Feature Selection Comparison

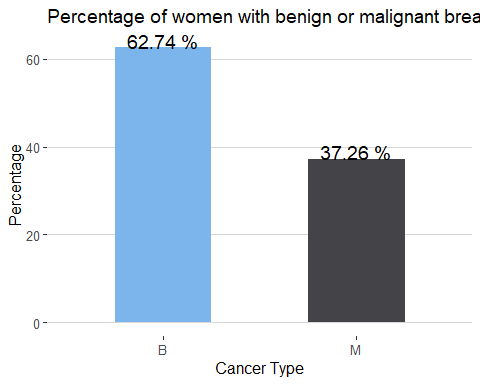
9/22/2020

library(tidyverse)  
library(data.table)  
library(knitr)  
library(caret)  
library(glmnet)  
library(ggthemes)  
  
cancer <- fread("data.csv")  
  
cancer[, V33 := NULL]  
cancer[, diagnosis := factor(diagnosis)]  
nms <- names(cancer)  
nms <- gsub(" ", "\_", nms)  
names(cancer) <- nms  
str(cancer)

## Classes 'data.table' and 'data.frame': 569 obs. of 32 variables:  
## $ id : int 842302 842517 84300903 84348301 84358402 843786 844359 84458202 844981 84501001 ...  
## $ diagnosis : Factor w/ 2 levels "B","M": 2 2 2 2 2 2 2 2 2 2 ...  
## $ radius\_mean : num 18 20.6 19.7 11.4 20.3 ...  
## $ texture\_mean : num 10.4 17.8 21.2 20.4 14.3 ...  
## $ perimeter\_mean : num 122.8 132.9 130 77.6 135.1 ...  
## $ area\_mean : num 1001 1326 1203 386 1297 ...  
## $ smoothness\_mean : num 0.1184 0.0847 0.1096 0.1425 0.1003 ...  
## $ compactness\_mean : num 0.2776 0.0786 0.1599 0.2839 0.1328 ...  
## $ concavity\_mean : num 0.3001 0.0869 0.1974 0.2414 0.198 ...  
## $ concave\_points\_mean : num 0.1471 0.0702 0.1279 0.1052 0.1043 ...  
## $ symmetry\_mean : num 0.242 0.181 0.207 0.26 0.181 ...  
## $ fractal\_dimension\_mean : num 0.0787 0.0567 0.06 0.0974 0.0588 ...  
## $ radius\_se : num 1.095 0.543 0.746 0.496 0.757 ...  
## $ texture\_se : num 0.905 0.734 0.787 1.156 0.781 ...  
## $ perimeter\_se : num 8.59 3.4 4.58 3.44 5.44 ...  
## $ area\_se : num 153.4 74.1 94 27.2 94.4 ...  
## $ smoothness\_se : num 0.0064 0.00522 0.00615 0.00911 0.01149 ...  
## $ compactness\_se : num 0.049 0.0131 0.0401 0.0746 0.0246 ...  
## $ concavity\_se : num 0.0537 0.0186 0.0383 0.0566 0.0569 ...  
## $ concave\_points\_se : num 0.0159 0.0134 0.0206 0.0187 0.0188 ...  
## $ symmetry\_se : num 0.03 0.0139 0.0225 0.0596 0.0176 ...  
## $ fractal\_dimension\_se : num 0.00619 0.00353 0.00457 0.00921 0.00511 ...  
## $ radius\_worst : num 25.4 25 23.6 14.9 22.5 ...  
## $ texture\_worst : num 17.3 23.4 25.5 26.5 16.7 ...  
## $ perimeter\_worst : num 184.6 158.8 152.5 98.9 152.2 ...  
## $ area\_worst : num 2019 1956 1709 568 1575 ...  
## $ smoothness\_worst : num 0.162 0.124 0.144 0.21 0.137 ...  
## $ compactness\_worst : num 0.666 0.187 0.424 0.866 0.205 ...  
## $ concavity\_worst : num 0.712 0.242 0.45 0.687 0.4 ...  
## $ concave\_points\_worst : num 0.265 0.186 0.243 0.258 0.163 ...  
## $ symmetry\_worst : num 0.46 0.275 0.361 0.664 0.236 ...  
## $ fractal\_dimension\_worst: num 0.1189 0.089 0.0876 0.173 0.0768 ...  
## - attr(\*, ".internal.selfref")=<externalptr>

