Airline passenger satisfaction

Caricamento delle librerie necessarie

```
library(MASS)
library(boot)
library(caret)
```

```
## Caricamento del pacchetto richiesto: ggplot2

## Caricamento del pacchetto richiesto: lattice

##
## Caricamento pacchetto: 'lattice'

## Il seguente oggetto è mascherato da 'package:boot':
##
##     melanoma
```

```
library(ggplot2)
library(readr)
library(dplyr)
```

```
##
## Caricamento pacchetto: 'dplyr'

## Il seguente oggetto è mascherato da 'package:MASS':
##
##     select

## I seguenti oggetti sono mascherati da 'package:stats':
##
##     filter, lag

## I seguenti oggetti sono mascherati da 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(glmnet)
```

```
## Caricamento del pacchetto richiesto: Matrix

## Loaded glmnet 4.1-8
```

```
library(effects)
```

```
## Caricamento del pacchetto richiesto: carData
```

```
## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```

```
library(summarytools)
library(glmnetUtils)
```

```
##
## Caricamento pacchetto: 'glmnetUtils'
```

```
## I seguenti oggetti sono mascherati da 'package:glmnet':
##
##     cv.glmnet, glmnet
```

```
library(pROC)
```

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

```
##
## Caricamento pacchetto: 'pROC'
```

```
## I seguenti oggetti sono mascherati da 'package:stats':
##
##     cov, smooth, var
```

```
library(car)
```

```
##
## Caricamento pacchetto: 'car'
```

```
## Il seguente oggetto è mascherato da 'package:dplyr':
##
##     recode
```

```
## Il seguente oggetto è mascherato da 'package:boot':
##
##     logit
```

Preparation Dataset

```r
data <- read.csv("Dataset/airline_passenger_satisfaction.csv")
head(data)
```

```
##   ID Gender Age Customer.Type Type.of.Travel    Class Flight.Distance
## 1  1   Male  48    First-time       Business Business             821
## 2  2 Female  35     Returning       Business Business             821
## 3  3   Male  41     Returning       Business Business             853
## 4  4   Male  50     Returning       Business Business            1905
## 5  5 Female  49     Returning       Business Business            3470
## 6  6   Male  43     Returning       Business Business            3788
##   Departure.Delay Arrival.Delay Departure.and.Arrival.Time.Convenience
## 1               2             5                                      3
## 2              26            39                                      2
## 3               0             0                                      4
## 4               0             0                                      2
## 5               0             1                                      3
## 6               0             0                                      4
##   Ease.of.Online.Booking Check.in.Service Online.Boarding Gate.Location
## 1                      3                4               3             3
## 2                      2                3               5             2
## 3                      4                4               5             4
## 4                      2                3               4             2
## 5                      3                3               5             3
## 6                      4                3               5             4
##   On.board.Service Seat.Comfort Leg.Room.Service Cleanliness Food.and.Drink
## 1                3            5                2           5              5
## 2                5            4                5           5              3
## 3                3            5                3           5              5
## 4                5            5                5           4              4
## 5                3            4                4           5              4
## 6                4            4                4           3              3
##   In.flight.Service In.flight.Wifi.Service In.flight.Entertainment
## 1                 5                      3                       5
## 2                 5                      2                       5
## 3                 3                      4                       3
## 4                 5                      2                       5
## 5                 3                      3                       3
## 6                 4                      4                       4
##   Baggage.Handling            Satisfaction
## 1                5 Neutral or Dissatisfied
## 2                5               Satisfied
## 3                3               Satisfied
## 4                5               Satisfied
## 5                3               Satisfied
## 6                4               Satisfied
```

```r
attach(data)
dimension_data <- dim(data)
summary(data)
```

```
##        ID             Gender                Age         Customer.Type
##  Min.   :     1   Length:129880      Min.   : 7.00   Length:129880
```

```
##  1st Qu.: 32471   Class :character   1st Qu.:27.00   Class :character
##  Median : 64941   Mode  :character   Median :40.00   Mode  :character
##  Mean   : 64941                      Mean   :39.43
##  3rd Qu.: 97410                      3rd Qu.:51.00
##  Max.   :129880                      Max.   :85.00
##
##  Type.of.Travel        Class           Flight.Distance Departure.Delay
##  Length:129880      Length:129880      Min.   :  31    Min.   :   0.00
##  Class :character   Class :character   1st Qu.: 414    1st Qu.:   0.00
##  Mode  :character   Mode  :character   Median : 844    Median :   0.00
##                                        Mean   :1190    Mean   :  14.71
##                                        3rd Qu.:1744    3rd Qu.:  12.00
##                                        Max.   :4983    Max.   :1592.00
##
##  Arrival.Delay     Departure.and.Arrival.Time.Convenience
##  Min.   :   0.00   Min.   :0.000
##  1st Qu.:   0.00   1st Qu.:2.000
##  Median :   0.00   Median :3.000
##  Mean   :  15.09   Mean   :3.058
##  3rd Qu.:  13.00   3rd Qu.:4.000
##  Max.   :1584.00   Max.   :5.000
##  NA's   :393
##  Ease.of.Online.Booking Check.in.Service Online.Boarding Gate.Location
##  Min.   :0.000          Min.   :0.000    Min.   :0.000   Min.   :0.000
##  1st Qu.:2.000          1st Qu.:3.000    1st Qu.:2.000   1st Qu.:2.000
##  Median :3.000          Median :3.000    Median :3.000   Median :3.000
##  Mean   :2.757          Mean   :3.306    Mean   :3.253   Mean   :2.977
##  3rd Qu.:4.000          3rd Qu.:4.000    3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000          Max.   :5.000    Max.   :5.000   Max.   :5.000
##
##  On.board.Service  Seat.Comfort   Leg.Room.Service  Cleanliness
##  Min.   :0.000    Min.   :0.000   Min.   :0.000    Min.   :0.000
##  1st Qu.:2.000    1st Qu.:2.000   1st Qu.:2.000    1st Qu.:2.000
##  Median :4.000    Median :4.000   Median :4.000    Median :3.000
##  Mean   :3.383    Mean   :3.441   Mean   :3.351    Mean   :3.286
##  3rd Qu.:4.000    3rd Qu.:5.000   3rd Qu.:4.000    3rd Qu.:4.000
##  Max.   :5.000    Max.   :5.000   Max.   :5.000    Max.   :5.000
##
##  Food.and.Drink  In.flight.Service In.flight.Wifi.Service
##  Min.   :0.000   Min.   :0.000     Min.   :0.000
##  1st Qu.:2.000   1st Qu.:3.000     1st Qu.:2.000
##  Median :3.000   Median :4.000     Median :3.000
##  Mean   :3.205   Mean   :3.642     Mean   :2.729
##  3rd Qu.:4.000   3rd Qu.:5.000     3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000     Max.   :5.000
##
##  In.flight.Entertainment Baggage.Handling Satisfaction
##  Min.   :0.000           Min.   :1.000    Length:129880
##  1st Qu.:2.000           1st Qu.:3.000    Class :character
##  Median :4.000           Median :4.000    Mode  :character
##  Mean   :3.358           Mean   :3.632
##  3rd Qu.:4.000           3rd Qu.:5.000
##  Max.   :5.000           Max.   :5.000
##
```

```
print(dfSummary(data), method = 'render')   #'viewer' ,'browser', 'pander' and 'plain'
```

```
names(data)
```

```
##  [1] "ID"
##  [2] "Gender"
##  [3] "Age"
##  [4] "Customer.Type"
##  [5] "Type.of.Travel"
##  [6] "Class"
##  [7] "Flight.Distance"
##  [8] "Departure.Delay"
##  [9] "Arrival.Delay"
## [10] "Departure.and.Arrival.Time.Convenience"
## [11] "Ease.of.Online.Booking"
## [12] "Check.in.Service"
## [13] "Online.Boarding"
## [14] "Gate.Location"
## [15] "On.board.Service"
## [16] "Seat.Comfort"
## [17] "Leg.Room.Service"
## [18] "Cleanliness"
## [19] "Food.and.Drink"
## [20] "In.flight.Service"
## [21] "In.flight.Wifi.Service"
## [22] "In.flight.Entertainment"
## [23] "Baggage.Handling"
## [24] "Satisfaction"
```

Remove usefull column of the dataset

```
data<- data %>% select(-ID, -Gender)
head(data)
```

```
##   Age Customer.Type Type.of.Travel    Class Flight.Distance Departure.Delay
## 1  48    First-time       Business Business             821               2
## 2  35     Returning       Business Business             821              26
## 3  41     Returning       Business Business             853               0
## 4  50     Returning       Business Business            1905               0
## 5  49     Returning       Business Business            3470               0
## 6  43     Returning       Business Business            3788               0
##   Arrival.Delay Departure.and.Arrival.Time.Convenience Ease.of.Online.Booking
## 1             5                                     3                      3
## 2            39                                     2                      2
## 3             0                                     4                      4
## 4             0                                     2                      2
## 5             1                                     3                      3
## 6             0                                     4                      4
##   Check.in.Service Online.Boarding Gate.Location On.board.Service Seat.Comfort
## 1                4               3             3                3           5
## 2                3               5             2                5           4
## 3                4               5             4                3           5
```

```
## 4                3              4             2                 5            5
## 5                3              5             3                 3            4
## 6                3              5             4                 4            4
##   Leg.Room.Service Cleanliness Food.and.Drink In.flight.Service
## 1                2           5              5                 5
## 2                5           5              3                 5
## 3                3           5              5                 3
## 4                5           4              4                 5
## 5                4           5              4                 3
## 6                4           3              3                 4
##   In.flight.Wifi.Service In.flight.Entertainment Baggage.Handling
## 1                      3                       5                5
## 2                      2                       5                5
## 3                      4                       3                3
## 4                      2                       5                5
## 5                      3                       3                3
## 6                      4                       4                4
##              Satisfaction
## 1 Neutral or Dissatisfied
## 2               Satisfied
## 3               Satisfied
## 4               Satisfied
## 5               Satisfied
## 6               Satisfied
```

Remove observations with missing values

```
data <- na.omit(data)
```

Trasformo la satisfaction in 0 e 1 0 = neutral or dissatisfied 1 = satisfied

```
data$Satisfaction <- ifelse(data$Satisfaction %in% c("Neutral or Dissatisfied"),0, ifelse(data$Satisfac
data$Satisfaction <- as.factor(data$Satisfaction)
head(data)
```

```
##   Age Customer.Type Type.of.Travel    Class Flight.Distance Departure.Delay
## 1  48    First-time       Business Business             821               2
## 2  35     Returning       Business Business             821              26
## 3  41     Returning       Business Business             853               0
## 4  50     Returning       Business Business            1905               0
## 5  49     Returning       Business Business            3470               0
## 6  43     Returning       Business Business            3788               0
##   Arrival.Delay Departure.and.Arrival.Time.Convenience Ease.of.Online.Booking
## 1             5                                     3                      3
## 2            39                                     2                      2
## 3             0                                     4                      4
## 4             0                                     2                      2
## 5             1                                     3                      3
## 6             0                                     4                      4
##   Check.in.Service Online.Boarding Gate.Location On.board.Service Seat.Comfort
## 1                4               3             3                3           5
## 2                3               5             2                5           4
## 3                4               5             4                3           5
```

```
## 4                3               4             2                5            5
## 5                3               5             3                3            4
## 6                3               5             4                4            4
##   Leg.Room.Service Cleanliness Food.and.Drink In.flight.Service
## 1                2           5              5                 5
## 2                5           5              3                 5
## 3                3           5              5                 3
## 4                5           4              4                 5
## 5                4           5              4                 3
## 6                4           3              3                 4
##   In.flight.Wifi.Service In.flight.Entertainment Baggage.Handling Satisfaction
## 1                      3                       5                5            0
## 2                      2                       5                5            1
## 3                      4                       3                3            1
## 4                      2                       5                5            1
## 5                      3                       3                3            1
## 6                      4                       4                4            1
```

```
data <- data %>%
  mutate_if(is.character, as.factor)
summary(data)
```

```
##       Age           Customer.Type     Type.of.Travel           Class
##  Min.   : 7.00   First-time: 23714   Business:89445   Business    :61990
##  1st Qu.:27.00   Returning :105773   Personal:40042   Economy     :58117
##  Median :40.00                                        Economy Plus: 9380
##  Mean   :39.43
##  3rd Qu.:51.00
##  Max.   :85.00
##  Flight.Distance Departure.Delay   Arrival.Delay
##  Min.   :  31    Min.   :   0.00   Min.   :   0.00
##  1st Qu.: 414    1st Qu.:   0.00   1st Qu.:   0.00
##  Median : 844    Median :   0.00   Median :   0.00
##  Mean   :1190    Mean   :  14.64   Mean   :  15.09
##  3rd Qu.:1744    3rd Qu.:  12.00   3rd Qu.:  13.00
##  Max.   :4983    Max.   :1592.00   Max.   :1584.00
##  Departure.and.Arrival.Time.Convenience Ease.of.Online.Booking Check.in.Service
##  Min.   :0.000                          Min.   :0.000          Min.   :0.000
##  1st Qu.:2.000                          1st Qu.:2.000          1st Qu.:3.000
##  Median :3.000                          Median :3.000          Median :3.000
##  Mean   :3.057                          Mean   :2.757          Mean   :3.306
##  3rd Qu.:4.000                          3rd Qu.:4.000          3rd Qu.:4.000
##  Max.   :5.000                          Max.   :5.000          Max.   :5.000
##  Online.Boarding Gate.Location   On.board.Service  Seat.Comfort
##  Min.   :0.000   Min.   :0.000   Min.   :0.000    Min.   :0.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000    1st Qu.:2.000
##  Median :3.000   Median :3.000   Median :4.000    Median :4.000
##  Mean   :3.253   Mean   :2.977   Mean   :3.383    Mean   :3.442
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000    3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000    Max.   :5.000
##  Leg.Room.Service  Cleanliness    Food.and.Drink  In.flight.Service
##  Min.   :0.000    Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:2.000    1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000
##  Median :4.000    Median :3.000   Median :3.000   Median :4.000
```

