512_project

Dongru Jia

4/19/2020

```
library(foreach)
library(doParallel)
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

```
library(pROC)
```

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

```
##
## Attaching package: 'pROC'
```

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

```
library(bestglm)
```

```
## Loading required package: leaps
```

```
library(MASS)
library(ggplot2)
```

```
df = read.csv("fifa.csv")
head(df)
```

```
##   X     ID              Name Age                                          Photo
## 1 0 158023          L. Messi  31 https://cdn.sofifa.org/players/4/19/158023.png
## 2 1  20801 Cristiano Ronaldo  33  https://cdn.sofifa.org/players/4/19/20801.png
## 3 2 190871         Neymar Jr  26 https://cdn.sofifa.org/players/4/19/190871.png
## 4 3 193080            De Gea  27 https://cdn.sofifa.org/players/4/19/193080.png
## 5 4 192985      K. De Bruyne  27 https://cdn.sofifa.org/players/4/19/192985.png
## 6 5 183277         E. Hazard  27 https://cdn.sofifa.org/players/4/19/183277.png
##   Nationality                                Flag Overall Potential
## 1   Argentina https://cdn.sofifa.org/flags/52.png      94        94
## 2    Portugal https://cdn.sofifa.org/flags/38.png      94        94
## 3      Brazil https://cdn.sofifa.org/flags/54.png      92        93
## 4       Spain https://cdn.sofifa.org/flags/45.png      91        93
## 5     Belgium  https://cdn.sofifa.org/flags/7.png      91        92
## 6     Belgium  https://cdn.sofifa.org/flags/7.png      91        91
##                  Club                                     Club.Logo    Value
## 1        FC Barcelona https://cdn.sofifa.org/teams/2/light/241.png €110.5M
## 2            Juventus  https://cdn.sofifa.org/teams/2/light/45.png    €77M
## 3 Paris Saint-Germain  https://cdn.sofifa.org/teams/2/light/73.png €118.5M
```

```
## 4   Manchester United  https://cdn.sofifa.org/teams/2/light/11.png    €72M
## 5     Manchester City  https://cdn.sofifa.org/teams/2/light/10.png   €102M
## 6             Chelsea   https://cdn.sofifa.org/teams/2/light/5.png    €93M
##     Wage Special Preferred.Foot International.Reputation Weak.Foot Skill.Moves
## 1 €565K    2202           Left                        5         4           4
## 2 €405K    2228          Right                        5         4           5
## 3 €290K    2143          Right                        5         5           5
## 4 €260K    1471          Right                        4         3           1
## 5 €355K    2281          Right                        4         5           4
## 6 €340K    2142          Right                        4         4           4
##        Work.Rate  Body.Type Real.Face Position Jersey.Number       Joined
## 1 Medium/ Medium      Messi       Yes       RF            10  Jul 1, 2004
## 2      High/ Low C. Ronaldo       Yes       ST             7 Jul 10, 2018
## 3   High/ Medium     Neymar       Yes       LW            10  Aug 3, 2017
## 4 Medium/ Medium       Lean       Yes       GK             1  Jul 1, 2011
## 5     High/ High     Normal       Yes      RCM             7 Aug 30, 2015
## 6   High/ Medium     Normal       Yes       LF            10  Jul 1, 2012
##   Loaned.From Contract.Valid.Until Height Weight   LS   ST   RS   LW   LF   CF
## 1                            2021    5'7 159lbs 88+2 88+2 88+2 92+2 93+2 93+2
## 2                            2022    6'2 183lbs 91+3 91+3 91+3 89+3 90+3 90+3
## 3                            2022    5'9 150lbs 84+3 84+3 84+3 89+3 89+3 89+3
## 4                            2020    6'4 168lbs
## 5                            2023   5'11 154lbs 82+3 82+3 82+3 87+3 87+3 87+3
## 6                            2020    5'8 163lbs 83+3 83+3 83+3 89+3 88+3 88+3
##     RF   RW  LAM  CAM  RAM   LM  LCM   CM  RCM   RM  LWB  LDM  CDM  RDM  RWB
## 1 93+2 92+2 93+2 93+2 93+2 91+2 84+2 84+2 84+2 91+2 64+2 61+2 61+2 61+2 64+2
## 2 90+3 89+3 88+3 88+3 88+3 88+3 81+3 81+3 81+3 88+3 65+3 61+3 61+3 61+3 65+3
## 3 89+3 89+3 89+3 89+3 89+3 88+3 81+3 81+3 81+3 88+3 65+3 60+3 60+3 60+3 65+3
## 4
## 5 87+3 87+3 88+3 88+3 88+3 88+3 87+3 87+3 87+3 88+3 77+3 77+3 77+3 77+3 77+3
## 6 88+3 89+3 89+3 89+3 89+3 89+3 82+3 82+3 82+3 89+3 66+3 63+3 63+3 63+3 66+3
##     LB  LCC   CB  RCB   RB Crossing Finishing HeadingAccuracy ShortPassing
## 1 59+2 47+2 47+2 47+2 59+2       84        95              70           90
## 2 61+3 53+3 53+3 53+3 61+3       84        94              89           81
## 3 60+3 47+3 47+3 47+3 60+3       79        87              62           84
## 4                            17        13              21           50
## 5 73+3 66+3 66+3 66+3 73+3       93        82              55           92
## 6 60+3 49+3 49+3 49+3 60+3       81        84              61           89
##   Volleys Dribbling Curve FKAccuracy LongPassing BallControl Acceleration
## 1      86        97    93         94          87          96           91
## 2      87        88    81         76          77          94           89
## 3      84        96    88         87          78          95           94
## 4      13        18    21         19          51          42           57
## 5      82        86    85         83          91          91           78
## 6      80        95    83         79          83          94           94
##   SprintSpeed Agility Reactions Balance ShotPower Jumping Stamina Strength
## 1          86      91        95      95        85      68      72       59
## 2          91      87        96      70        95      95      88       79
## 3          90      96        94      84        80      61      81       49
## 4          58      60        90      43        31      67      43       64
## 5          76      79        91      77        91      63      90       75
## 6          88      95        90      94        82      56      83       66
##   LongShots Aggression Interceptions Positioning Vision Penalties Composure
## 1        94         48            22          94     94        75        96
```