cancer[, id := NULL]

cancer[, .(freq = .N),  
 by = diagnosis] %>%   
 .[, perc := round(100 \* freq/sum(freq), 2)] %>%  
   
ggplot(aes(x=diagnosis, y=perc, fill = diagnosis)) +   
 geom\_bar(stat = "identity", width = 0.5)+ theme\_hc() +  
 geom\_text(aes(x=diagnosis, y=perc, label = paste(perc, "%")),  
 position = position\_dodge(width = 0.5),  
 vjust = 0.05, hjust = 0.5, size = 5)+  
 scale\_fill\_hc(name = "")+  
 labs(x = "Cancer Type",  
 y = "Percentage",  
 title = "Percentage of women with benign or malignant breast bancer")+  
 theme(legend.position = "none",  
 axis.title = element\_text(size =12))



## Test train

set.seed(100)  
train\_sample <- sample(1:nrow(cancer), round(0.7\*nrow(cancer)))  
train\_set <- cancer[train\_sample,]  
test\_set <- cancer[-train\_sample,]

## Fit model

library(broom)  
glm\_mod <- glm(diagnosis ~ .,  
 data = train\_set,   
 family = binomial())  
  
  
tidy(glm\_mod) %>% kable

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | 1.051554e+02 | 4.009638e+05 | 0.0002623 | 0.9997907 |
| radius\_mean | -9.903775e+02 | 1.560410e+05 | -0.0063469 | 0.9949359 |
| texture\_mean | 1.145513e+01 | 3.530283e+03 | 0.0032448 | 0.9974110 |
| perimeter\_mean | 9.743985e+01 | 2.559560e+04 | 0.0038069 | 0.9969625 |
| area\_mean | 2.585920e+00 | 5.745781e+02 | 0.0045006 | 0.9964091 |
| smoothness\_mean | 2.947579e+03 | 1.035082e+06 | 0.0028477 | 0.9977279 |
| compactness\_mean | -8.526539e+03 | 8.904460e+05 | -0.0095756 | 0.9923599 |
| concavity\_mean | 2.219474e+03 | 4.943034e+05 | 0.0044901 | 0.9964174 |
| concave\_points\_mean | 1.138650e+04 | 9.282659e+05 | 0.0122664 | 0.9902131 |
| symmetry\_mean | -2.836263e+03 | 2.343713e+05 | -0.0121016 | 0.9903446 |
| fractal\_dimension\_mean | -3.493666e+03 | 1.652806e+06 | -0.0021138 | 0.9983135 |
| radius\_se | -1.635119e+03 | 5.034541e+05 | -0.0032478 | 0.9974086 |
| texture\_se | -1.992183e+01 | 2.387207e+04 | -0.0008345 | 0.9993341 |
| perimeter\_se | 5.366880e+01 | 2.473653e+04 | 0.0021696 | 0.9982689 |
| area\_se | 2.032495e+01 | 3.870656e+03 | 0.0052510 | 0.9958103 |
| smoothness\_se | -2.378366e+04 | 2.990972e+06 | -0.0079518 | 0.9936554 |
| compactness\_se | 1.631593e+04 | 3.105827e+06 | 0.0052533 | 0.9958085 |
| concavity\_se | -6.128921e+03 | 5.893287e+05 | -0.0103998 | 0.9917023 |
| concave\_points\_se | 3.931166e+04 | 2.950471e+06 | 0.0133239 | 0.9893694 |
| symmetry\_se | -2.073166e+04 | 2.814956e+06 | -0.0073648 | 0.9941238 |
| fractal\_dimension\_se | -1.088166e+05 | 1.888605e+07 | -0.0057617 | 0.9954028 |
| radius\_worst | 3.877885e+02 | 4.104385e+04 | 0.0094482 | 0.9924616 |
| texture\_worst | 1.384641e+00 | 3.685304e+03 | 0.0003757 | 0.9997002 |
| perimeter\_worst | -2.947268e+01 | 5.744130e+03 | -0.0051309 | 0.9959061 |
| area\_worst | -1.043073e+00 | 3.184540e+02 | -0.0032754 | 0.9973866 |
| smoothness\_worst | -1.177138e+03 | 4.179253e+05 | -0.0028166 | 0.9977527 |
| compactness\_worst | -1.178895e+03 | 2.761519e+05 | -0.0042690 | 0.9965938 |
| concavity\_worst | 3.930469e+02 | 1.677193e+05 | 0.0023435 | 0.9981302 |
| concave\_points\_worst | -1.419982e+03 | 5.219040e+05 | -0.0027208 | 0.9978291 |
| symmetry\_worst | 3.534347e+03 | 2.750049e+05 | 0.0128519 | 0.9897459 |
| fractal\_dimension\_worst | 1.190558e+04 | 1.485167e+06 | 0.0080163 | 0.9936040 |