```
##  Mean   :3.351    Mean   :3.286   Mean   :3.205   Mean   :3.642
##  3rd Qu.:4.000    3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :5.000    Max.   :5.000   Max.   :5.000   Max.   :5.000
##  In.flight.Wifi.Service In.flight.Entertainment Baggage.Handling Satisfaction
##  Min.   :0.000          Min.   :0.000           Min.   :1.000    0:73225
##  1st Qu.:2.000          1st Qu.:2.000           1st Qu.:3.000    1:56262
##  Median :3.000          Median :4.000           Median :4.000
##  Mean   :2.729          Mean   :3.358           Mean   :3.632
##  3rd Qu.:4.000          3rd Qu.:4.000           3rd Qu.:5.000
##  Max.   :5.000          Max.   :5.000           Max.   :5.000
```

```
print(dfSummary(data), method = 'render')
```

Dividio il dataset in train e test set

```
partition_data <- function(data, target_var, train_ratio = 0.8, seed = NULL) {
  if (!is.null(seed)) {
    set.seed(seed)
  }

  nrow <- nrow(data)
  sample <- sample(c(TRUE, FALSE), nrow, replace = TRUE, prob = c(train_ratio, 1 - train_ratio))

  train <- data[sample, ]
  test <- data[!sample, ]

  trainY <- train[, target_var]
  testY <- test[, target_var]

  trainX <- train[, -which(names(train) == target_var)]
  testX <- test[, -which(names(test) == target_var)]

  return(list(trainX = trainX, trainY = trainY, testX = testX, testY = testY,train = train, test = test
}



split <- partition_data(data, target_var = "Satisfaction", train_ratio = 0.8, seed = 123)

dim(split$trainX)
```

```
## [1] 103612     21
```

```
dim(split$testX)
```

```
## [1] 25875    21
```

```
head(split$trainX)
```

```
##   Age Customer.Type Type.of.Travel    Class Flight.Distance Departure.Delay
## 1  48    First-time       Business Business             821               2
```

```
## 2  35     Returning       Business Business             821              26
## 3  41     Returning       Business Business             853               0
## 6  43     Returning       Business Business            3788               0
## 7  43     Returning       Business Business            1963               0
## 9  50     Returning       Business Business            2607               0
##   Arrival.Delay Departure.and.Arrival.Time.Convenience Ease.of.Online.Booking
## 1             5                                     3                      3
## 2            39                                     2                      2
## 3             0                                     4                      4
## 6             0                                     4                      4
## 7             0                                     3                      3
## 9             0                                     1                      1
##   Check.in.Service Online.Boarding Gate.Location On.board.Service Seat.Comfort
## 1                4               3             3                3            5
## 2                3               5             2                5            4
## 3                4               5             4                3            5
## 6                3               5             4                4            4
## 7                4               4             3                5            5
## 9                3               2             1                4            3
##   Leg.Room.Service Cleanliness Food.and.Drink In.flight.Service
## 1                2           5              5                 5
## 2                5           5              3                 5
## 3                3           5              5                 3
## 6                4           3              3                 4
## 7                5           4              5                 5
## 9                4           3              3                 4
##   In.flight.Wifi.Service In.flight.Entertainment Baggage.Handling
## 1                      3                       5                5
## 2                      2                       5                5
## 3                      4                       3                3
## 6                      4                       4                4
## 7                      3                       5                5
## 9                      4                       4                4
```

```
head(split$testX)
```

```
##    Age Customer.Type Type.of.Travel    Class Flight.Distance Departure.Delay
## 4   50     Returning       Business Business            1905               0
## 5   49     Returning       Business Business            3470               0
## 8   60     Returning       Business Business             853               0
## 11  28    First-time       Business Business             821               0
## 16  70     Returning       Personal  Economy             821               0
## 20  42     Returning       Personal  Economy             821               4
##    Arrival.Delay Departure.and.Arrival.Time.Convenience Ease.of.Online.Booking
## 4              0                                     2                      2
## 5              1                                     3                      3
## 8              3                                     3                      4
## 11             5                                     1                      1
## 16             0                                     5                      2
## 20             0                                     3                      3
##    Check.in.Service Online.Boarding Gate.Location On.board.Service Seat.Comfort
## 4                 3               4             2                5            5
## 5                 3               5             3                3            4
## 8                 3               4             4                3            4
```

```
## 11                3               1             3                 2            2
## 16                4               2             1                 4            5
## 20                3               3             4                 1            4
##    Leg.Room.Service Cleanliness Food.and.Drink In.flight.Service
## 4                 5           4              4                 5
## 5                 4           5              4                 3
## 8                 4           4              4                 3
## 11                5           2              2                 4
## 16                2           5              5                 5
## 20                3           3              1                 1
##    In.flight.Wifi.Service In.flight.Entertainment Baggage.Handling
## 4                       2                       5                5
## 5                       3                       3                3
## 8                       4                       3                3
## 11                      1                       2                3
## 16                      2                       5                4
## 20                      3                       1                1
```

```
head(split$trainY)
```

```
## [1] 0 1 1 1 1 0
## Levels: 0 1
```

```
head(split$testY)
```

```
## [1] 1 1 1 0 0 0
## Levels: 0 1
```

```
train <- split$train
test <- split$test
testX <- split$testX
testy <- split$testY
```

Implementation of models

Cosa voglio fare eccc....

Logistic Regression

```
model_with_all_predictors <- glm(Satisfaction ~ ., data = split$train, family = binomial)
summary(model_with_all_predictors)
```

```
##
## Call:
## glm(formula = Satisfaction ~ ., family = binomial, data = split$train)
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                               -7.812e+00  7.823e-02 -99.852  < 2e-16
## Age                                       -8.375e-03  7.113e-04 -11.773  < 2e-16
## Customer.TypeReturning                     2.037e+00  2.986e-02  68.235  < 2e-16
## Type.of.TravelPersonal                    -2.720e+00  3.142e-02 -86.564  < 2e-16
## ClassEconomy                              -7.125e-01  2.563e-02 -27.797  < 2e-16
## ClassEconomy Plus                         -8.091e-01  4.137e-02 -19.557  < 2e-16
## Flight.Distance                           -1.001e-05  1.127e-05  -0.888  0.37457
## Departure.Delay                            3.977e-03  9.849e-04   4.038 5.39e-05
## Arrival.Delay                             -8.848e-03  9.706e-04  -9.116  < 2e-16
## Departure.and.Arrival.Time.Convenience -1.371e-01  8.183e-03 -16.748  < 2e-16
## Ease.of.Online.Booking                    -1.478e-01  1.129e-02 -13.087  < 2e-16
## Check.in.Service                           3.289e-01  8.554e-03  38.448  < 2e-16
## Online.Boarding                            6.095e-01  1.023e-02  59.575  < 2e-16
## Gate.Location                              2.536e-02  9.185e-03   2.761  0.00576
## On.board.Service                           2.922e-01  1.020e-02  28.642  < 2e-16
## Seat.Comfort                               6.435e-02  1.119e-02   5.749 8.97e-09
## Leg.Room.Service                           2.489e-01  8.521e-03  29.214  < 2e-16
## Cleanliness                                2.251e-01  1.208e-02  18.632  < 2e-16
## Food.and.Drink                            -2.937e-02  1.070e-02  -2.745  0.00606
## In.flight.Service                          1.156e-01  1.206e-02   9.590  < 2e-16
## In.flight.Wifi.Service                     4.038e-01  1.144e-02  35.286  < 2e-16
## In.flight.Entertainment                    6.127e-02  1.422e-02   4.308 1.65e-05
## Baggage.Handling                           1.491e-01  1.142e-02  13.054  < 2e-16
##
## (Intercept)                               ***
## Age                                       ***
## Customer.TypeReturning                    ***
## Type.of.TravelPersonal                    ***
## ClassEconomy                              ***
## ClassEconomy Plus                         ***
## Flight.Distance
## Departure.Delay                           ***
## Arrival.Delay                             ***
## Departure.and.Arrival.Time.Convenience ***
## Ease.of.Online.Booking                    ***
## Check.in.Service                          ***
## Online.Boarding                           ***
## Gate.Location                             **
## On.board.Service                          ***
## Seat.Comfort                              ***
## Leg.Room.Service                          ***
## Cleanliness                               ***
## Food.and.Drink                            **
## In.flight.Service                         ***
## In.flight.Wifi.Service                    ***
## In.flight.Entertainment                   ***
## Baggage.Handling                          ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69367  on 103589  degrees of freedom
```

```
## AIC: 69413
##
## Number of Fisher Scoring iterations: 5
```

```
new_model <- glm(Satisfaction ~ Age +Customer.Type + Type.of.Travel + Class + Flight.Distance + Departu
+ Departure.and.Arrival.Time.Convenience + Ease.of.Online.Booking + Check.in.Service + Online.Boarding
 Food.and.Drink + In.flight.Service + In.flight.Wifi.Service + In.flight.Entertainment + Baggage.Handli
 , data = split$train, family = binomial)
summary(new_model)
```

```
##
## Call:
## glm(formula = Satisfaction ~ Age + Customer.Type + Type.of.Travel +
##     Class + Flight.Distance + Departure.Delay + Arrival.Delay +
##     Departure.and.Arrival.Time.Convenience + Ease.of.Online.Booking +
##     Check.in.Service + Online.Boarding + Gate.Location + On.board.Service +
##     Seat.Comfort + Leg.Room.Service + Cleanliness + Food.and.Drink +
##     In.flight.Service + In.flight.Wifi.Service + In.flight.Entertainment +
##     Baggage.Handling, family = binomial, data = split$train)
##
## Coefficients:
##                                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                               -7.812e+00  7.823e-02 -99.852  < 2e-16
## Age                                       -8.375e-03  7.113e-04 -11.773  < 2e-16
## Customer.TypeReturning                     2.037e+00  2.986e-02  68.235  < 2e-16
## Type.of.TravelPersonal                    -2.720e+00  3.142e-02 -86.564  < 2e-16
## ClassEconomy                              -7.125e-01  2.563e-02 -27.797  < 2e-16
## ClassEconomy Plus                         -8.091e-01  4.137e-02 -19.557  < 2e-16
## Flight.Distance                           -1.001e-05  1.127e-05  -0.888  0.37457
## Departure.Delay                            3.977e-03  9.849e-04   4.038 5.39e-05
## Arrival.Delay                             -8.848e-03  9.706e-04  -9.116  < 2e-16
## Departure.and.Arrival.Time.Convenience    -1.371e-01  8.183e-03 -16.748  < 2e-16
## Ease.of.Online.Booking                    -1.478e-01  1.129e-02 -13.087  < 2e-16
## Check.in.Service                           3.289e-01  8.554e-03  38.448  < 2e-16
## Online.Boarding                            6.095e-01  1.023e-02  59.575  < 2e-16
## Gate.Location                              2.536e-02  9.185e-03   2.761  0.00576
## On.board.Service                           2.922e-01  1.020e-02  28.642  < 2e-16
## Seat.Comfort                               6.435e-02  1.119e-02   5.749 8.97e-09
## Leg.Room.Service                           2.489e-01  8.521e-03  29.214  < 2e-16
## Cleanliness                                2.251e-01  1.208e-02  18.632  < 2e-16
## Food.and.Drink                            -2.937e-02  1.070e-02  -2.745  0.00606
## In.flight.Service                          1.156e-01  1.206e-02   9.590  < 2e-16
## In.flight.Wifi.Service                     4.038e-01  1.144e-02  35.286  < 2e-16
## In.flight.Entertainment                    6.127e-02  1.422e-02   4.308 1.65e-05
## Baggage.Handling                           1.491e-01  1.142e-02  13.054  < 2e-16
##
## (Intercept)                               ***
## Age                                       ***
## Customer.TypeReturning                    ***
## Type.of.TravelPersonal                    ***
## ClassEconomy                              ***
## ClassEconomy Plus                         ***
## Flight.Distance
## Departure.Delay                           ***
```

```
## Arrival.Delay                           ***
## Departure.and.Arrival.Time.Convenience ***
## Ease.of.Online.Booking                  ***
## Check.in.Service                        ***
## Online.Boarding                         ***
## Gate.Location                           **
## On.board.Service                        ***
## Seat.Comfort                            ***
## Leg.Room.Service                        ***
## Cleanliness                             ***
## Food.and.Drink                          **
## In.flight.Service                       ***
## In.flight.Wifi.Service                  ***
## In.flight.Entertainment                 ***
## Baggage.Handling                        ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69367  on 103589  degrees of freedom
## AIC: 69413
##
## Number of Fisher Scoring iterations: 5
```

Most of the predictors seem to be useful for the calculation of the final result, in fact all predictors have three asterisks. We now go on to calculate the collinearity between the different predictors:

```
check_collinearity <- vif(new_model)
check_collinearity
```

```
##                                        GVIF Df GVIF^(1/(2*Df))
## Age                                1.184136  1        1.088180
## Customer.Type                      1.610252  1        1.268957
## Type.of.Travel                     1.854266  1        1.361714
## Class                              1.669581  2        1.136716
## Flight.Distance                    1.353564  1        1.163428
## Departure.Delay                   13.831235  1        3.719037
## Arrival.Delay                     13.861757  1        3.723138
## Departure.and.Arrival.Time.Convenience  1.715222  1        1.309665
## Ease.of.Online.Booking             2.587823  1        1.608671
## Check.in.Service                   1.206622  1        1.098464
## Online.Boarding                    1.480741  1        1.216857
## Gate.Location                      1.527318  1        1.235847
## On.board.Service                   1.646149  1        1.283023
## Seat.Comfort                       2.037790  1        1.427512
## Leg.Room.Service                   1.217567  1        1.103434
## Cleanliness                        2.452767  1        1.566131
## Food.and.Drink                     2.016920  1        1.420183
## In.flight.Service                  2.023745  1        1.422584
## In.flight.Wifi.Service             2.215943  1        1.488604
## In.flight.Entertainment            3.243813  1        1.801059
## Baggage.Handling                   1.823892  1        1.350515
```

The image shows the results of a function vif() applied to a regression model. The results include the columns GVIF, Df, and GVIF^(1/(2*Df)).

- GVIF (Generalized Variance Inflation Factor): Extension of VIF that takes into account the number of levels of a categorical variable. Larger values indicate greater collinearity.
- Df (Degrees of Freedom): Degrees of Freedom associated with the variable. For continuous variables it is usually 1.
- GVIF^(1/(2*Df)): Correction of GVIF to facilitate interpretation, especially when having categorical variables with multiple levels.

Analysis: + Departure.Delay and Arrival.Delay: Both have very high GVIF values (13.47 and 13.50), indicating strong collinearity. This could be due to the fact that these two predictors are strongly correlated with each other.

- Ease.of.Online.Booking and Cleanliness: They have moderate GVIF values (2.58 and 2.46), indicating moderate collinearity.
- In.flight.Service and In.flight.Wifi.Service: Moderate GVIF values (2.00 and 2.22).

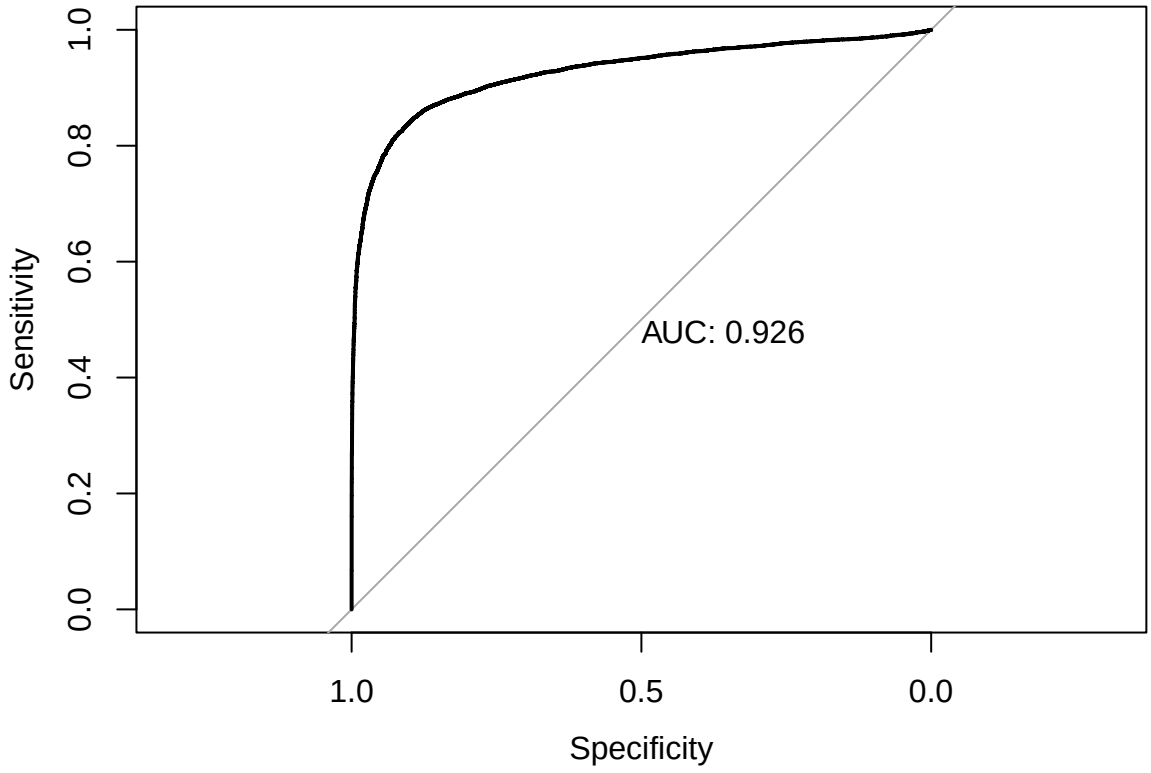
Below i run a test using the test set of mine dataset to see the capacity of the model:

```
new_model.probs <- predict(new_model, testX, type = "response")

new_model_roc <- roc(test$Satisfaction ~ new_model.probs, plot=TRUE, print.auc=TRUE)
```

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

## Setting direction: controls < cases
```



The image represents a ROC (Receiver Operating Characteristic) curve used for evaluating the performance of a binary classification model. An ROC curve is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:\ True Positive Rate False Positive Rate

Axes:

X-axis (Specificity): This typically ranges from 0 to 1, representing the True Negative Rate (TNR). However, the scale here seems reversed, going from 1.5 to -0.5, which is unconventional. Y-axis (Sensitivity): This ranges from 0 to 1, representing the True Positive Rate (TPR). ROC Curve:

The curve plots Sensitivity against 1 - Specificity at various threshold settings. AUC (Area Under the Curve):

The AUC value of 0.927 indicates a high performance of the model, with an excellent ability to distinguish between the positive and negative classes. The conventional interpretation of AUC values is as follows:

0.5: No discrimination (random classifier) 0.7 - 0.8: Acceptable 0.8 - 0.9: Excellent 0.9: Outstanding

Given the AUC of 0.926, the model in question is performing outstandingly in distinguishing between the two classes.

```
coords(new_model_roc, x=0.5, ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold       0.5   0.9086611    0.831132 0.8748599 13261 9376 1905 1333
##                 npv       ppv       fdr        fpr      tpr       tnr      fnr
## threshold 0.8743901 0.8755253 0.1244747 0.09133891 0.831132 0.9086611 0.168868
##           1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv precision
## threshold    0.09133891      0.168868  0.1251401 0.1256099 0.1244747 0.8755253
##             recall   youden closest.topleft
## threshold 0.831132 1.739793       0.0368592
```

```
coords(new_model_roc, x="best", ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold 0.4698187   0.8972866   0.8435422 0.8738551 13095 9516 1765 1499
##                 npv       ppv       fdr       fpr       tpr       tnr       fnr
## threshold 0.8812248 0.8639128 0.1360872 0.1027134 0.8435422 0.8972866 0.1564578
##           1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv precision
## threshold     0.1027134     0.1564578  0.1261449 0.1187752 0.1360872 0.8639128
##              recall   youden closest.topleft
## threshold 0.8435422 1.740829      0.03502908
```

Stepwise Regression models

In this section I am going to implement the stepwise regression model using three different modes: forward, backward and both.

```
null_model <- glm(Satisfaction ~ 1, data = train, family = binomial)
full_model <- glm(Satisfaction ~ ., data = train, family = binomial)
summary(full_model)
```

```
##
## Call:
## glm(formula = Satisfaction ~ ., family = binomial, data = train)
##
## Coefficients:
##                                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                               -7.812e+00  7.823e-02 -99.852  < 2e-16
## Age                                       -8.375e-03  7.113e-04 -11.773  < 2e-16
## Customer.TypeReturning                     2.037e+00  2.986e-02  68.235  < 2e-16
## Type.of.TravelPersonal                    -2.720e+00  3.142e-02 -86.564  < 2e-16
## ClassEconomy                              -7.125e-01  2.563e-02 -27.797  < 2e-16
## ClassEconomy Plus                         -8.091e-01  4.137e-02 -19.557  < 2e-16
## Flight.Distance                           -1.001e-05  1.127e-05  -0.888  0.37457
## Departure.Delay                            3.977e-03  9.849e-04   4.038 5.39e-05
## Arrival.Delay                             -8.848e-03  9.706e-04  -9.116  < 2e-16
## Departure.and.Arrival.Time.Convenience    -1.371e-01  8.183e-03 -16.748  < 2e-16
## Ease.of.Online.Booking                    -1.478e-01  1.129e-02 -13.087  < 2e-16
## Check.in.Service                           3.289e-01  8.554e-03  38.448  < 2e-16
## Online.Boarding                            6.095e-01  1.023e-02  59.575  < 2e-16
## Gate.Location                              2.536e-02  9.185e-03   2.761  0.00576
## On.board.Service                           2.922e-01  1.020e-02  28.642  < 2e-16
## Seat.Comfort                               6.435e-02  1.119e-02   5.749 8.97e-09
## Leg.Room.Service                           2.489e-01  8.521e-03  29.214  < 2e-16
## Cleanliness                                2.251e-01  1.208e-02  18.632  < 2e-16
## Food.and.Drink                            -2.937e-02  1.070e-02  -2.745  0.00606
## In.flight.Service                          1.156e-01  1.206e-02   9.590  < 2e-16
## In.flight.Wifi.Service                     4.038e-01  1.144e-02  35.286  < 2e-16
## In.flight.Entertainment                    6.127e-02  1.422e-02   4.308 1.65e-05
## Baggage.Handling                           1.491e-01  1.142e-02  13.054  < 2e-16
##
## (Intercept)                               ***
## Age                                       ***
```

```
## Customer.TypeReturning                  ***
## Type.of.TravelPersonal                  ***
## ClassEconomy                            ***
## ClassEconomy Plus                       ***
## Flight.Distance
## Departure.Delay                         ***
## Arrival.Delay                           ***
## Departure.and.Arrival.Time.Convenience ***
## Ease.of.Online.Booking                  ***
## Check.in.Service                        ***
## Online.Boarding                         ***
## Gate.Location                           **
## On.board.Service                        ***
## Seat.Comfort                            ***
## Leg.Room.Service                        ***
## Cleanliness                             ***
## Food.and.Drink                          **
## In.flight.Service                       ***
## In.flight.Wifi.Service                  ***
## In.flight.Entertainment                 ***
## Baggage.Handling                        ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69367  on 103589  degrees of freedom
## AIC: 69413
##
## Number of Fisher Scoring iterations: 5
```

```
stepwise_model_forward <- step(null_model, scope = list(lower = null_model, upper = full_model), direct
```