```
## 2        93         63           29          95     82        85        95
## 3        82         56           36          89     87        81        94
## 4        12         38           30          12     68        40        68
## 5        91         76           61          87     94        79        88
## 6        80         54           41          87     89        86        91
##   Marking StandingTackle SlidingTackle GKDiving GKHandling GKKicking
## 1      33             28            26        6          11        15
## 2      28             31            23        7          11        15
## 3      27             24            33        9           9        15
## 4      15             21            13       90          85        87
## 5      68             58            51       15          13         5
## 6      34             27            22       11          12         6
##   GKPositioning GKReflexes Release.Clause
## 1            14          8       €226.5M
## 2            14         11       €127.1M
## 3            15         11       €228.1M
## 4            88         94       €138.6M
## 5            10         13       €196.4M
## 6             8          8       €172.1M
```

```
feature_list = c('Age', 'Nationality', 'Club', 'Value', 'Wage', 'Preferred.Foot', 'Position', 'Jersey.Nu
                 'Joined', 'Contract.Valid.Until', 'Height', 'Weight', 'Overall', "Potential")
fifa = subset(df, select = feature_list)
head(fifa)
```

```
##   Age Nationality               Club    Value   Wage Preferred.Foot Position
## 1  31   Argentina       FC Barcelona €110.5M €565K           Left       RF
## 2  33    Portugal           Juventus    €77M €405K          Right       ST
## 3  26      Brazil Paris Saint-Germain €118.5M €290K          Right       LW
## 4  27       Spain   Manchester United    €72M €260K          Right       GK
## 5  27     Belgium     Manchester City   €102M €355K          Right      RCM
## 6  27     Belgium             Chelsea    €93M €340K          Right       LF
##   Jersey.Number       Joined Contract.Valid.Until Height Weight Overall
## 1            10  Jul 1, 2004                 2021    5'7 159lbs      94
## 2             7 Jul 10, 2018                 2022    6'2 183lbs      94
## 3            10  Aug 3, 2017                 2022    5'9 150lbs      92
## 4             1  Jul 1, 2011                 2020    6'4 168lbs      91
## 5             7 Aug 30, 2015                 2023   5'11 154lbs      91
## 6            10  Jul 1, 2012                 2020    5'8 163lbs      91
##   Potential
## 1        94
## 2        94
## 3        93
## 4        93
## 5        92
## 6        91
```

```
## Register multi-cores computing
numCores <- detectCores()
cl <- makeCluster(numCores)
registerDoParallel(cl)
```

```
## Data Cleaning

# Clean Value column
fifa$Value = gsub("[\\€]", "", fifa$Value)
```

```
fifa$Value = foreach (i=fifa$Value, .combine=c) %dopar% {
  if(grepl("M",i)){
    as.numeric(gsub("\\M", "", i))*1000000
  }else if(grepl("K",i)){
    as.numeric(gsub("\\K", "", i))*1000
  }else{
    as.numeric(i)
  }
}
# Log transform Value column
fifa$Value = log(fifa$Value)

# Clean Wage column
fifa$Wage = gsub("[\\€]", "", fifa$Wage)
fifa$Wage = foreach (i=fifa$Wage, .combine=c) %dopar% {
  if(grepl("K",i)){
    as.numeric(gsub("\\K", "", i))*1000
  }else{
    as.numeric(i)
  }
}
# Log transform Wage column
fifa$Wage = log(fifa$Wage)

# Set Jersey.Number as factor
fifa$Jersey.Number = as.factor(fifa$Jersey.Number)

# Convert Height to meters, divide by 3.281
fifa$Height = gsub("[\\']", ".", fifa$Height)
fifa$Height = foreach (i=fifa$Height, .combine=c) %dopar% {
  round(as.numeric(i)/3.281, 2)
}

# Clean Weight column
fifa$Weight = as.numeric(gsub("\\lbs", "", fifa$Weight))

# Build a new feature, "Contract.Duration", by using Contract.Valid.Until to subtract year of Joined
fifa$Contract.Duration = as.numeric(sub(".*(\\d{4})$", "\\1", fifa$Contract.Valid.Until)) -
  as.numeric(sub(".*(\\d{4})$", "\\1", fifa$Joined))

# Build a new feature, "Improved", if Potential is higher than Overall then True, otherwise False
fifa$Improved = ifelse(fifa$Potential-fifa$Overall>0, T, F)

# Imputate missing values in "Contract.Duration" with median
fifa[is.na(fifa$Contract.Duration),"Contract.Duration"] = median(fifa$Contract.Duration, na.rm = T)

# Drop unnecessary columns
fifa = subset(fifa, select = -c(Joined, Contract.Valid.Until))

# Drop rows with Inf values in Value and Wage columns
fifa = fifa[!is.infinite(fifa$Value),]
fifa = fifa[!is.infinite(fifa$Wage),]
```