## Forward

forward\_select <- step(glm\_mod, direction = "forward")

## Start: AIC=62  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + radius\_se + texture\_se +   
## perimeter\_se + area\_se + smoothness\_se + compactness\_se +   
## concavity\_se + concave\_points\_se + symmetry\_se + fractal\_dimension\_se +   
## radius\_worst + texture\_worst + perimeter\_worst + area\_worst +   
## smoothness\_worst + compactness\_worst + concavity\_worst +   
## concave\_points\_worst + symmetry\_worst + fractal\_dimension\_worst

## Backward

back\_select <- step(glm\_mod, direction = "backward")

## Start: AIC=62  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + radius\_se + texture\_se +   
## perimeter\_se + area\_se + smoothness\_se + compactness\_se +   
## concavity\_se + concave\_points\_se + symmetry\_se + fractal\_dimension\_se +   
## radius\_worst + texture\_worst + perimeter\_worst + area\_worst +   
## smoothness\_worst + compactness\_worst + concavity\_worst +   
## concave\_points\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - texture\_worst 1 0.00 60.00  
## - radius\_se 1 0.00 60.00  
## - perimeter\_se 1 0.00 60.00  
## - texture\_se 1 0.00 60.00  
## - concave\_points\_worst 1 0.00 60.00  
## - concavity\_worst 1 0.00 60.00  
## - smoothness\_worst 1 0.00 60.00  
## - smoothness\_mean 1 0.00 60.00  
## - fractal\_dimension\_mean 1 0.00 60.00  
## - area\_worst 1 0.00 60.00  
## - area\_se 1 0.00 60.00  
## - texture\_mean 1 0.00 60.00  
## - compactness\_worst 1 0.00 60.00  
## - area\_mean 1 0.00 60.00  
## - radius\_worst 1 0.00 60.00  
## - perimeter\_worst 1 0.00 60.00  
## - concavity\_mean 1 0.00 60.00  
## - smoothness\_se 1 0.00 60.00  
## - perimeter\_mean 1 0.00 60.00  
## - radius\_mean 1 0.00 60.00  
## - compactness\_se 1 0.00 60.00  
## - concavity\_se 1 0.00 60.00  
## - symmetry\_mean 1 0.00 60.00  
## - concave\_points\_se 1 0.00 60.00  
## - symmetry\_se 1 0.00 60.00  
## - compactness\_mean 1 0.00 60.00  
## - fractal\_dimension\_se 1 0.00 60.00  
## - symmetry\_worst 1 0.00 60.00  
## <none> 0.00 62.00  
## - concave\_points\_mean 1 432.52 492.52  
## - fractal\_dimension\_worst 1 865.05 925.05  
##   
## Step: AIC=60  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + radius\_se + texture\_se +   
## perimeter\_se + area\_se + smoothness\_se + compactness\_se +   
## concavity\_se + concave\_points\_se + symmetry\_se + fractal\_dimension\_se +   
## radius\_worst + perimeter\_worst + area\_worst + smoothness\_worst +   
## compactness\_worst + concavity\_worst + concave\_points\_worst +   
## symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - texture\_se 1 0.00 58.00  
## - area\_worst 1 0.00 58.00  
## - radius\_se 1 0.00 58.00  
## - perimeter\_se 1 0.00 58.00  
## - concavity\_worst 1 0.00 58.00  
## - smoothness\_worst 1 0.00 58.00  
## - fractal\_dimension\_mean 1 0.00 58.00  
## - concave\_points\_worst 1 0.00 58.00  
## - smoothness\_mean 1 0.00 58.00  
## - compactness\_worst 1 0.00 58.00  
## - area\_mean 1 0.00 58.00  
## - concavity\_mean 1 0.00 58.00  