Forward

```
## Start:  AIC=141835.2
## Satisfaction ~ 1
##
##                                      Df Deviance    AIC
## + Online.Boarding                     1   111866 111870
## + Class                               2   114373 114379
## + Type.of.Travel                      1   118175 118179
## + In.flight.Entertainment             1   124047 124051
## + Seat.Comfort                        1   128375 128379
## + On.board.Service                    1   130620 130624
## + Leg.Room.Service                    1   131227 131231
## + Cleanliness                         1   131668 131672
## + Flight.Distance                     1   132314 132318
## + In.flight.Wifi.Service              1   133215 133219
## + Baggage.Handling                    1   135111 135115
```

```
## + In.flight.Service                          1   135421 135425
## + Check.in.Service                           1   135817 135821
## + Food.and.Drink                             1   137077 137081
## + Customer.Type                              1   138001 138005
## + Ease.of.Online.Booking                     1   138830 138834
## + Age                                        1   139940 139944
## + Arrival.Delay                              1   141492 141496
## + Departure.and.Arrival.Time.Convenience     1   141514 141518
## + Departure.Delay                            1   141574 141578
## <none>                                           141833 141835
## + Gate.Location                              1   141832 141836
##
## Step:  AIC=111870.2
## Satisfaction ~ Online.Boarding
##
##                                           Df Deviance    AIC
## + Type.of.Travel                           1    93786  93792
## + Class                                    2    95650  95658
## + In.flight.Entertainment                  1   102223 102229
## + Leg.Room.Service                         1   103128 103134
## + On.board.Service                         1   103787 103793
## + Baggage.Handling                         1   106195 106201
## + In.flight.Service                        1   106234 106240
## + Flight.Distance                          1   106775 106781
## + Seat.Comfort                             1   109048 109054
## + Cleanliness                              1   109148 109154
## + Check.in.Service                         1   109583 109589
## + Customer.Type                            1   110621 110627
## + Departure.and.Arrival.Time.Convenience   1   110710 110716
## + Food.and.Drink                           1   110717 110723
## + In.flight.Wifi.Service                   1   111151 111157
## + Arrival.Delay                            1   111591 111597
## + Departure.Delay                          1   111653 111659
## + Age                                      1   111742 111748
## + Ease.of.Online.Booking                   1   111826 111832
## <none>                                           111866 111870
## + Gate.Location                            1   111865 111871
##
## Step:  AIC=93792.47
## Satisfaction ~ Online.Boarding + Type.of.Travel
##
##                                           Df Deviance   AIC
## + On.board.Service                         1    85557 85565
## + In.flight.Entertainment                  1    86068 86076
## + Leg.Room.Service                         1    87251 87259
## + In.flight.Service                        1    87685 87693
## + Baggage.Handling                         1    87814 87822
## + Customer.Type                            1    87907 87915
## + Class                                    2    89100 89110
## + Check.in.Service                         1    90206 90214
## + Cleanliness                              1    91223 91231
## + Seat.Comfort                             1    91538 91546
## + Flight.Distance                          1    92095 92103
## + Food.and.Drink                           1    92941 92949
```

```
## + In.flight.Wifi.Service                   1    93019 93027
## + Arrival.Delay                            1    93416 93424
## + Departure.Delay                          1    93488 93496
## + Age                                      1    93559 93567
## + Ease.of.Online.Booking                   1    93726 93734
## + Gate.Location                            1    93762 93770
## + Departure.and.Arrival.Time.Convenience   1    93784 93792
## <none>                                          93786 93792
##
## Step:  AIC=85565.23
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service
##
##                                           Df Deviance   AIC
## + Customer.Type                            1    79300 79310
## + In.flight.Entertainment                  1    82603 82613
## + Class                                    2    82777 82789
## + Leg.Room.Service                         1    82939 82949
## + Cleanliness                              1    83510 83520
## + Check.in.Service                         1    83726 83736
## + Seat.Comfort                             1    83769 83779
## + Flight.Distance                          1    84271 84281
## + Baggage.Handling                         1    84335 84345
## + In.flight.Service                        1    84466 84476
## + Food.and.Drink                           1    84690 84700
## + In.flight.Wifi.Service                   1    85127 85137
## + Arrival.Delay                            1    85285 85295
## + Departure.Delay                          1    85338 85348
## + Age                                      1    85424 85434
## + Ease.of.Online.Booking                   1    85497 85507
## + Departure.and.Arrival.Time.Convenience   1    85545 85555
## + Gate.Location                            1    85548 85558
## <none>                                          85557 85565
##
## Step:  AIC=79310.18
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type
##
##                                           Df Deviance   AIC
## + Check.in.Service                         1    76801 76813
## + Leg.Room.Service                         1    76952 76964
## + In.flight.Entertainment                  1    77180 77192
## + Cleanliness                              1    77352 77364
## + Baggage.Handling                         1    77360 77372
## + In.flight.Service                        1    77514 77526
## + Class                                    2    78158 78172
## + Seat.Comfort                             1    78242 78254
## + In.flight.Wifi.Service                   1    78360 78372
## + Food.and.Drink                           1    78549 78561
## + Arrival.Delay                            1    78981 78993
## + Age                                      1    79022 79034
## + Departure.Delay                          1    79043 79055
## + Departure.and.Arrival.Time.Convenience   1    79170 79182
## + Flight.Distance                          1    79180 79192
## + Gate.Location                            1    79283 79295
```

```
## + Ease.of.Online.Booking                   1    79293 79305
## <none>                                         79300 79310
##
## Step:  AIC=76812.67
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service
##
##                                           Df Deviance   AIC
## + In.flight.Entertainment                  1    74561 74575
## + Leg.Room.Service                         1    74717 74731
## + Cleanliness                              1    75254 75268
## + Baggage.Handling                         1    75289 75303
## + In.flight.Service                        1    75421 75435
## + In.flight.Wifi.Service                   1    75607 75621
## + Seat.Comfort                             1    76049 76063
## + Class                                    2    76073 76089
## + Food.and.Drink                           1    76124 76138
## + Arrival.Delay                            1    76429 76443
## + Departure.Delay                          1    76497 76511
## + Age                                      1    76515 76529
## + Departure.and.Arrival.Time.Convenience   1    76642 76656
## + Flight.Distance                          1    76736 76750
## + Gate.Location                            1    76791 76805
## <none>                                         76801 76813
## + Ease.of.Online.Booking                   1    76801 76815
##
## Step:  AIC=74574.71
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment
##
##                                           Df Deviance   AIC
## + Leg.Room.Service                         1    73092 73108
## + Class                                    2    73658 73676
## + In.flight.Wifi.Service                   1    73762 73778
## + Baggage.Handling                         1    73765 73781
## + In.flight.Service                        1    73941 73957
## + Arrival.Delay                            1    74223 74239
## + Age                                      1    74265 74281
## + Departure.Delay                          1    74289 74305
## + Departure.and.Arrival.Time.Convenience   1    74427 74443
## + Cleanliness                              1    74441 74457
## + Flight.Distance                          1    74477 74493
## + Gate.Location                            1    74547 74563
## + Food.and.Drink                           1    74550 74566
## + Ease.of.Online.Booking                   1    74550 74566
## + Seat.Comfort                             1    74558 74574
## <none>                                         74561 74575
##
## Step:  AIC=73107.55
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service
##
##                                           Df Deviance   AIC
```

```
## + Class                                     2    72279 72299
## + In.flight.Wifi.Service                    1    72560 72578
## + Baggage.Handling                          1    72661 72679
## + Arrival.Delay                             1    72700 72718
## + Departure.Delay                           1    72770 72788
## + In.flight.Service                         1    72774 72792
## + Age                                       1    72846 72864
## + Cleanliness                               1    72872 72890
## + Departure.and.Arrival.Time.Convenience    1    72931 72949
## + Flight.Distance                           1    73028 73046
## + Seat.Comfort                              1    73063 73081
## + Gate.Location                             1    73079 73097
## + Ease.of.Online.Booking                    1    73089 73107
## <none>                                           73092 73108
## + Food.and.Drink                            1    73090 73108
##
## Step:  AIC=72299.12
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class
##
##                                           Df Deviance   AIC
## + In.flight.Wifi.Service                   1    71463 71485
## + Arrival.Delay                            1    71907 71929
## + Baggage.Handling                         1    71940 71962
## + Departure.Delay                          1    71972 71994
## + Age                                      1    72021 72043
## + In.flight.Service                        1    72047 72069
## + Cleanliness                              1    72078 72100
## + Departure.and.Arrival.Time.Convenience   1    72123 72145
## + Gate.Location                            1    72268 72290
## + Seat.Comfort                             1    72271 72293
## + Flight.Distance                          1    72273 72295
## <none>                                          72279 72299
## + Food.and.Drink                           1    72279 72301
## + Ease.of.Online.Booking                   1    72279 72301
##
## Step:  AIC=71485.39
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service
##
##                                           Df Deviance   AIC
## + Departure.and.Arrival.Time.Convenience   1    70895 70919
## + Ease.of.Online.Booking                   1    70939 70963
## + Arrival.Delay                            1    71118 71142
## + Cleanliness                              1    71153 71177
## + Departure.Delay                          1    71179 71203
## + Baggage.Handling                         1    71219 71243
## + Age                                      1    71269 71293
## + In.flight.Service                        1    71295 71319
## + Gate.Location                            1    71301 71325
## + Seat.Comfort                             1    71408 71432
## <none>                                          71463 71485
```

```
## + Food.and.Drink                          1    71461 71485
## + Flight.Distance                         1    71462 71486
##
## Step:  AIC=70919.49
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience
##
##                           Df Deviance   AIC
## + Arrival.Delay            1    70553 70579
## + Cleanliness              1    70565 70591
## + Departure.Delay          1    70616 70642
## + Baggage.Handling         1    70645 70671
## + Ease.of.Online.Booking   1    70691 70717
## + In.flight.Service        1    70711 70737
## + Age                      1    70722 70748
## + Seat.Comfort             1    70830 70856
## + Gate.Location            1    70891 70917
## + Food.and.Drink           1    70893 70919
## <none>                          70895 70919
## + Flight.Distance          1    70895 70921
##
## Step:  AIC=70579.22
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay
##
##                           Df Deviance   AIC
## + Cleanliness              1    70188 70216
## + Baggage.Handling         1    70290 70318
## + Ease.of.Online.Booking   1    70338 70366
## + Age                      1    70372 70400
## + In.flight.Service        1    70395 70423
## + Seat.Comfort             1    70483 70511
## + Departure.Delay          1    70535 70563
## + Gate.Location            1    70549 70577
## + Food.and.Drink           1    70551 70579
## <none>                          70553 70579
## + Flight.Distance          1    70553 70581
##
## Step:  AIC=70216.24
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness
##
##                           Df Deviance   AIC
## + Baggage.Handling         1    69821 69851
## + In.flight.Service        1    69913 69943
## + Ease.of.Online.Booking   1    70004 70034
## + Age                      1    70029 70059
## + Food.and.Drink           1    70168 70198
## + Departure.Delay          1    70171 70201
```

```
## + Seat.Comfort            1    70183 70213
## + Gate.Location           1    70184 70214
## <none>                         70188 70216
## + Flight.Distance         1    70188 70218
##
## Step:  AIC=69850.55
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling
##
##                           Df Deviance   AIC
## + Ease.of.Online.Booking  1    69658 69690
## + Age                     1    69697 69729
## + In.flight.Service       1    69715 69747
## + Seat.Comfort            1    69802 69834
## + Departure.Delay         1    69804 69836
## + Food.and.Drink          1    69815 69847
## + Gate.Location           1    69818 69850
## <none>                         69821 69851
## + Flight.Distance         1    69820 69852
##
## Step:  AIC=69689.77
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking
##
##                     Df Deviance   AIC
## + Age                1    69519 69553
## + In.flight.Service  1    69555 69589
## + Departure.Delay    1    69642 69676
## + Seat.Comfort       1    69645 69679
## + Gate.Location      1    69651 69685
## + Food.and.Drink     1    69652 69686
## <none>                    69658 69690
## + Flight.Distance    1    69658 69692
##
## Step:  AIC=69552.85
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age
##
##                     Df Deviance   AIC
## + In.flight.Service  1    69431 69467
## + Seat.Comfort       1    69497 69533
## + Departure.Delay    1    69502 69538
## + Gate.Location      1    69510 69546
## + Food.and.Drink     1    69511 69547
## <none>                    69519 69553
## + Flight.Distance    1    69518 69554
##
```

```
## Step:  AIC=69466.61
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service
##
##                   Df Deviance   AIC
## + Seat.Comfort     1    69400 69438
## + Departure.Delay  1    69414 69452
## + Gate.Location    1    69421 69459
## + Food.and.Drink   1    69427 69465
## <none>                  69431 69467
## + Flight.Distance  1    69430 69468
##
## Step:  AIC=69437.6
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service + Seat.Comfort
##
##                   Df Deviance   AIC
## + Departure.Delay  1    69383 69423
## + Gate.Location    1    69392 69432
## + Food.and.Drink   1    69392 69432
## <none>                  69400 69438
## + Flight.Distance  1    69399 69439
##
## Step:  AIC=69423.42
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service + Seat.Comfort + Departure.Delay
##
##                   Df Deviance   AIC
## + Gate.Location    1    69375 69417
## + Food.and.Drink   1    69376 69418
## <none>                  69383 69423
## + Flight.Distance  1    69383 69425
##
## Step:  AIC=69417.38
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service + Seat.Comfort + Departure.Delay +
##     Gate.Location
##
##                   Df Deviance   AIC
## + Food.and.Drink   1    69368 69412
## <none>                  69375 69417
## + Flight.Distance  1    69375 69419
```

```
##
## Step:  AIC=69411.93
## Satisfaction ~ Online.Boarding + Type.of.Travel + On.board.Service +
##     Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service + Seat.Comfort + Departure.Delay +
##     Gate.Location + Food.and.Drink
##
##                   Df Deviance   AIC
## <none>                 69368 69412
## + Flight.Distance  1    69367 69413
```

```
summary(stepwise_model_forward)
```

```
##
## Call:
## glm(formula = Satisfaction ~ Online.Boarding + Type.of.Travel +
##     On.board.Service + Customer.Type + Check.in.Service + In.flight.Entertainment +
##     Leg.Room.Service + Class + In.flight.Wifi.Service + Departure.and.Arrival.Time.Convenience +
##     Arrival.Delay + Cleanliness + Baggage.Handling + Ease.of.Online.Booking +
##     Age + In.flight.Service + Seat.Comfort + Departure.Delay +
##     Gate.Location + Food.and.Drink, family = binomial, data = train)
##
## Coefficients:
##                                      Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                        -7.8232629  0.0771300 -101.430  < 2e-16
## Online.Boarding                     0.6091938  0.0102259   59.573  < 2e-16
## Type.of.TravelPersonal             -2.7166692  0.0312090  -87.048  < 2e-16
## On.board.Service                    0.2920703  0.0101986   28.638  < 2e-16
## Customer.TypeReturning              2.0303950  0.0288025   70.494  < 2e-16
## Check.in.Service                    0.3288308  0.0085534   38.444  < 2e-16
## In.flight.Entertainment             0.0612187  0.0142220    4.305 1.67e-05
## Leg.Room.Service                    0.2486636  0.0085151   29.203  < 2e-16
## ClassEconomy                       -0.7054170  0.0243630  -28.954  < 2e-16
## ClassEconomy Plus                  -0.8006458  0.0402537  -19.890  < 2e-16
## In.flight.Wifi.Service              0.4043205  0.0114272   35.382  < 2e-16
## Departure.and.Arrival.Time.Convenience -0.1371571  0.0081820  -16.763  < 2e-16
## Arrival.Delay                      -0.0088439  0.0009705   -9.113  < 2e-16
## Cleanliness                         0.2250334  0.0120806   18.628  < 2e-16
## Baggage.Handling                    0.1492217  0.0114204   13.066  < 2e-16
## Ease.of.Online.Booking             -0.1479513  0.0112912  -13.103  < 2e-16
## Age                                -0.0083330  0.0007098  -11.741  < 2e-16
## In.flight.Service                   0.1157939  0.0120512    9.609  < 2e-16
## Seat.Comfort                        0.0641651  0.0111906    5.734 9.82e-09
## Departure.Delay                     0.0039722  0.0009848    4.033 5.50e-05
## Gate.Location                       0.0254399  0.0091837    2.770  0.00560
## Food.and.Drink                     -0.0292072  0.0107003   -2.730  0.00634
##
## (Intercept)                        ***
## Online.Boarding                    ***
## Type.of.TravelPersonal             ***
## On.board.Service                   ***
## Customer.TypeReturning             ***
```

```
## Check.in.Service                        ***
## In.flight.Entertainment                 ***
## Leg.Room.Service                        ***
## ClassEconomy                            ***
## ClassEconomy Plus                       ***
## In.flight.Wifi.Service                  ***
## Departure.and.Arrival.Time.Convenience ***
## Arrival.Delay                           ***
## Cleanliness                             ***
## Baggage.Handling                        ***
## Ease.of.Online.Booking                  ***
## Age                                     ***
## In.flight.Service                       ***
## Seat.Comfort                            ***
## Departure.Delay                         ***
## Gate.Location                           **
## Food.and.Drink                          **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69368  on 103590  degrees of freedom
## AIC: 69412
##
## Number of Fisher Scoring iterations: 5
```

AIC is a measure of the relative entropy of a model, i.e. the amount of information lost when using the model to describe the data. A lower AIC value indicates a better model. During model selection, the AIC values of different models are compared. The model with the lowest AIC is considered the best of those compared.