```r
# Drop rows with white space value in Preferred.Foot
fifa = fifa[!(fifa$Preferred.Foot != "Left" & fifa$Preferred.Foot != "Right"), ]
```

```r
## Standardize Value and Wage
mean_value = mean(fifa$Value)
sd_value = sd(fifa$Value)
fifa$Value = (fifa$Value-mean_value)/sd_value

mean_wage = mean(fifa$Wage)
sd_wage = sd(fifa$Wage)
fifa$Wage = (fifa$Wage-mean_wage)/sd_wage

write.csv(fifa, "fifa_cleaned_dj.csv", row.names = F)
```

```r
# Fit logistic regression
fifa = read.csv("fifa_cleaned_dj.csv")
fifa = subset(fifa, select = -c(Nationality, Club, Potential, Jersey.Number))
colnames(fifa)[10] = "y"

# Split into train and test
set.seed(1)
train_index = sample(nrow(fifa), 0.8*nrow(fifa))
train = fifa[train_index,]
test = fifa[-train_index,]
lr1 = glm(y~., data = train, family = binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(lr1)
```

```
##
## Call:
## glm(formula = y ~ ., family = binomial, data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7723  -0.0049   0.0005   0.0248   2.7737
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         109.759833   4.998230  21.960  < 2e-16 ***
## Age                  -2.021733   0.060414 -33.465  < 2e-16 ***
## Value                 6.205362   0.455193  13.632  < 2e-16 ***
## Wage                 -0.420757   0.087387  -4.815 1.47e-06 ***
## Preferred.FootRight   0.291143   0.133388   2.183 0.029059 *
## PositionCB            3.007420   0.312299   9.630  < 2e-16 ***
## PositionCDM           1.946970   0.336268   5.790 7.04e-09 ***
## PositionCF            0.423955   0.967962   0.438 0.661395
## PositionCM            0.257785   0.302465   0.852 0.394058
## PositionGK            5.147473   0.337758  15.240  < 2e-16 ***
## PositionLAM          -0.486966   1.067122  -0.456 0.648148
## PositionLB            1.364818   0.303030   4.504 6.67e-06 ***
## PositionLCB           2.963183   0.357689   8.284  < 2e-16 ***
## PositionLCM           1.324593   0.385707   3.434 0.000594 ***
## PositionLDM           1.183742   0.445148   2.659 0.007832 **
```

```
## PositionLF            0.926443   2.363237   0.392 0.695041
## PositionLM           -1.060277   0.297348  -3.566 0.000363 ***
## PositionLS            0.190236   0.435396   0.437 0.662166
## PositionLW           -0.516516   0.436862  -1.182 0.237074
## PositionLWB           1.177586   0.752758   1.564 0.117732
## PositionRAM          -2.624120   0.950552  -2.761 0.005769 **
## PositionRB            1.113653   0.289899   3.842 0.000122 ***
## PositionRCB           2.905966   0.349056   8.325  < 2e-16 ***
## PositionRCM           0.051630   0.382897   0.135 0.892739
## PositionRDM           0.697252   0.414116   1.684 0.092238 .
## PositionRF           -0.686024   1.378139  -0.498 0.618632
## PositionRM           -0.973960   0.282531  -3.447 0.000566 ***
## PositionRS           -0.462808   0.452575  -1.023 0.306493
## PositionRW           -0.809267   0.426689  -1.897 0.057878 .
## PositionRWB           1.039209   0.758643   1.370 0.170741
## PositionST           -0.052606   0.266062  -0.198 0.843264
## Height                1.687862   0.423232   3.988 6.66e-05 ***
## Weight                0.005344   0.004123   1.296 0.194927
## Overall              -0.907692   0.072738 -12.479  < 2e-16 ***
## Contract.Duration     0.050814   0.022431   2.265 0.023488 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18304.7  on 14324  degrees of freedom
## Residual deviance:  2667.2  on 14290  degrees of freedom
## AIC: 2737.2
##
## Number of Fisher Scoring iterations: 9
```

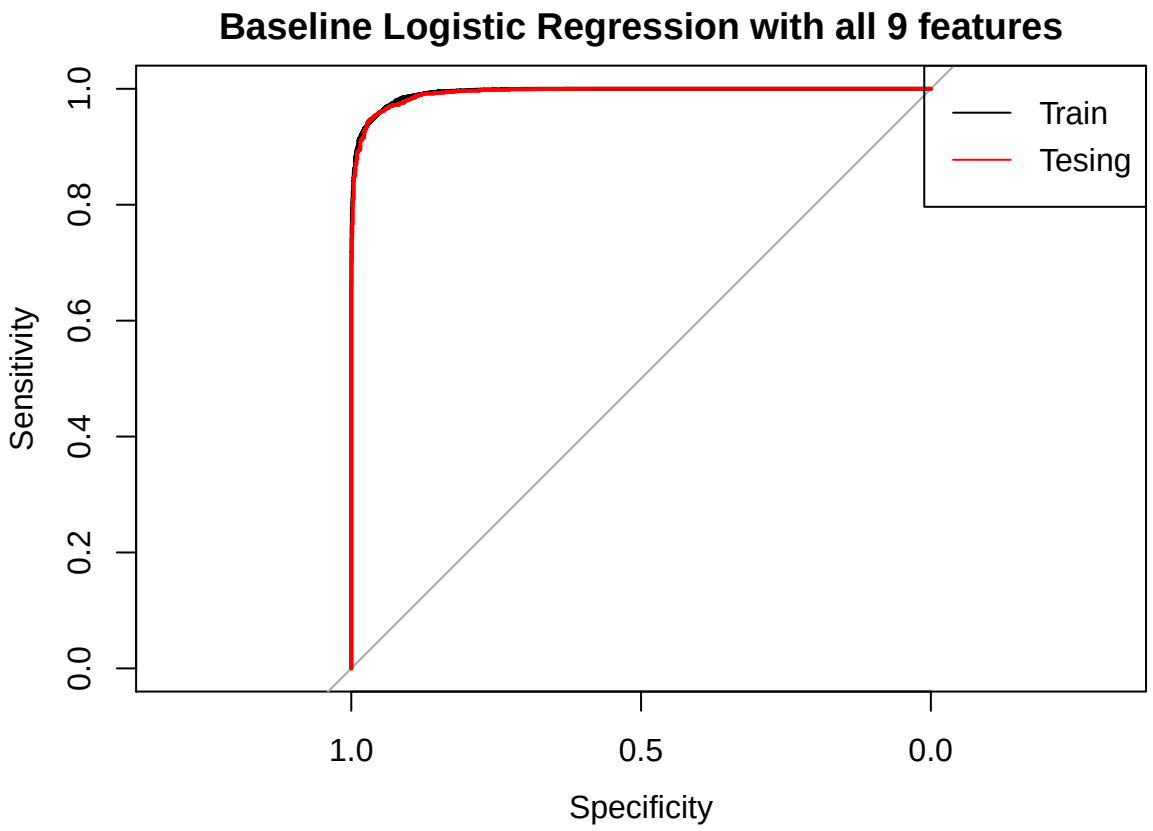
```
# Evaluate result
pred.lr1.train <- predict(lr1, type = "r")
pred.lr1.test <- predict(lr1, newdata = test, type = "r")