## - perimeter\_worst 1 0.00 58.00  
## - area\_se 1 0.00 58.00  
## - perimeter\_mean 1 0.00 58.00  
## - radius\_worst 1 0.00 58.00  
## - radius\_mean 1 0.00 58.00  
## - compactness\_se 1 0.00 58.00  
## - concavity\_se 1 0.00 58.00  
## - fractal\_dimension\_worst 1 0.00 58.00  
## - concave\_points\_se 1 0.00 58.00  
## - smoothness\_se 1 0.00 58.00  
## - symmetry\_mean 1 0.00 58.00  
## - texture\_mean 1 0.00 58.00  
## - symmetry\_worst 1 0.00 58.00  
## - fractal\_dimension\_se 1 0.00 58.00  
## - compactness\_mean 1 1.51 59.51  
## <none> 0.00 60.00  
## - concave\_points\_mean 1 792.96 850.96  
## - symmetry\_se 1 865.05 923.05  
##   
## Step: AIC=58  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + radius\_se + perimeter\_se +   
## area\_se + smoothness\_se + compactness\_se + concavity\_se +   
## concave\_points\_se + symmetry\_se + fractal\_dimension\_se +   
## radius\_worst + perimeter\_worst + area\_worst + smoothness\_worst +   
## compactness\_worst + concavity\_worst + concave\_points\_worst +   
## symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - radius\_se 1 0.00 56.00  
## - area\_worst 1 0.00 56.00  
## - concavity\_worst 1 0.00 56.00  
## - perimeter\_se 1 0.00 56.00  
## - smoothness\_worst 1 0.00 56.00  
## - fractal\_dimension\_mean 1 0.00 56.00  
## - concave\_points\_worst 1 0.00 56.00  
## - smoothness\_mean 1 0.00 56.00  
## - compactness\_worst 1 0.00 56.00  
## - concavity\_mean 1 0.00 56.00  
## - perimeter\_worst 1 0.00 56.00  
## - area\_mean 1 0.00 56.00  
## - radius\_worst 1 0.00 56.00  
## - area\_se 1 0.00 56.00  
## - perimeter\_mean 1 0.00 56.00  
## - compactness\_se 1 0.00 56.00  
## - radius\_mean 1 0.00 56.00  
## - smoothness\_se 1 0.00 56.00  
## - concavity\_se 1 0.00 56.00  
## - concave\_points\_se 1 0.00 56.00  
## - fractal\_dimension\_worst 1 0.00 56.00  
## - compactness\_mean 1 0.00 56.00  
## - symmetry\_worst 1 0.00 56.00  
## <none> 0.00 58.00  
## - texture\_mean 1 27.05 83.05  
## - symmetry\_mean 1 648.79 704.79  
## - concave\_points\_mean 1 792.96 848.96  
## - fractal\_dimension\_se 1 792.96 848.96  
## - symmetry\_se 1 937.13 993.13  
##   
## Step: AIC=56  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + perimeter\_se + area\_se +   
## smoothness\_se + compactness\_se + concavity\_se + concave\_points\_se +   
## symmetry\_se + fractal\_dimension\_se + radius\_worst + perimeter\_worst +   
## area\_worst + smoothness\_worst + compactness\_worst + concavity\_worst +   
## concave\_points\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - area\_worst 1 0.00 54.00  
## - smoothness\_mean 1 0.00 54.00  
## - smoothness\_worst 1 0.00 54.00  
## - concave\_points\_worst 1 0.00 54.00  
## - fractal\_dimension\_mean 1 0.00 54.00  
## - perimeter\_se 1 0.00 54.00  
## - concavity\_worst 1 0.00 54.00  
## - area\_mean 1 0.00 54.00  
## - concavity\_mean 1 0.00 54.00  
## - perimeter\_worst 1 0.00 54.00  
## - radius\_worst 1 0.00 54.00  
## - area\_se 1 0.00 54.00  
## - perimeter\_mean 1 0.00 54.00  
## - compactness\_worst 1 0.00 54.00  