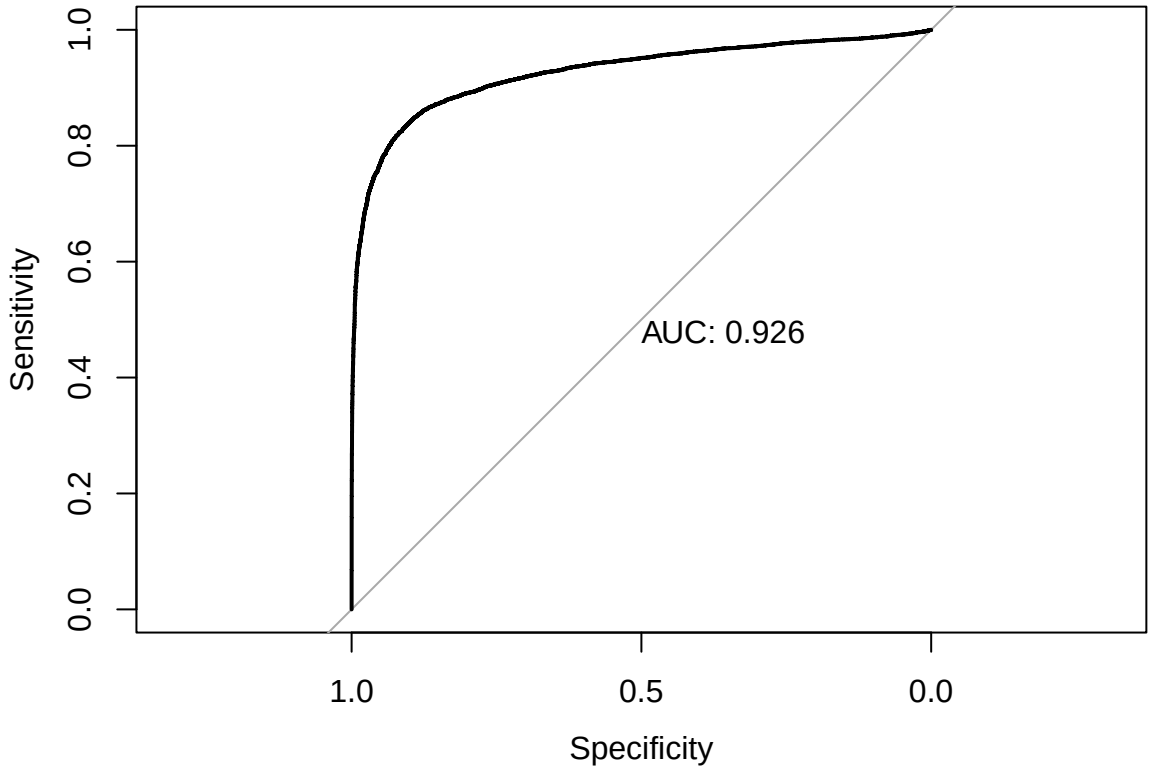
Below i run a test using the test set of mine dataset to see the capacity of the model:

```
stepwise_model_forward_probs <- predict(stepwise_model_forward, testX, type = "response")

stepwise_model_forward_roc <- roc(test$Satisfaction ~ stepwise_model_forward_probs, plot=TRUE, print.au
```

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

## Setting direction: controls < cases
```




```
coords(stepwise_model_forward_roc, x=0.5, ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold       0.5   0.9087981   0.8312206 0.8749758 13263 9377 1904 1331
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8744643 0.8757004 0.1242996 0.09120186 0.8312206 0.9087981
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1687794    0.09120186     0.1687794  0.1250242 0.1255357 0.1242996
##           precision    recall   youden closest.topleft
## threshold 0.8757004 0.8312206 1.740019      0.03680425
```

And the following considering the best threshold:

```
coords(stepwise_model_forward_roc, x="best", ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold 0.4890552   0.9051665   0.8358302 0.8749372 13210 9429 1852 1384
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8770416 0.8720059 0.1279941 0.09483349 0.8358302 0.9051665
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1641698    0.09483349     0.1641698  0.1250628 0.1229584 0.1279941
##           precision    recall   youden closest.topleft
## threshold 0.8720059 0.8358302 1.740997      0.03594513
```

```r
stepwise_model_backward <- step(full_model, scope = list(lower = null_model, upper = full_model), direc
```

Backward

```
## Start:  AIC=69413.14
## Satisfaction ~ Age + Customer.Type + Type.of.Travel + Class +
##     Flight.Distance + Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling
##
##                                        Df Deviance   AIC
## - Flight.Distance                       1    69368 69412
## <none>                                       69367 69413
## - Food.and.Drink                        1    69375 69419
## - Gate.Location                         1    69375 69419
## - Departure.Delay                       1    69384 69428
## - In.flight.Entertainment               1    69386 69430
## - Seat.Comfort                          1    69400 69444
## - Arrival.Delay                         1    69451 69495
## - In.flight.Service                     1    69460 69504
## - Age                                   1    69506 69550
## - Baggage.Handling                      1    69539 69583
## - Ease.of.Online.Booking                1    69539 69583
## - Departure.and.Arrival.Time.Convenience 1    69646 69690
## - Cleanliness                           1    69716 69760
## - On.board.Service                      1    70202 70246
## - Leg.Room.Service                      1    70226 70270
## - Class                                 2    70238 70280
## - In.flight.Wifi.Service                1    70656 70700
## - Check.in.Service                      1    70896 70940
## - Online.Boarding                       1    73139 73183
## - Customer.Type                         1    74486 74530
## - Type.of.Travel                        1    78171 78215
##
## Step:  AIC=69411.93
## Satisfaction ~ Age + Customer.Type + Type.of.Travel + Class +
##     Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling
##
##                                        Df Deviance   AIC
## <none>                                       69368 69412
## - Food.and.Drink                        1    69375 69417
## - Gate.Location                         1    69376 69418
## - Departure.Delay                       1    69384 69426
## - In.flight.Entertainment               1    69386 69428
## - Seat.Comfort                          1    69401 69443
## - Arrival.Delay                         1    69452 69494
```

```
## - In.flight.Service                         1    69461 69503
## - Age                                       1    69506 69548
## - Baggage.Handling                          1    69540 69582
## - Ease.of.Online.Booking                    1    69540 69582
## - Departure.and.Arrival.Time.Convenience    1    69648 69690
## - Cleanliness                               1    69717 69759
## - On.board.Service                          1    70202 70244
## - Leg.Room.Service                          1    70226 70268
## - Class                                     2    70329 70369
## - In.flight.Wifi.Service                    1    70664 70706
## - Check.in.Service                          1    70897 70939
## - Online.Boarding                           1    73140 73182
## - Customer.Type                             1    74830 74872
## - Type.of.Travel                            1    78298 78340
```

```
summary(stepwise_model_backward)
```

```
##
## Call:
## glm(formula = Satisfaction ~ Age + Customer.Type + Type.of.Travel +
##     Class + Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling, family = binomial,
##     data = train)
##
## Coefficients:
##                                           Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                             -7.8232629  0.0771300 -101.430  < 2e-16
## Age                                     -0.0083330  0.0007098  -11.741  < 2e-16
## Customer.TypeReturning                   2.0303950  0.0288025   70.494  < 2e-16
## Type.of.TravelPersonal                  -2.7166692  0.0312090  -87.048  < 2e-16
## ClassEconomy                            -0.7054170  0.0243630  -28.954  < 2e-16
## ClassEconomy Plus                       -0.8006458  0.0402537  -19.890  < 2e-16
## Departure.Delay                          0.0039722  0.0009848    4.033 5.50e-05
## Arrival.Delay                           -0.0088439  0.0009705   -9.113  < 2e-16
## Departure.and.Arrival.Time.Convenience -0.1371571  0.0081820  -16.763  < 2e-16
## Ease.of.Online.Booking                  -0.1479513  0.0112912  -13.103  < 2e-16
## Check.in.Service                         0.3288308  0.0085534   38.444  < 2e-16
## Online.Boarding                          0.6091938  0.0102259   59.573  < 2e-16
## Gate.Location                            0.0254399  0.0091837    2.770  0.00560
## On.board.Service                         0.2920703  0.0101986   28.638  < 2e-16
## Seat.Comfort                             0.0641651  0.0111906    5.734 9.82e-09
## Leg.Room.Service                         0.2486636  0.0085151   29.203  < 2e-16
## Cleanliness                              0.2250334  0.0120806   18.628  < 2e-16
## Food.and.Drink                          -0.0292072  0.0107003   -2.730  0.00634
## In.flight.Service                        0.1157939  0.0120512    9.609  < 2e-16
## In.flight.Wifi.Service                   0.4043205  0.0114272   35.382  < 2e-16
## In.flight.Entertainment                  0.0612187  0.0142220    4.305 1.67e-05
## Baggage.Handling                         0.1492217  0.0114204   13.066  < 2e-16
##
## (Intercept)                             ***
## Age                                     ***
```

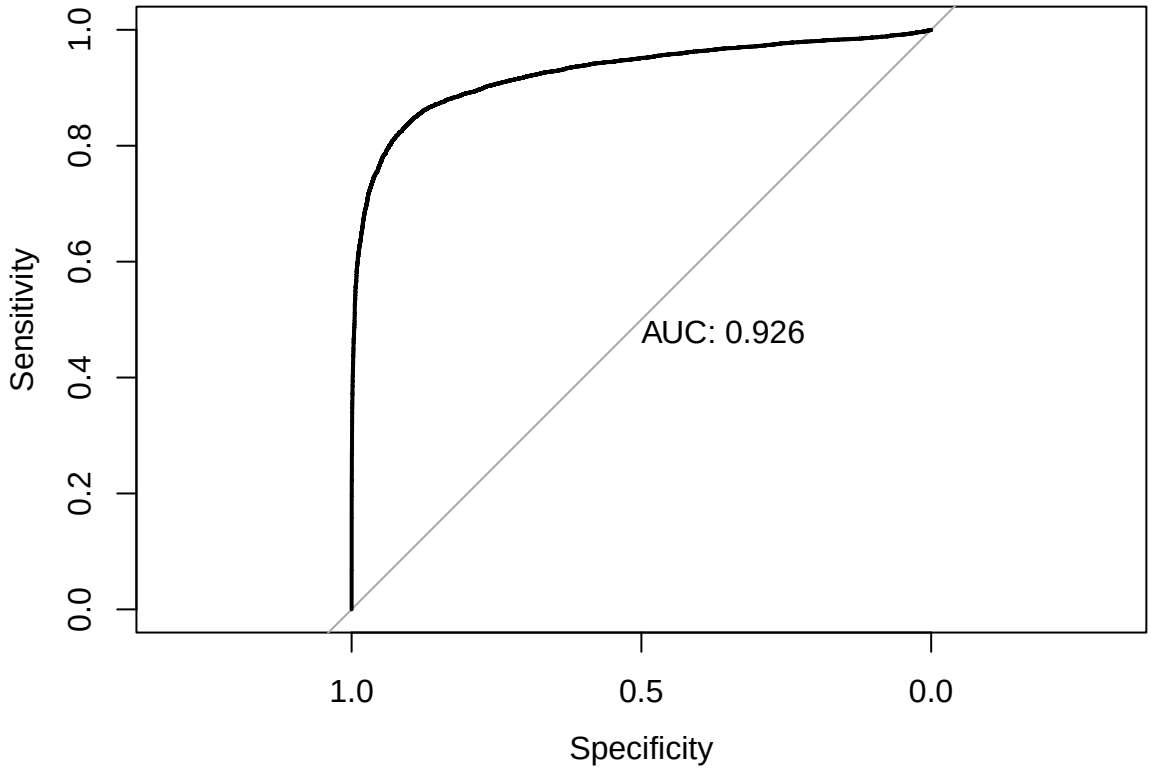
```
## Customer.TypeReturning                  ***
## Type.of.TravelPersonal                  ***
## ClassEconomy                            ***
## ClassEconomy Plus                       ***
## Departure.Delay                         ***
## Arrival.Delay                           ***
## Departure.and.Arrival.Time.Convenience ***
## Ease.of.Online.Booking                  ***
## Check.in.Service                        ***
## Online.Boarding                         ***
## Gate.Location                           **
## On.board.Service                        ***
## Seat.Comfort                            ***
## Leg.Room.Service                        ***
## Cleanliness                             ***
## Food.and.Drink                          **
## In.flight.Service                       ***
## In.flight.Wifi.Service                  ***
## In.flight.Entertainment                 ***
## Baggage.Handling                        ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69368  on 103590  degrees of freedom
## AIC: 69412
##
## Number of Fisher Scoring iterations: 5
```

```
stepwise_model_backward_probs <- predict(stepwise_model_backward, testX, type = "response")

stepwise_model_backward_roc <- roc(test$Satisfaction ~ stepwise_model_backward_probs, plot=TRUE, print.a
```

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

## Setting direction: controls < cases
```



roc function reports the following results with a threshold of 0.5:

```
coords(stepwise_model_backward_roc, x=0.5, ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold       0.5   0.9087981   0.8312206 0.8749758 13263 9377 1904 1331
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8744643 0.8757004 0.1242996 0.09120186 0.8312206 0.9087981
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1687794    0.09120186     0.1687794  0.1250242 0.1255357 0.1242996
##           precision    recall   youden closest.topleft
## threshold 0.8757004 0.8312206 1.740019      0.03680425
```

And the following considering the best threshold:

```
coords(stepwise_model_backward_roc, x="best", ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold 0.4890552   0.9051665   0.8358302 0.8749372 13210 9429 1852 1384
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8770416 0.8720059 0.1279941 0.09483349 0.8358302 0.9051665
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1641698    0.09483349     0.1641698  0.1250628 0.1229584 0.1279941
##           precision    recall   youden closest.topleft
## threshold 0.8720059 0.8358302 1.740997      0.03594513
```

```r
stepwise_model_both <- step(full_model, scope = list(lower = null_model, upper = full_model), direction
```