plot(roc(train$y, pred.lr1.train), main = "Baseline Logistic Regression with all 9 features")
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
lines(roc(test$y, pred.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)
```




```
auc(roc(train$y, pred.lr1.train))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.9941
```

```
auc(roc(test$y, pred.lr1.test))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.9932
```

```
pred.lr1.test.class = ifelse(pred.lr1.test>0.5,T,F)
table(pred.lr1.test.class, test$y)
```

```
##
## pred.lr1.test.class FALSE TRUE
##               FALSE  1133   74
##               TRUE     80 2295
```

```
cat("Best subset model accuracy is", mean(pred.lr1.test.class==test$y))
```

```
## Best subset model accuracy is 0.9570073
```

```
# Perform best subset feature selection

best_lr = bestglm(train, family = binomial, IC = "BIC", nvmax = length(train)-1)
```

```
## Morgan-Tatar search since family is non-gaussian.
```

```
## Note: factors present with more than 2 levels.
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

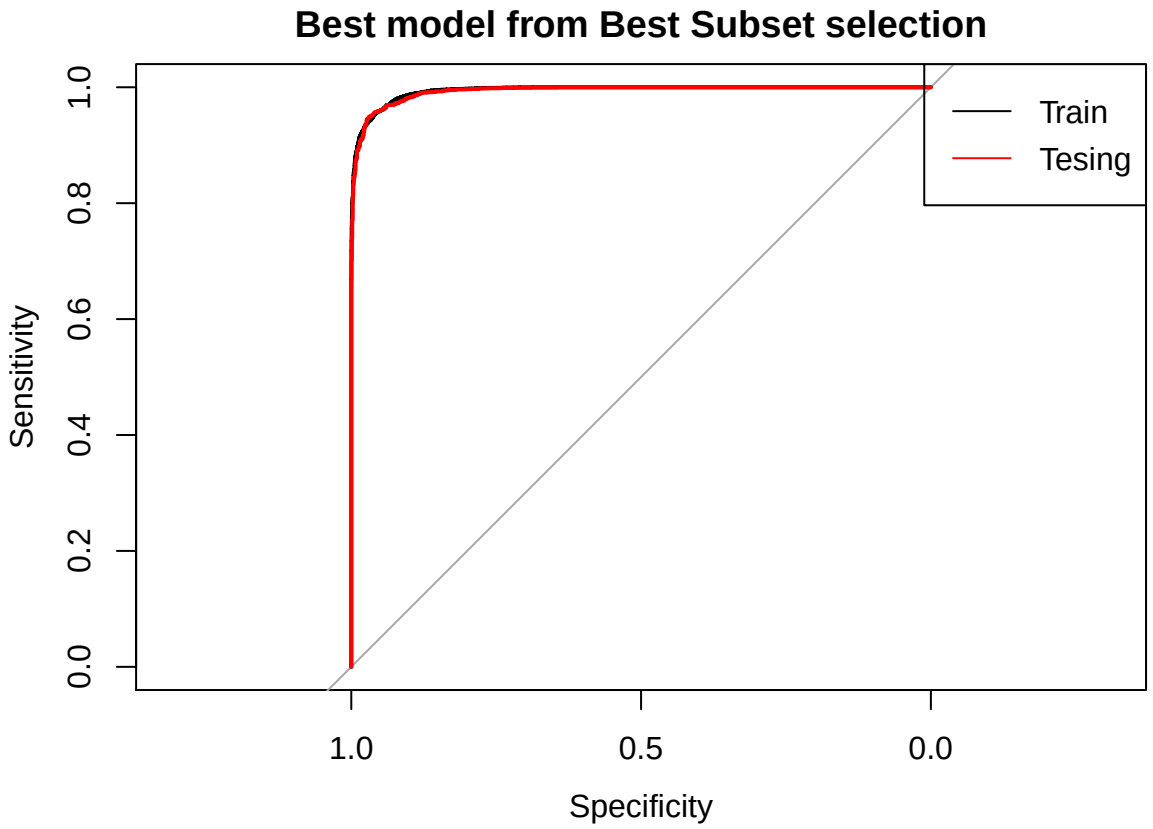
```
best_lr1 = best_lr$BestModel
best_lr1_sum = summary(best_lr1)

# Evaluate result
pred.best_lr.train <- predict(best_lr1, type = "r")
pred.best_lr.test <- predict(best_lr1, newdata = test, type = "r")

plot(roc(train$y, pred.best_lr.train), main = "Best model from Best Subset selection")
```

```
## Setting levels: control = FALSE, case = TRUE

## Setting direction: controls < cases
```

```
lines(roc(test$y, pred.best_lr.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)
```




```
auc(roc(train$y, pred.best_lr.train))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.994
```

```
auc(roc(test$y, pred.best_lr.test))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.9933
```

