Scoring iterations: 5
```

comment interpretation of the estimates of 1 , 2, 3 and 4 .

1= -1.31392, then e 1 = 0.26877 and the interpretation becomes: An increase of one unite in Pclass multiplies the odds of class 1 by 0.26877.meaning that an increase of one unite in Pclass is associated with an decrease of 73.12% in the odds of class 1. decreases by 73.12% (0.26877 - 1 = 0.5386).

2= -2.61477, then e = 0.07318 and the interpretation becomes: An increase of one unite in sex multiplies the odds of class 1 by 0.07318.meaning that an increase of one unite in sex is associated with an decrease of 92.68% in the odds of class 1. decreases by 92.68% (0.07318 - 1 = -0.9268).

3= -0.04459, with one unite increase in Age leads to a decreases of 0.04459 in the log-odds of class 1. 3= -0.04459, then e 3 = 0.9564 and the interpretation becomes: An increase of one unite in Age multiplies the odds of class by 0.9564. an increase of one unite in Age is associated with a decrease of 4.36% in the odds of class 1. decreases by 4.36.8% (0.9564 - 1 = -0.0436).

4= -0.37465, with one unite increase in number of siblings leads to a decreases of 0.37465 in the log-odds of class 1. 4= -0.37465, then e 4 = 0.6875 and the interpretation becomes: An increase of one unite in number of siblings multiplies the odds of class by 0.6875. an increase of one unite in number of siblings is associated with a decrease of 31.25% in the odds of class 1.decreases by 31.25% (0.6875 - 1 = -0.3125).

comment Based on this model we see that men have less chances to survive a Titanic sinking compared to women.in other words if we remain all the other variable constant there is 92.68% chances that a woman might survive the Titanic sinking compared to a man.

```r
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```r
m <- twm%>%
        filter(Survived == "1" , Sex == "male")
mn <- nrow(m)
mn
```

```
## [1] 93
```

```r
print("93 of men survived Titinc sinking")
```

```
## [1] "93 of men survived Titinc sinking"
```

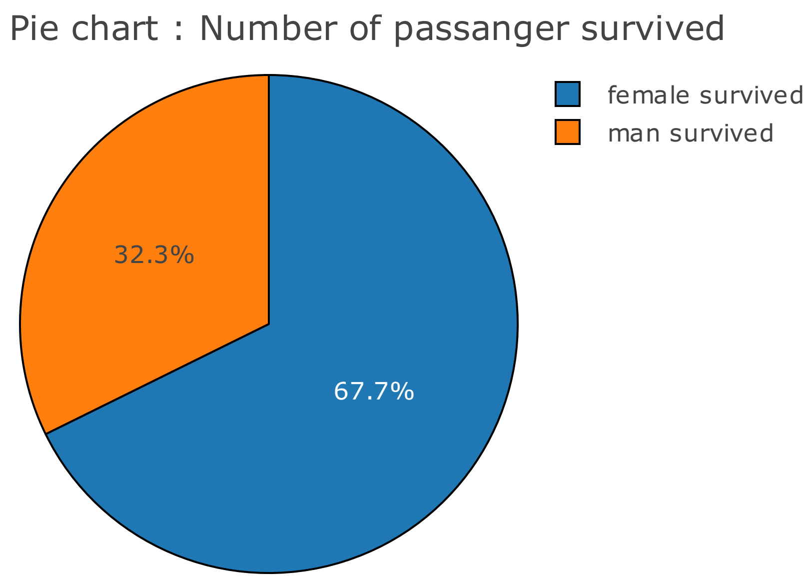
```r
w <- twm%>%
        filter(Survived == "1" , Sex == "female")
wn <- nrow(w)
wn
```

```
## [1] 195
```

```r
print("195 of women survived Titinc sinking")
```

```
## [1] "195 of women survived Titinc sinking"
```

```r
d <- data.frame(game = c("female survived", "man survived" ),
                number=c(wn, mn))
pi <- plot_ly(data = d, labels = ~game, values = ~number,
                type = 'pie', sort= FALSE,
                marker= list(colors=colors, line = list(color="black", width=1))) %%
  layout(title="Pie chart : Number of passanger survived ")
pi
```




```
m1 <- twm%>%
        filter(Survived == "0" , Sex == "male")
mn1 <- nrow(m1)
mn1
```

```
## [1] 360
```

```
print("360 of men did not survive Titinc sinking")
```

```
## [1] "360 of men did not survive Titinc sinking"
```

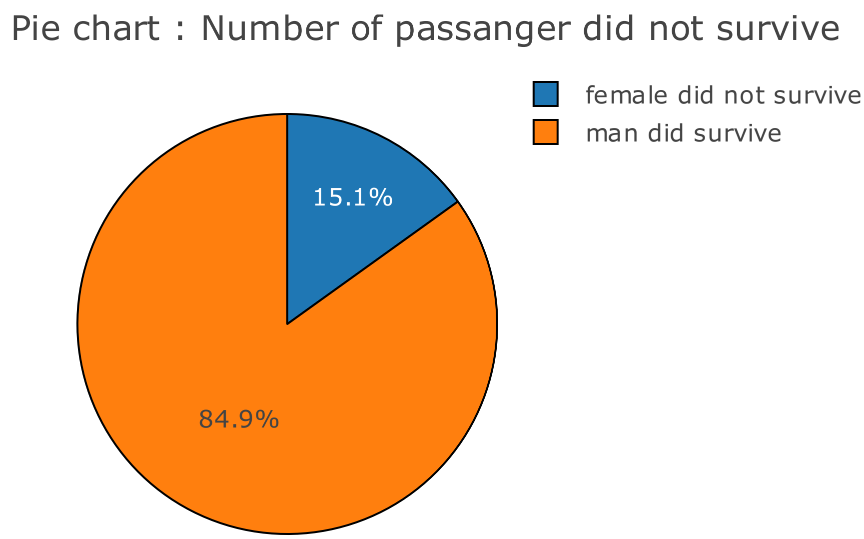
```
w1 <- twm%>%
        filter(Survived == "0" , Sex == "female")
wn1 <- nrow(w1)
wn1
```

```
## [1] 64
```

```
print("64 of women did not survive Titinc sinking")
```

```
## [1] "64 of women did not survive Titinc sinking"
```

```
d1 <- data.frame(game = c("female did not survive", "man did survive " ),
                number=c(wn1, mn1))
pi1 <- plot_ly(data = d1, labels = ~game, values = ~number,
                type = 'pie', sort= FALSE,
                marker= list(colors=colors, line = list(color="black", width=1))) %>%
  layout(title="Pie chart : Number of passanger did not survive ")
pi1
```


Pie chart : Number of passanger did not survive
female did not survive
man did survive
15.1%
84.9%