Both

```
## Start:  AIC=69413.14
## Satisfaction ~ Age + Customer.Type + Type.of.Travel + Class +
##     Flight.Distance + Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling
##
##                                        Df Deviance   AIC
## - Flight.Distance                       1    69368 69412
## <none>                                       69367 69413
## - Food.and.Drink                        1    69375 69419
## - Gate.Location                         1    69375 69419
## - Departure.Delay                       1    69384 69428
## - In.flight.Entertainment               1    69386 69430
## - Seat.Comfort                          1    69400 69444
## - Arrival.Delay                         1    69451 69495
## - In.flight.Service                     1    69460 69504
## - Age                                   1    69506 69550
## - Baggage.Handling                      1    69539 69583
## - Ease.of.Online.Booking                1    69539 69583
## - Departure.and.Arrival.Time.Convenience 1    69646 69690
## - Cleanliness                           1    69716 69760
## - On.board.Service                      1    70202 70246
## - Leg.Room.Service                      1    70226 70270
## - Class                                 2    70238 70280
## - In.flight.Wifi.Service                1    70656 70700
## - Check.in.Service                      1    70896 70940
## - Online.Boarding                       1    73139 73183
## - Customer.Type                         1    74486 74530
## - Type.of.Travel                        1    78171 78215
##
## Step:  AIC=69411.93
## Satisfaction ~ Age + Customer.Type + Type.of.Travel + Class +
##     Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling
##
##                                        Df Deviance   AIC
## <none>                                       69368 69412
## + Flight.Distance                       1    69367 69413
## - Food.and.Drink                        1    69375 69417
## - Gate.Location                         1    69376 69418
## - Departure.Delay                       1    69384 69426
## - In.flight.Entertainment               1    69386 69428
## - Seat.Comfort                          1    69401 69443
```

```
## - Arrival.Delay                             1    69452 69494
## - In.flight.Service                         1    69461 69503
## - Age                                       1    69506 69548
## - Baggage.Handling                          1    69540 69582
## - Ease.of.Online.Booking                    1    69540 69582
## - Departure.and.Arrival.Time.Convenience    1    69648 69690
## - Cleanliness                               1    69717 69759
## - On.board.Service                          1    70202 70244
## - Leg.Room.Service                          1    70226 70268
## - Class                                     2    70329 70369
## - In.flight.Wifi.Service                    1    70664 70706
## - Check.in.Service                          1    70897 70939
## - Online.Boarding                           1    73140 73182
## - Customer.Type                             1    74830 74872
## - Type.of.Travel                            1    78298 78340
```

```
summary(stepwise_model_both)
```

```
##
## Call:
## glm(formula = Satisfaction ~ Age + Customer.Type + Type.of.Travel +
##     Class + Departure.Delay + Arrival.Delay + Departure.and.Arrival.Time.Convenience +
##     Ease.of.Online.Booking + Check.in.Service + Online.Boarding +
##     Gate.Location + On.board.Service + Seat.Comfort + Leg.Room.Service +
##     Cleanliness + Food.and.Drink + In.flight.Service + In.flight.Wifi.Service +
##     In.flight.Entertainment + Baggage.Handling, family = binomial,
##     data = train)
##
## Coefficients:
##                                       Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                         -7.8232629  0.0771300 -101.430  < 2e-16
## Age                                 -0.0083330  0.0007098  -11.741  < 2e-16
## Customer.TypeReturning               2.0303950  0.0288025   70.494  < 2e-16
## Type.of.TravelPersonal              -2.7166692  0.0312090  -87.048  < 2e-16
## ClassEconomy                        -0.7054170  0.0243630  -28.954  < 2e-16
## ClassEconomy Plus                   -0.8006458  0.0402537  -19.890  < 2e-16
## Departure.Delay                      0.0039722  0.0009848    4.033 5.50e-05
## Arrival.Delay                       -0.0088439  0.0009705   -9.113  < 2e-16
## Departure.and.Arrival.Time.Convenience -0.1371571  0.0081820  -16.763  < 2e-16
## Ease.of.Online.Booking              -0.1479513  0.0112912  -13.103  < 2e-16
## Check.in.Service                     0.3288308  0.0085534   38.444  < 2e-16
## Online.Boarding                      0.6091938  0.0102259   59.573  < 2e-16
## Gate.Location                        0.0254399  0.0091837    2.770  0.00560
## On.board.Service                     0.2920703  0.0101986   28.638  < 2e-16
## Seat.Comfort                         0.0641651  0.0111906    5.734 9.82e-09
## Leg.Room.Service                     0.2486636  0.0085151   29.203  < 2e-16
## Cleanliness                          0.2250334  0.0120806   18.628  < 2e-16
## Food.and.Drink                      -0.0292072  0.0107003   -2.730  0.00634
## In.flight.Service                    0.1157939  0.0120512    9.609  < 2e-16
## In.flight.Wifi.Service               0.4043205  0.0114272   35.382  < 2e-16
## In.flight.Entertainment              0.0612187  0.0142220    4.305 1.67e-05
## Baggage.Handling                     0.1492217  0.0114204   13.066  < 2e-16
##
## (Intercept)                         ***
```

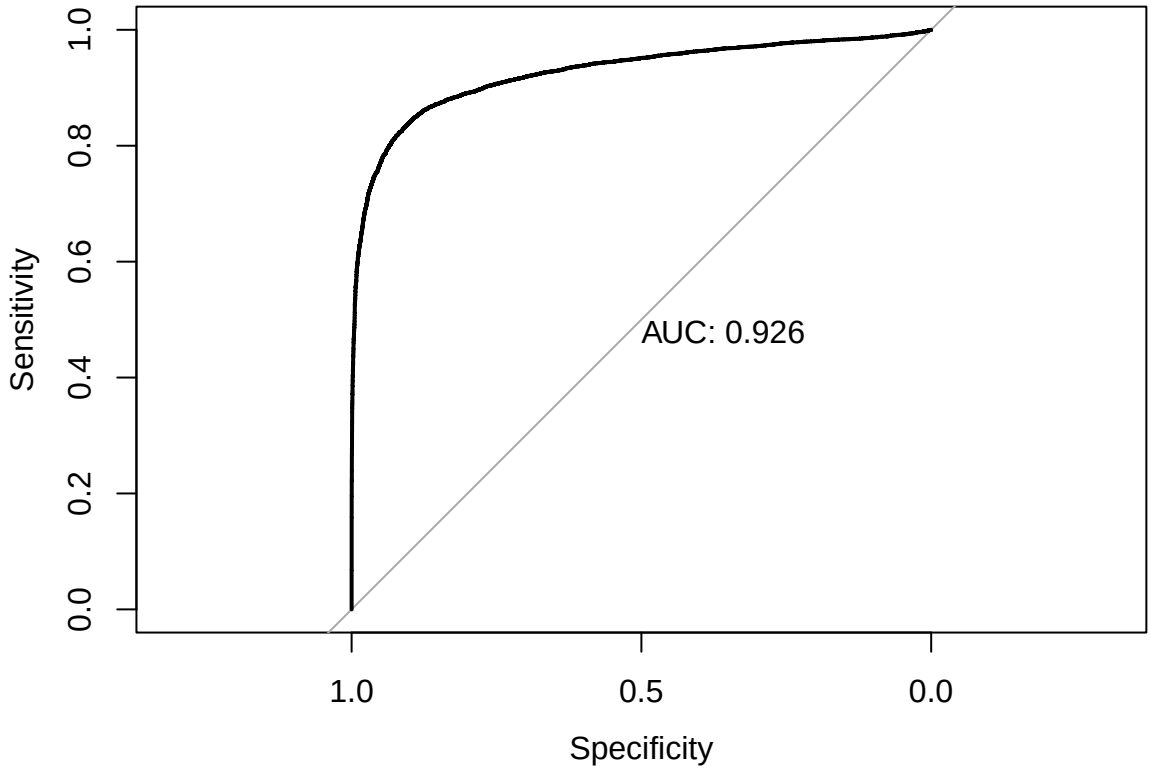
```
## Age                                       ***
## Customer.TypeReturning                    ***
## Type.of.TravelPersonal                    ***
## ClassEconomy                              ***
## ClassEconomy Plus                         ***
## Departure.Delay                           ***
## Arrival.Delay                             ***
## Departure.and.Arrival.Time.Convenience    ***
## Ease.of.Online.Booking                    ***
## Check.in.Service                          ***
## Online.Boarding                           ***
## Gate.Location                             **
## On.board.Service                          ***
## Seat.Comfort                              ***
## Leg.Room.Service                          ***
## Cleanliness                               ***
## Food.and.Drink                            **
## In.flight.Service                         ***
## In.flight.Wifi.Service                    ***
## In.flight.Entertainment                   ***
## Baggage.Handling                          ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69368  on 103590  degrees of freedom
## AIC: 69412
##
## Number of Fisher Scoring iterations: 5
```

```
stepwise_model_both_probs <- predict(stepwise_model_both, testX, type = "response")

stepwise_model_both_roc <- roc(test$Satisfaction ~ stepwise_model_both_probs, plot=TRUE, print.auc=TRUE
```

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

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



roc function reports the following results with a threshold of 0.5:

```
coords(stepwise_model_backward_roc, x=0.5, ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold       0.5   0.9087981   0.8312206 0.8749758 13263 9377 1904 1331
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8744643 0.8757004 0.1242996 0.09120186 0.8312206 0.9087981
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1687794    0.09120186     0.1687794  0.1250242 0.1255357 0.1242996
##           precision    recall   youden closest.topleft
## threshold 0.8757004 0.8312206 1.740019      0.03680425
```

And the following considering the best threshold:

```
coords(stepwise_model_backward_roc, x="best", ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold 0.4890552   0.9051665   0.8358302 0.8749372 13210 9429 1852 1384
##                 npv       ppv       fdr        fpr       tpr       tnr
## threshold 0.8770416 0.8720059 0.1279941 0.09483349 0.8358302 0.9051665
##                 fnr 1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv
## threshold 0.1641698    0.09483349     0.1641698  0.1250628 0.1229584 0.1279941
##           precision    recall   youden closest.topleft
## threshold 0.8720059 0.8358302 1.740997      0.03594513
```

Comparison between models

```
AIC(new_model)
```

```
## [1] 69413.14
```

```
AIC(stepwise_model_forward)
```

```
## [1] 69411.93
```

```
AIC(stepwise_model_backward)
```

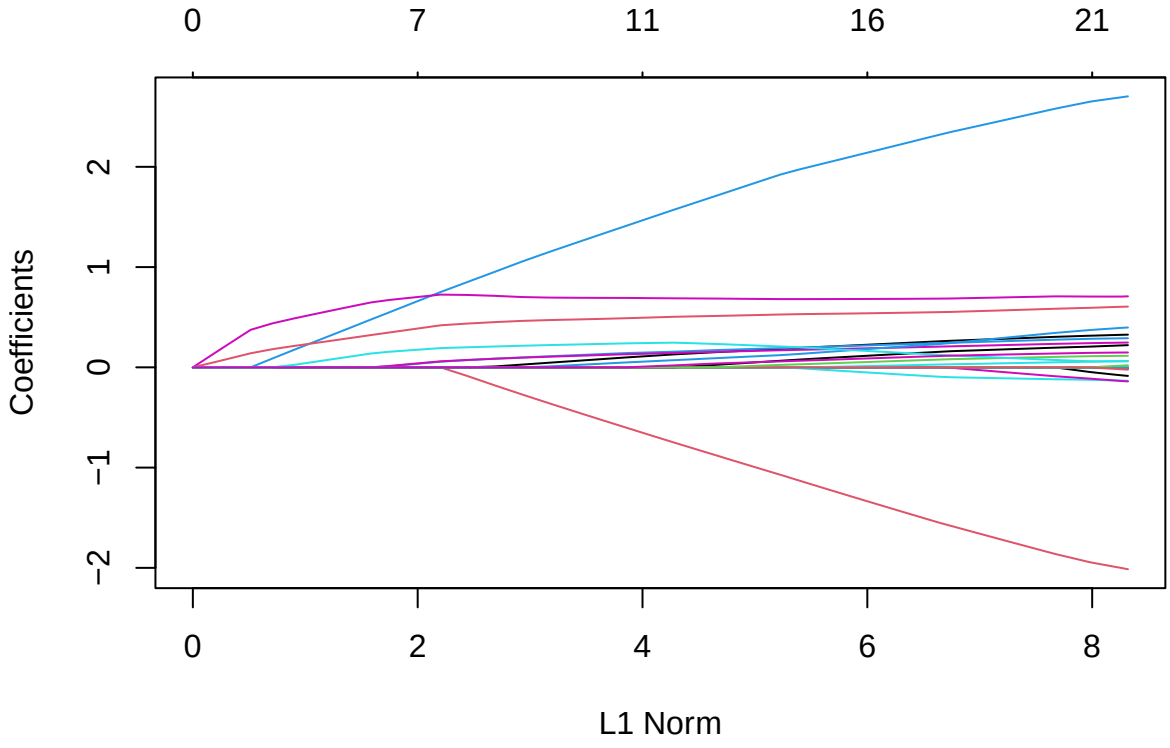
```
## [1] 69411.93
```

```
AIC(stepwise_model_both)
```

```
## [1] 69411.93
```