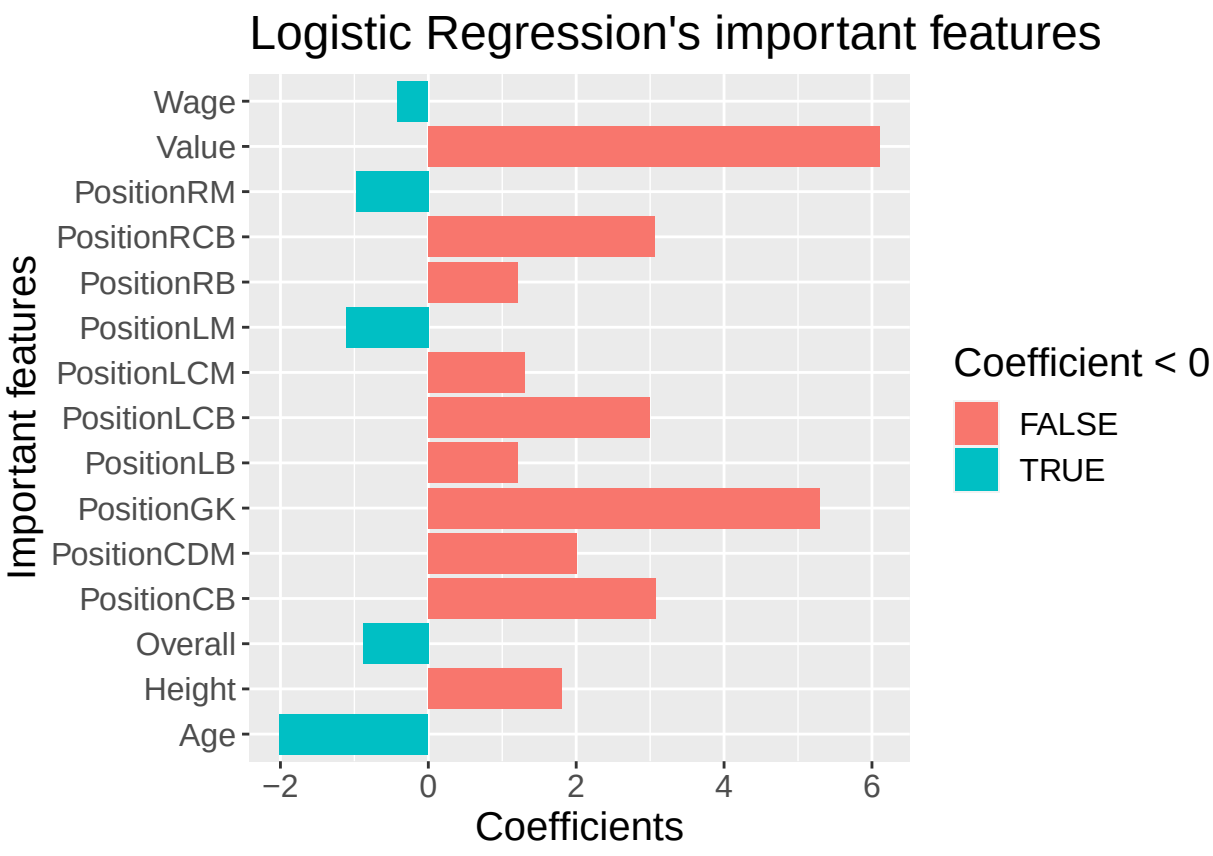
```
pred.best_lr.test.class = ifelse(pred.best_lr.test>0.5,T,F)
table(pred.best_lr.test.class, test$y)
```

```
##
## pred.best_lr.test.class FALSE TRUE
##                   FALSE  1130   73
##                   TRUE     83 2296
```

```
cat("Best subset model accuracy is", mean(pred.best_lr.test.class==test$y), "\n")
```

```
## Best subset model accuracy is 0.9564489
```

```r
# Visualize important features
best_lr1_coef = as.data.frame(best_lr1_sum$coefficients)
best_lr1_coef$features = rownames(best_lr1_coef)
best_lr1_coef = best_lr1_coef[-1,]
best_lr1_coef = best_lr1_coef[best_lr1_coef$`Pr(>|z|)`<0.001, ]

ggplot(data = best_lr1_coef,aes(y = Estimate,
                 x = features,
                 fill = Estimate < 0)) +
                 geom_col() +
                 coord_flip() +
                 ylab("Coefficients") +
                 xlab("Important features") +
                 labs(fill = "Coefficient < 0") +
                 theme(text = element_text(size=15)) +
                 ggtitle("Logistic Regression's important features")
```




```r
## Perform stepwise feature selection I
# Backward
back.bic.lr1 = bestglm(train, family = binomial, IC = "BIC", method = "backward", trace = F)
```

```
## Morgan-Tatar search since family is non-gaussian.

## Note: factors present with more than 2 levels.

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
back.bic.lr = best_lr$BestModel
summary(back.bic.lr)
```

```
##
## Call:
## glm(formula = y ~ ., family = family, data = Xi, weights = weights)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7568  -0.0050   0.0005   0.0255   2.7812
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 109.28141    4.95620  22.049  < 2e-16 ***
## Age          -2.01466    0.06012 -33.511  < 2e-16 ***
## Value         6.10600    0.45215  13.504  < 2e-16 ***
## Wage         -0.41486    0.08696  -4.771 1.84e-06 ***
## PositionCB    3.07649    0.30658  10.035  < 2e-16 ***
```

```
## PositionCDM   2.00648    0.33482   5.993 2.06e-09 ***
## PositionCF    0.51368    0.93705   0.548 0.583558
## PositionCM    0.29858    0.30128   0.991 0.321672
## PositionGK    5.29535    0.32883  16.104  < 2e-16 ***
## PositionLAM  -0.59587    1.04696  -0.569 0.569262
## PositionLB    1.20456    0.28960   4.159 3.19e-05 ***
## PositionLCB   2.99270    0.34881   8.580  < 2e-16 ***
## PositionLCM   1.30251    0.38576   3.376 0.000734 ***
## PositionLDM   1.18618    0.44384   2.673 0.007528 **
## PositionLF    0.95999    2.49215   0.385 0.700085
## PositionLM   -1.11160    0.29573  -3.759 0.000171 ***
## PositionLS    0.25288    0.43126   0.586 0.557626
## PositionLW   -0.52845    0.43764  -1.207 0.227245
## PositionLWB   1.06730    0.73757   1.447 0.147880
## PositionRAM  -2.72262    0.94615  -2.878 0.004007 **
## PositionRB    1.21167    0.28753   4.214 2.51e-05 ***
## PositionRCB   3.06726    0.34195   8.970  < 2e-16 ***
## PositionRCM   0.13446    0.37987   0.354 0.723368
## PositionRDM   0.81652    0.41286   1.978 0.047960 *
## PositionRF   -0.62091    1.35180  -0.459 0.646004
## PositionRM   -0.97990    0.28123  -3.484 0.000493 ***
## PositionRS   -0.38710    0.44879  -0.863 0.388392
## PositionRW   -0.83121    0.42711  -1.946 0.051639 .
## PositionRWB   1.10599    0.74739   1.480 0.138927
## PositionST    0.01820    0.26115   0.070 0.944445
## Height        1.80366    0.40337   4.471 7.77e-06 ***
## Overall      -0.88762    0.07205 -12.320  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18304.7  on 14324  degrees of freedom
## Residual deviance:  2679.3  on 14293  degrees of freedom
## AIC: 2743.3
##
## Number of Fisher Scoring iterations: 9
```

```
# Evaluate result
pred.back.bic.lr1.train <- predict(back.bic.lr, type = "r")
pred.back.bic.lr1.test <- predict(back.bic.lr, newdata = test, type = "r")

plot(roc(train$y, pred.back.bic.lr1.train), main = "Best model from Backward Stepwise selection")
```