## - radius\_mean 1 0.00 54.00  
## - concave\_points\_se 1 0.00 54.00  
## - concavity\_se 1 0.00 54.00  
## - compactness\_se 1 0.00 54.00  
## - symmetry\_mean 1 0.00 54.00  
## - fractal\_dimension\_worst 1 0.00 54.00  
## - compactness\_mean 1 0.00 54.00  
## - symmetry\_se 1 0.00 54.00  
## <none> 0.00 56.00  
## - symmetry\_worst 1 31.57 85.57  
## - texture\_mean 1 34.43 88.43  
## - fractal\_dimension\_se 1 792.96 846.96  
## - concave\_points\_mean 1 865.05 919.05  
## - smoothness\_se 1 865.05 919.05  
##   
## Step: AIC=54  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + perimeter\_se + area\_se +   
## smoothness\_se + compactness\_se + concavity\_se + concave\_points\_se +   
## symmetry\_se + fractal\_dimension\_se + radius\_worst + perimeter\_worst +   
## smoothness\_worst + compactness\_worst + concavity\_worst +   
## concave\_points\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - perimeter\_se 1 0.00 52.00  
## - smoothness\_mean 1 0.00 52.00  
## - smoothness\_worst 1 0.00 52.00  
## - concave\_points\_worst 1 0.00 52.00  
## - fractal\_dimension\_mean 1 0.00 52.00  
## - area\_mean 1 0.00 52.00  
## - concavity\_mean 1 0.00 52.00  
## - perimeter\_worst 1 0.00 52.00  
## - concavity\_worst 1 0.00 52.00  
## - area\_se 1 0.00 52.00  
## - perimeter\_mean 1 0.00 52.00  
## - compactness\_worst 1 0.00 52.00  
## - radius\_worst 1 0.00 52.00  
## - radius\_mean 1 0.00 52.00  
## - concave\_points\_se 1 0.00 52.00  
## - compactness\_se 1 0.00 52.00  
## - concavity\_se 1 0.00 52.00  
## - fractal\_dimension\_worst 1 0.00 52.00  
## - symmetry\_mean 1 0.00 52.00  
## - fractal\_dimension\_se 1 0.00 52.00  
## - compactness\_mean 1 0.00 52.00  
## - symmetry\_se 1 0.00 52.00  
## <none> 0.00 54.00  
## - symmetry\_worst 1 31.84 83.84  
## - texture\_mean 1 39.46 91.46  
## - concave\_points\_mean 1 792.96 844.96  
## - smoothness\_se 1 865.05 917.05  
##   
## Step: AIC=52  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + area\_se + smoothness\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + smoothness\_worst +   
## compactness\_worst + concavity\_worst + concave\_points\_worst +   
## symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - concave\_points\_worst 1 0.00 50.00  
## - smoothness\_worst 1 0.00 50.00  
## - area\_mean 1 0.00 50.00  
## - fractal\_dimension\_mean 1 0.00 50.00  
## - concavity\_worst 1 0.00 50.00  
## - smoothness\_mean 1 0.00 50.00  
## - concavity\_mean 1 0.00 50.00  
## - area\_se 1 0.00 50.00  
## - smoothness\_se 1 0.00 50.00  
## - perimeter\_mean 1 0.00 50.00  
## - concave\_points\_se 1 0.00 50.00  
## - perimeter\_worst 1 0.00 50.00  
## - compactness\_worst 1 0.00 50.00  
## - concavity\_se 1 0.00 50.00  
## - radius\_mean 1 0.00 50.00  
## - concave\_points\_mean 1 0.00 50.00  
## <none> 0.00 52.00  
## - fractal\_dimension\_worst 1 28.26 78.26  
## - compactness\_se 1 28.98 78.98  
## - compactness\_mean 1 29.20 79.20  
## - fractal\_dimension\_se 1 33.47 83.47  
## - symmetry\_se 1 33.96 83.96  
## - symmetry\_worst 1 35.47 85.47  
## - radius\_worst 