Lasso

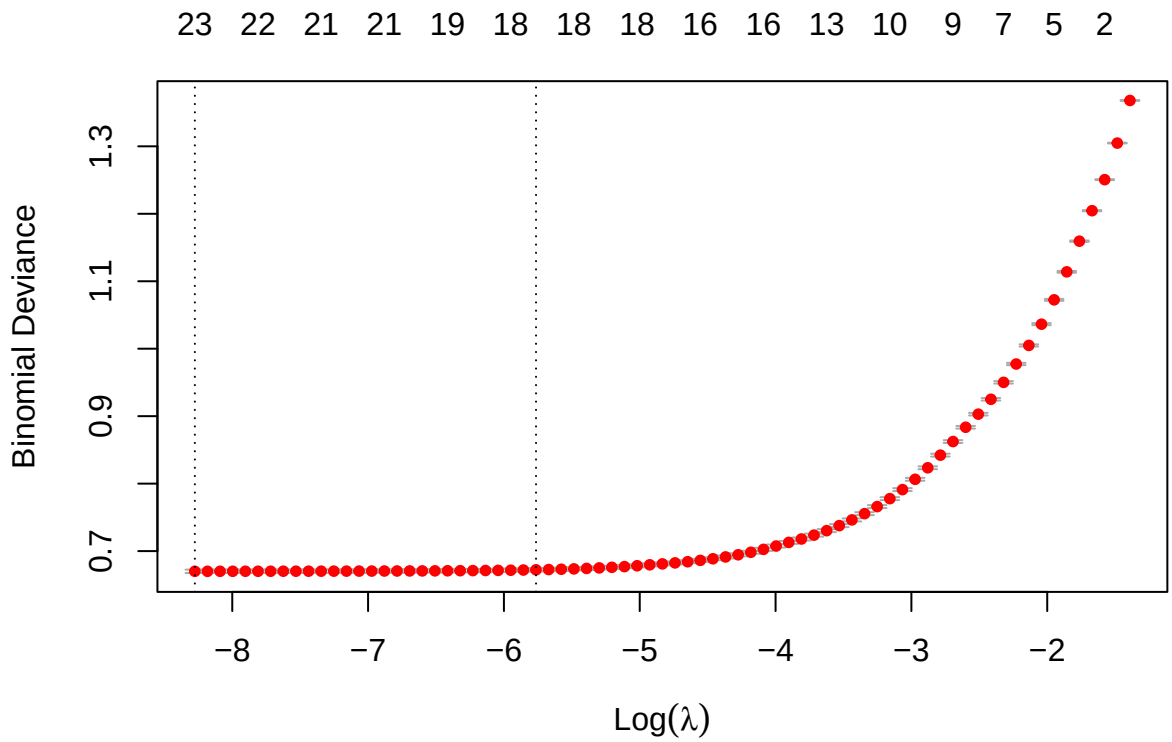
```
lasso_model<-glmnet(Satisfaction ~ ., data=train,family = "binomial", alpha = 1)
plot(lasso_model)
```



```
set.seed(1)
cv.out <- cv.glmnet(Satisfaction ~ ., data=train,family = "binomial", alpha = 1, K=5)
cv.out
```

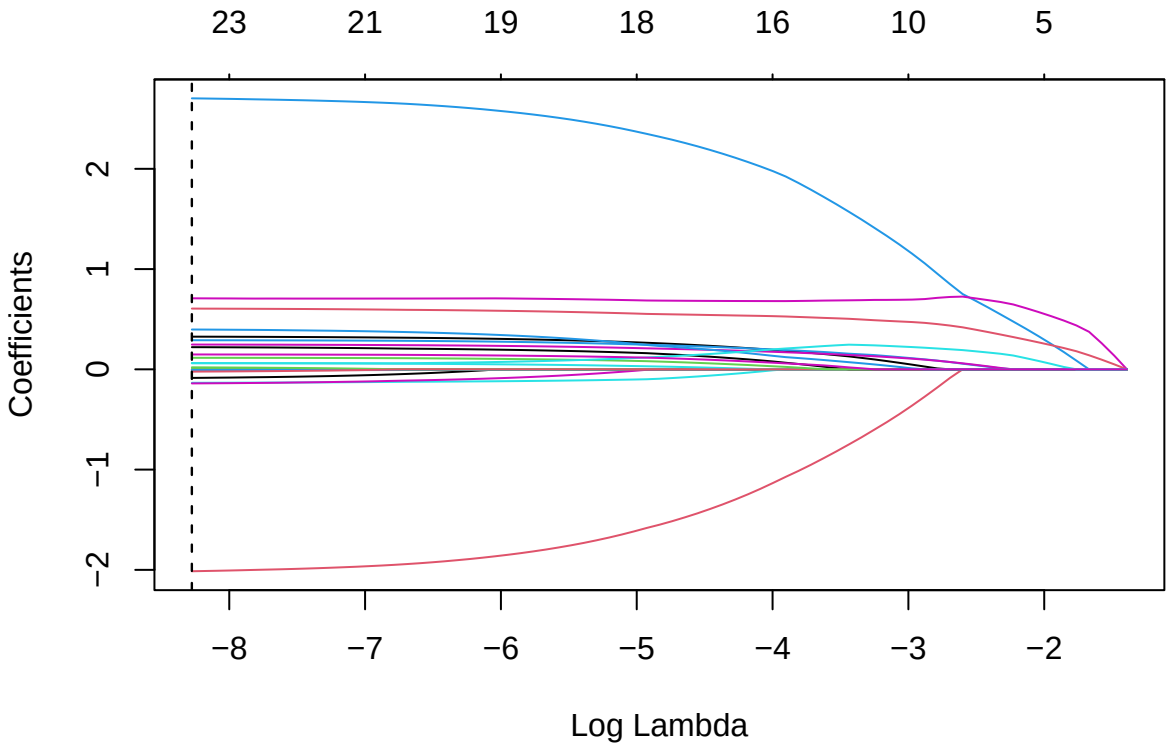
```
## Call:
## cv.glmnet.formula(formula = Satisfaction ~ ., data = train, alpha = 1,
##     family = "binomial", K = 5)
##
## Model fitting options:
##     Sparse model matrix: FALSE
##     Use model.frame: FALSE
##     Number of crossvalidation folds: 10
##     Alpha: 1
##     Deviance-minimizing lambda: 0.0002547004  (+1 SE): 0.003140066
```

```
plot(cv.out)
```




```
bestlam.lasso <- cv.out$lambda.min
lasso.final <- glmnet(Satisfaction ~ .,  data=train , family = "binomial", alpha = 1, lambda = bestlam.l
```

```
plot(lasso_model, xvar = "lambda")
abline(v = log(bestlam.lasso), lwd = 1.2, lty = "dashed")
```



The Lasso solution for the selected value of lambda is:

```
coef(lasso_model, bestlam.lasso)
```

```
## 26 x 1 sparse Matrix of class "dgCMatrix"
##                                                s1
## (Intercept)                          -9.177361e+00
## Age                                  -7.973963e-03
## Customer.TypeFirst-time              -2.013865e+00
## Customer.TypeReturning                .
## Type.of.TravelBusiness                2.702869e+00
## Type.of.TravelPersonal               -9.627720e-11
## ClassBusiness                         7.082433e-01
## ClassEconomy                          .
## ClassEconomy Plus                    -8.553711e-02
## Flight.Distance                      -3.680257e-06
## Departure.Delay                       1.836347e-03
## Arrival.Delay                        -6.717406e-03
## Departure.and.Arrival.Time.Convenience -1.342776e-01
## Ease.of.Online.Booking               -1.402984e-01
## Check.in.Service                      3.259034e-01
## Online.Boarding                       6.058646e-01
## Gate.Location                         1.991065e-02
## On.board.Service                      2.904374e-01
## Seat.Comfort                          6.248349e-02
## Leg.Room.Service                      2.471839e-01
```

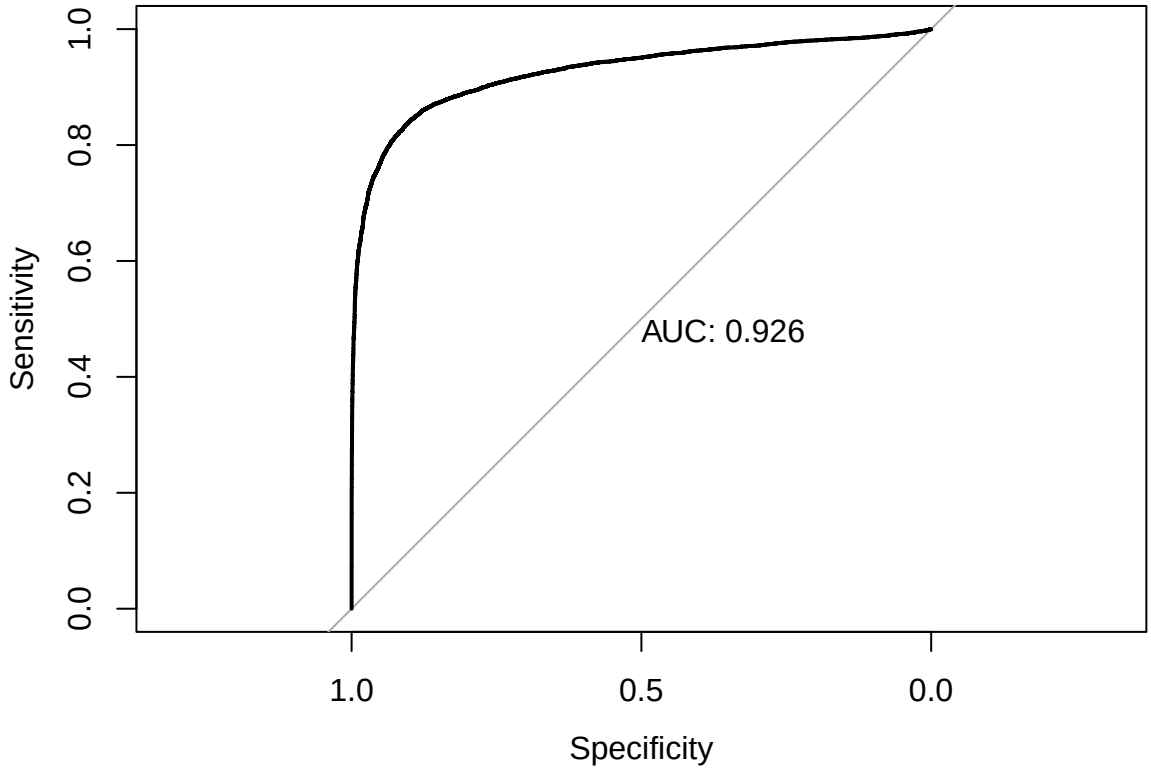
```
## Cleanliness                          2.210792e-01
## Food.and.Drink                      -2.277884e-02
## In.flight.Service                    1.148466e-01
## In.flight.Wifi.Service               3.971940e-01
## In.flight.Entertainment              6.080266e-02
## Baggage.Handling                     1.481939e-01
```

```
lasso_model_probs <- predict(lasso_model, s = bestlam.lasso, newdata=test, type="response")
lasso_model_roc <- roc(test$Satisfaction ~ lasso_model_probs, plot=TRUE, print.auc=TRUE)
```

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

```
## Warning in roc.default(response, predictors[, 1], ...): Deprecated use a matrix
## as predictor. Unexpected results may be produced, please pass a numeric vector.
```

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



Roc function reports the following results with a threshold of 0.5:

```
coords(lasso_model_roc, x=0.5, ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold       0.5   0.9087981   0.8313093 0.8750145 13263 9378 1903 1331
##               npv       ppv       fdr       fpr       tpr       tnr       fnr
```

```
## threshold 0.874522 0.875712 0.124288 0.09120186 0.8313093 0.9087981 0.1686907
##           1-specificity 1-sensitivity 1-accuracy    1-npv    1-ppv precision
## threshold    0.09120186     0.1686907  0.1249855 0.125478 0.124288  0.875712
##              recall   youden closest.topleft
## threshold 0.8313093 1.740107      0.03677434
```

And the following considering the best threshold:

```
coords(lasso_model_roc, x="best", ret="all")
```

```
##           threshold specificity sensitivity  accuracy    tn   tp   fn   fp
## threshold 0.4712345   0.8988625   0.8428331 0.8744348 13118 9508 1773 1476
##                 npv       ppv       fdr       fpr       tpr       tnr       fnr
## threshold 0.8809348 0.8656227 0.1343773 0.1011375 0.8428331 0.8988625 0.1571669
##           1-specificity 1-sensitivity 1-accuracy     1-npv     1-ppv precision
## threshold     0.1011375     0.1571669  0.1255652 0.1190652 0.1343773 0.8656227
##              recall   youden closest.topleft
## threshold 0.8428331 1.741696      0.03493022
```

We can also try to fit a logistic regression model choosing the predictors suggested by Lasso.