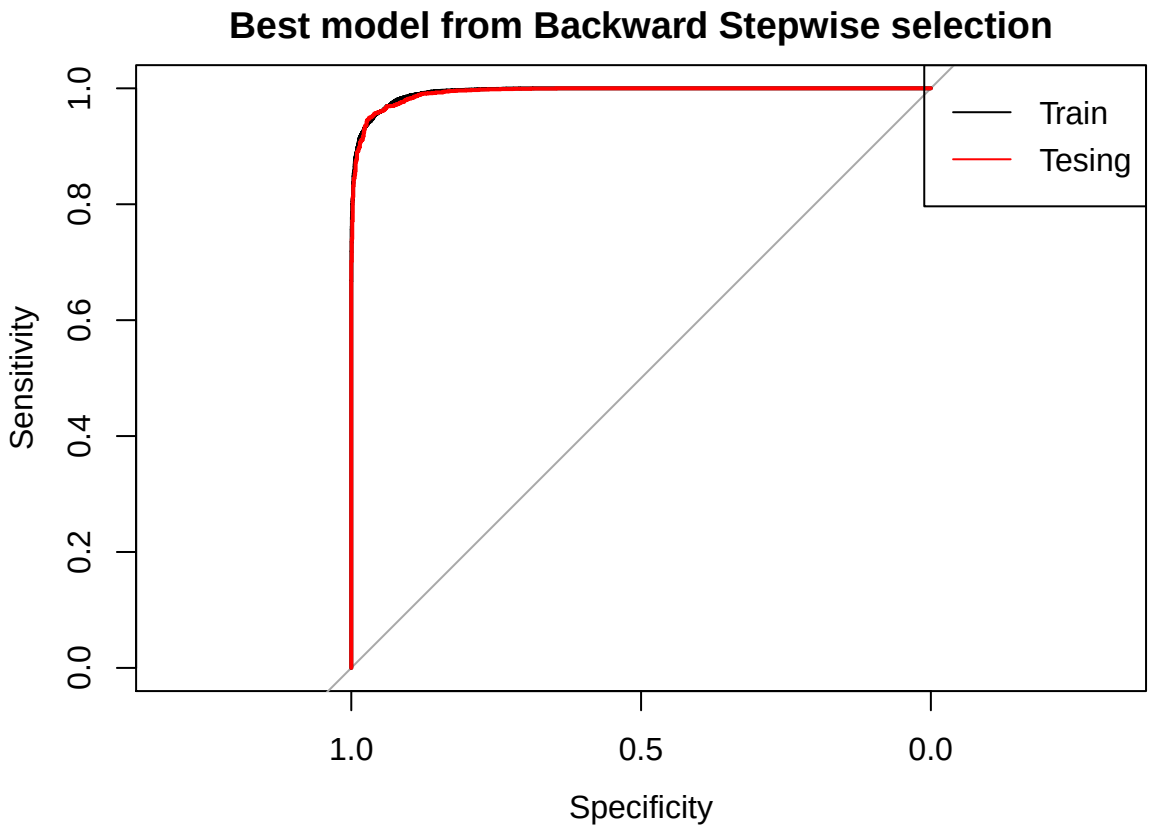
```
## Setting levels: control = FALSE, case = TRUE
```

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

```
lines(roc(test$y, pred.back.bic.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)
```




```
auc(roc(train$y, pred.back.bic.lr1.train))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.994
```

```
auc(roc(test$y, pred.back.bic.lr1.test))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.9933
```

```
pred.back.bic.lr1.test.class = ifelse(pred.back.bic.lr1.test>0.5,T,F)
table(pred.back.bic.lr1.test.class, test$y)
```

```
##
## pred.back.bic.lr1.test.class FALSE TRUE
##                        FALSE  1130   73
##                        TRUE     83 2296
```

```
cat("Backward stepwise AIC model accuracy is", mean(pred.back.bic.lr1.test.class==test$y))
```

```
## Backward stepwise AIC model accuracy is 0.9564489
```

```
## Perform stepwise feature selection II
# Forward AIC
for.bic.lr1 = bestglm(train, family = binomial, IC = "BIC", method = "forward")
```

```
## Morgan-Tatar search since family is non-gaussian.
## Note: factors present with more than 2 levels.
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

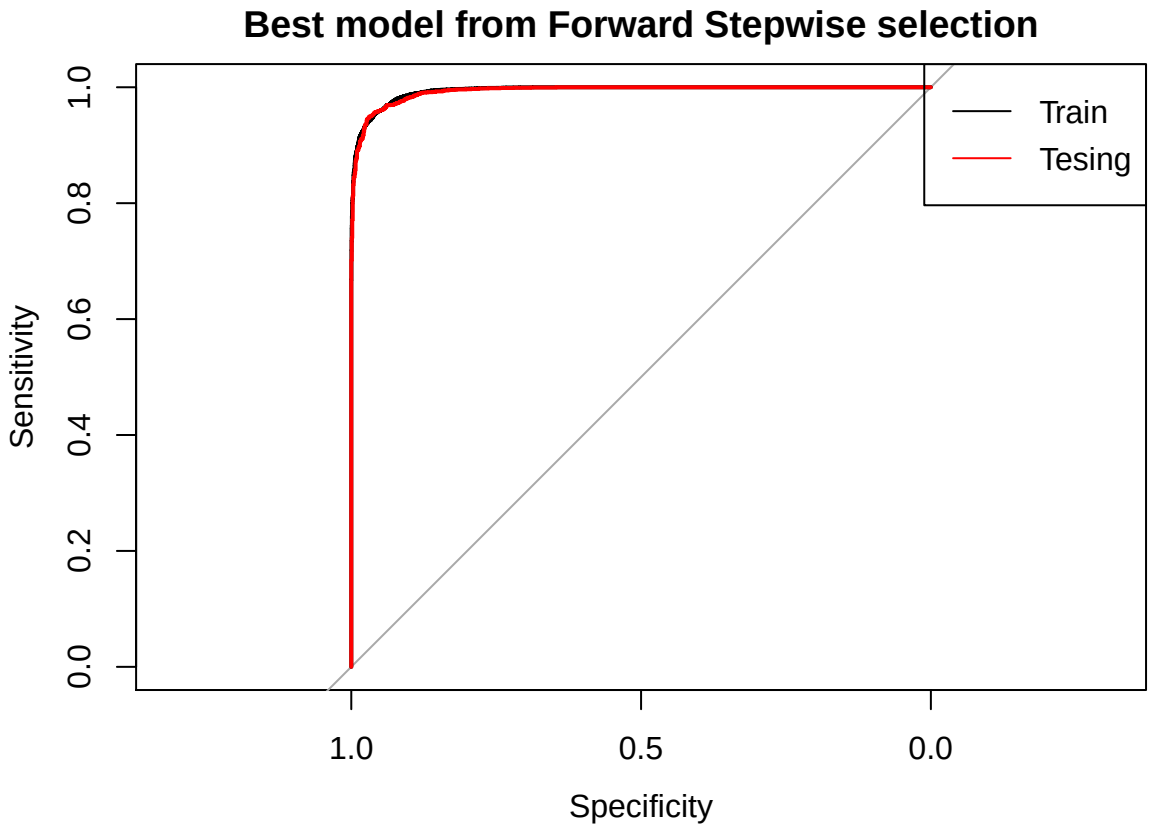
```r
for.bic.lr = for.bic.lr1$BestModel
for.bic.sum = summary(for.bic.lr)