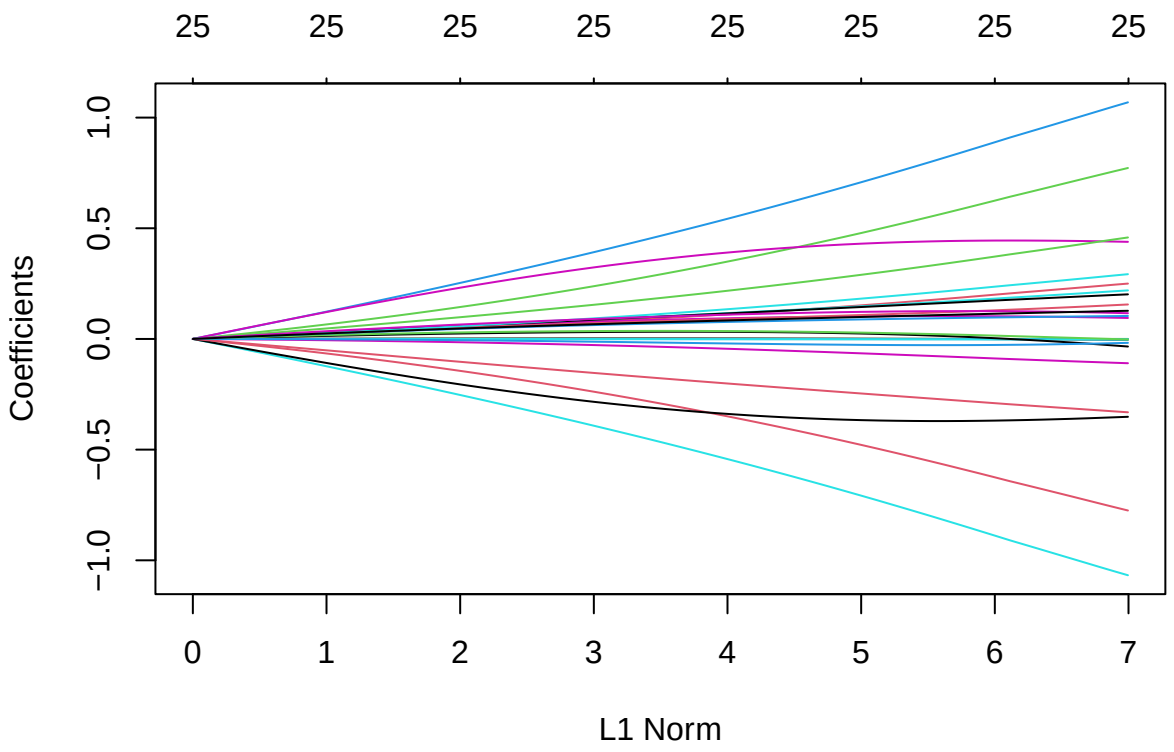
```
lasso_log <- glm(Satisfaction ~ Age +Customer.Type + Type.of.Travel + Class + Flight.Distance + Departu
+ Departure.and.Arrival.Time.Convenience + Ease.of.Online.Booking + Check.in.Service + Online.Boarding
 Food.and.Drink + In.flight.Service + In.flight.Wifi.Service + In.flight.Entertainment + Baggage.Handli
summary(lasso_log)
```

```
##
## Call:
## glm(formula = Satisfaction ~ Age + Customer.Type + Type.of.Travel +
##     Class + Flight.Distance + Departure.Delay + Arrival.Delay +
##     Departure.and.Arrival.Time.Convenience + Ease.of.Online.Booking +
##     Check.in.Service + Online.Boarding + Gate.Location + On.board.Service +
##     Seat.Comfort + Leg.Room.Service + Cleanliness + Food.and.Drink +
##     In.flight.Service + In.flight.Wifi.Service + In.flight.Entertainment +
##     Baggage.Handling, family = binomial, data = train)
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                            -7.812e+00  7.823e-02 -99.852  < 2e-16
## Age                                    -8.375e-03  7.113e-04 -11.773  < 2e-16
## Customer.TypeReturning                  2.037e+00  2.986e-02  68.235  < 2e-16
## Type.of.TravelPersonal                 -2.720e+00  3.142e-02 -86.564  < 2e-16
## ClassEconomy                           -7.125e-01  2.563e-02 -27.797  < 2e-16
## ClassEconomy Plus                      -8.091e-01  4.137e-02 -19.557  < 2e-16
## Flight.Distance                        -1.001e-05  1.127e-05  -0.888  0.37457
## Departure.Delay                         3.977e-03  9.849e-04   4.038 5.39e-05
## Arrival.Delay                          -8.848e-03  9.706e-04  -9.116  < 2e-16
## Departure.and.Arrival.Time.Convenience -1.371e-01  8.183e-03 -16.748  < 2e-16
## Ease.of.Online.Booking                 -1.478e-01  1.129e-02 -13.087  < 2e-16
## Check.in.Service                        3.289e-01  8.554e-03  38.448  < 2e-16
## Online.Boarding                         6.095e-01  1.023e-02  59.575  < 2e-16
## Gate.Location                           2.536e-02  9.185e-03   2.761  0.00576
```

```
## On.board.Service                          2.922e-01  1.020e-02  28.642  < 2e-16
## Seat.Comfort                              6.435e-02  1.119e-02   5.749 8.97e-09
## Leg.Room.Service                          2.489e-01  8.521e-03  29.214  < 2e-16
## Cleanliness                               2.251e-01  1.208e-02  18.632  < 2e-16
## Food.and.Drink                           -2.937e-02  1.070e-02  -2.745  0.00606
## In.flight.Service                         1.156e-01  1.206e-02   9.590  < 2e-16
## In.flight.Wifi.Service                    4.038e-01  1.144e-02  35.286  < 2e-16
## In.flight.Entertainment                   6.127e-02  1.422e-02   4.308 1.65e-05
## Baggage.Handling                          1.491e-01  1.142e-02  13.054  < 2e-16
##
## (Intercept)                              ***
## Age                                      ***
## Customer.TypeReturning                   ***
## Type.of.TravelPersonal                   ***
## ClassEconomy                             ***
## ClassEconomy Plus                        ***
## Flight.Distance
## Departure.Delay                          ***
## Arrival.Delay                            ***
## Departure.and.Arrival.Time.Convenience   ***
## Ease.of.Online.Booking                   ***
## Check.in.Service                         ***
## Online.Boarding                          ***
## Gate.Location                            **
## On.board.Service                         ***
## Seat.Comfort                             ***
## Leg.Room.Service                         ***
## Cleanliness                              ***
## Food.and.Drink                           **
## In.flight.Service                        ***
## In.flight.Wifi.Service                   ***
## In.flight.Entertainment                  ***
## Baggage.Handling                         ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141833  on 103611  degrees of freedom
## Residual deviance:  69367  on 103589  degrees of freedom
## AIC: 69413
##
## Number of Fisher Scoring iterations: 5
```

Ridge Regression

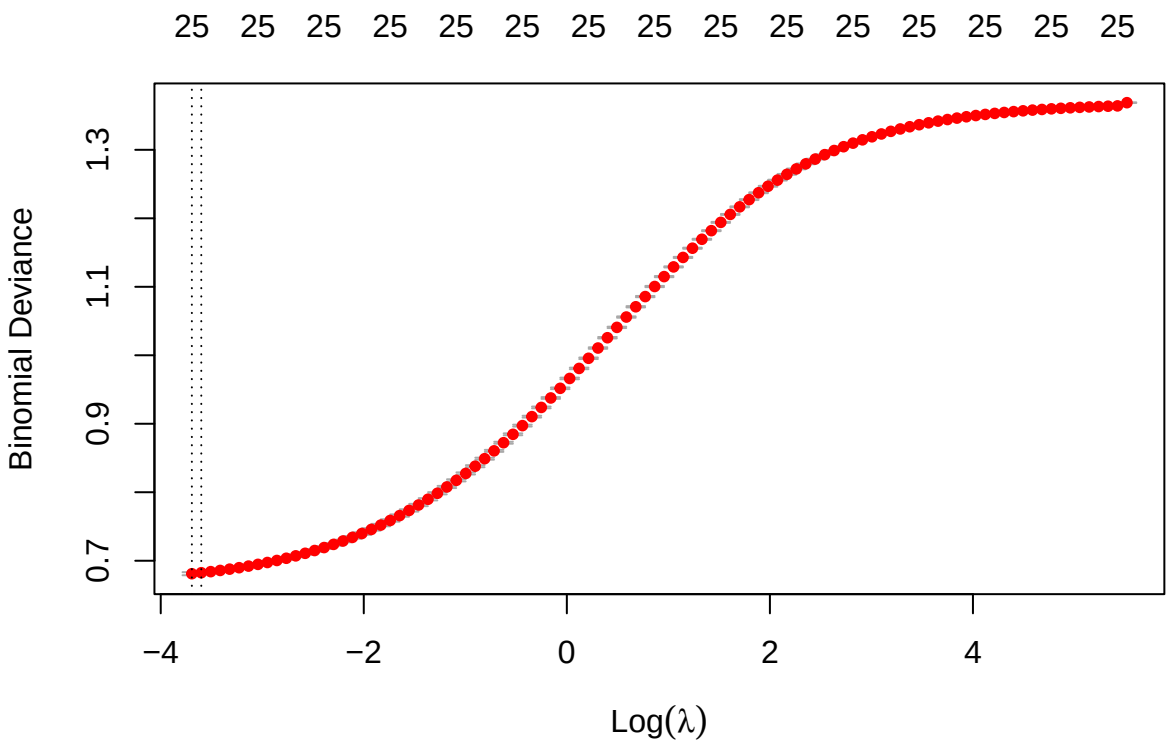
```
ridge_model<-glmnet(Satisfaction ~ ., data=train, family = "binomial", alpha = 0)
plot(ridge_model)
```




```
set.seed(1)
cv.out <- cv.glmnet(Satisfaction ~ ., data=train,family = "binomial", alpha = 0, K=5)
cv.out
```

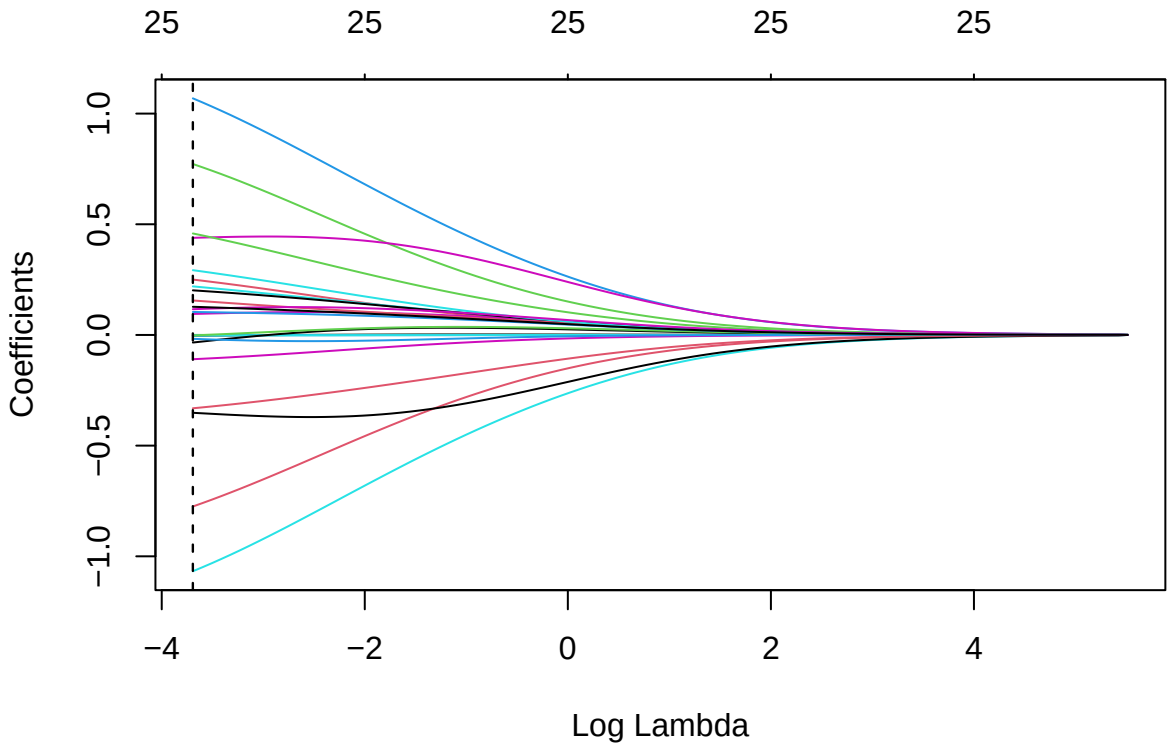
```
## Call:
## cv.glmnet.formula(formula = Satisfaction ~ ., data = train, alpha = 0,
##     family = "binomial", K = 5)
##
## Model fitting options:
##     Sparse model matrix: FALSE
##     Use model.frame: FALSE
##     Number of crossvalidation folds: 10
##     Alpha: 0
##     Deviance-minimizing lambda: 0.02488449  (+1 SE): 0.02731069
```

```
plot(cv.out)
```




```
bestlam_ridge <- cv.out$lambda.min
ridge_final <- glmnet(Satisfaction ~ .,  data=train , family = "binomial", alpha = 0, lambda = bestlam_
```

```
plot(ridge_model, xvar = "lambda")
abline(v = log(bestlam_ridge), lwd = 1.2, lty = "dashed")
```



The Lasso solution for the selected value of lambda is:

```
coef(ridge_model, bestlam_ridge)
```

```
## 26 x 1 sparse Matrix of class "dgCMatrix"
##                                                 s1
## (Intercept)                             -7.375106955
## Age                                     -0.003153999
## Customer.TypeFirst-time                 -0.775017980
## Customer.TypeReturning                   0.772296382
## Type.of.TravelBusiness                   1.068815868
## Type.of.TravelPersonal                  -1.067441746
## ClassBusiness                            0.438677534
## ClassEconomy                            -0.351676142
## ClassEconomy Plus                       -0.331346924
## Flight.Distance                          0.000045315
## Departure.Delay                         -0.001095376
## Arrival.Delay                           -0.002675885
## Departure.and.Arrival.Time.Convenience  -0.109661852
## Ease.of.Online.Booking                  -0.034267169
## Check.in.Service                         0.250084772
## Online.Boarding                          0.458477043
## Gate.Location                           -0.017580800
## On.board.Service                         0.219392018
## Seat.Comfort                             0.094743003
## Leg.Room.Service                         0.201541917
```

```
## Cleanliness                          0.155735887
## Food.and.Drink                      -0.002332838
## In.flight.Service                    0.103839049
## In.flight.Wifi.Service               0.292585507
## In.flight.Entertainment              0.116237488
## Baggage.Handling                     0.127236128
```

```
ridge_model_probs <- predict(ridge_model, s = bestlam_ridge, newdata=test, type="response")
ridge_model_roc <- roc(test$Satisfaction ~ ridge_model_probs, plot=TRUE, print.auc=TRUE)
```

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

```
## Warning in roc.default(response, predictors[, 1], ...): Deprecated use a matrix
## as predictor. Unexpected results may be produced, please pass a numeric vector.
```

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