# Evaluate result
pred.for.bic.lr1.train <- predict(for.bic.lr, type = "r")
pred.for.bic.lr1.test <- predict(for.bic.lr, newdata = test, type = "r")

plot(roc(train$y, pred.for.bic.lr1.train), main = "Best model from Forward Stepwise selection")
```

```
## Setting levels: control = FALSE, case = TRUE

## Setting direction: controls < cases
```

```
lines(roc(test$y, pred.for.bic.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)
```




```
auc(roc(train$y, pred.for.bic.lr1.train))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.994
```

```
auc(roc(test$y, pred.for.bic.lr1.test))
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Area under the curve: 0.9933
```

```
pred.for.bic.lr1.test.class = ifelse(pred.for.bic.lr1.test>0.5,T,F)
table(pred.for.bic.lr1.test.class, test$y)
```

```
##
## pred.for.bic.lr1.test.class FALSE TRUE
##                       FALSE  1130   73
##                       TRUE     83 2296
```

```
cat("Forward stepwise AIC model accuracy is", mean(pred.for.bic.lr1.test.class==test$y))
```

```
## Forward stepwise AIC model accuracy is 0.9564489
```

```r
# Visuals for report
par(mfrow = c(2,2))
# Baseline logistic regression
plot(roc(train$y, pred.lr1.train), main = "Baseline Logistic Regression")
```

```
## Setting levels: control = FALSE, case = TRUE
```

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

```r
lines(roc(test$y, pred.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```r
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)

# Best subset
plot(roc(train$y, pred.best_lr.train), main = "Best Subset Selection")
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```r
lines(roc(test$y, pred.best_lr.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```r
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)

# Forward
plot(roc(train$y, pred.for.bic.lr1.train), main = "Forward Stepwise Selection")
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```r
lines(roc(test$y, pred.for.bic.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

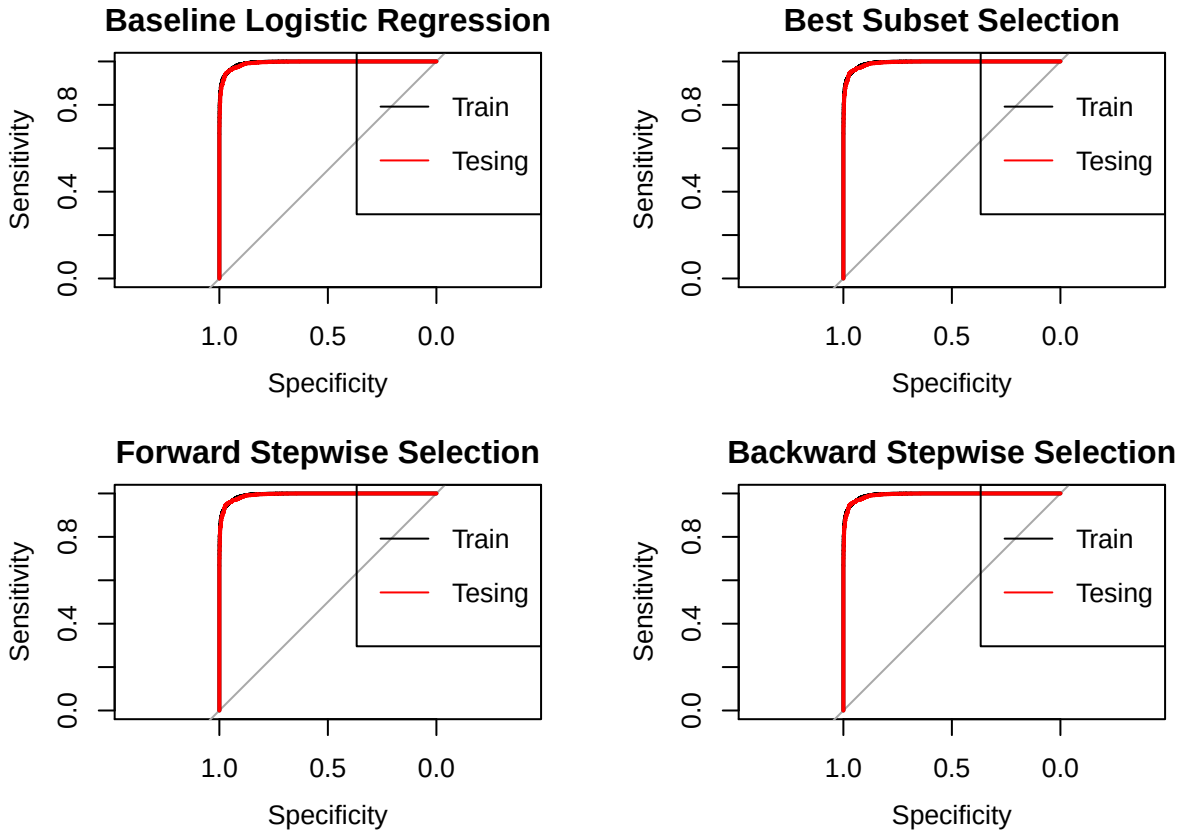
```r
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)

# Backward
plot(roc(train$y, pred.back.bic.lr1.train), main = "Backward Stepwise Selection")
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```r
lines(roc(test$y, pred.back.bic.lr1.test), col = 2)
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
legend("topright", c("Train", "Tesing"), col=c("black", "red"), cex=1, lty=1)
```




```
stopCluster(cl)
```