1 36.30 86.30  
## - texture\_mean 1 39.89 89.89  
## - symmetry\_mean 1 648.79 698.79  
##   
## Step: AIC=50  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + area\_se + smoothness\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + smoothness\_worst +   
## compactness\_worst + concavity\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - smoothness\_worst 1 0.00 48.00  
## - fractal\_dimension\_mean 1 0.00 48.00  
## - area\_mean 1 0.00 48.00  
## - concavity\_worst 1 0.00 48.00  
## - smoothness\_mean 1 0.00 48.00  
## - concavity\_mean 1 0.00 48.00  
## - smoothness\_se 1 0.00 48.00  
## - perimeter\_mean 1 0.00 48.00  
## - perimeter\_worst 1 0.00 48.00  
## - area\_se 1 0.00 48.00  
## - compactness\_worst 1 0.00 48.00  
## - concavity\_se 1 0.00 48.00  
## - radius\_mean 1 0.00 48.00  
## <none> 0.00 50.00  
## - symmetry\_mean 1 21.77 69.77  
## - fractal\_dimension\_worst 1 28.34 76.34  
## - compactness\_mean 1 30.82 78.82  
## - compactness\_se 1 31.00 79.00  
## - concave\_points\_se 1 32.09 80.09  
## - fractal\_dimension\_se 1 33.63 81.63  
## - symmetry\_se 1 34.73 82.73  
## - symmetry\_worst 1 35.59 83.59  
## - radius\_worst 1 36.48 84.48  
## - texture\_mean 1 40.72 88.72  
## - concave\_points\_mean 1 720.87 768.87  
##   
## Step: AIC=48  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + area\_mean +   
## smoothness\_mean + compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + area\_se + smoothness\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + compactness\_worst +   
## concavity\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - area\_mean 1 0.00 46.00  
## - concavity\_worst 1 0.00 46.00  
## - smoothness\_mean 1 0.00 46.00  
## - fractal\_dimension\_mean 1 0.00 46.00  
## - concavity\_mean 1 0.00 46.00  
## - perimeter\_mean 1 0.00 46.00  
## - compactness\_worst 1 0.00 46.00  
## - smoothness\_se 1 0.00 46.00  
## - concavity\_se 1 0.00 46.00  
## - perimeter\_worst 1 0.00 46.00  
## - radius\_mean 1 0.00 46.00  
## <none> 0.00 48.00  
## - symmetry\_mean 1 23.14 69.14  
## - fractal\_dimension\_worst 1 31.04 77.04  
## - compactness\_se 1 31.38 77.38  
## - compactness\_mean 1 31.60 77.60  
## - concave\_points\_se 1 32.61 78.61  
## - fractal\_dimension\_se 1 33.65 79.65  
## - symmetry\_se 1 34.81 80.81  
## - symmetry\_worst 1 35.63 81.63  
## - radius\_worst 1 37.03 83.03  
## - texture\_mean 1 40.73 86.73  
## - concave\_points\_mean 1 792.96 838.96  
## - area\_se 1 2667.23 2713.23  
##   
## Step: AIC=46  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + smoothness\_mean +   
## compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + area\_se + smoothness\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + compactness\_worst +   
## concavity\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - concavity\_worst 1 0.000 44.000  
## - smoothness\_mean 1 0.000 44.000  
## - fractal\_dimension\_mean 1 0.000 44.000  
## - concavity\_mean 1 0.000 44.000  
## - smoothness\_se 1 0.000 44.000  
## - compactness\_worst 1 0.000 44.000  
## - perimeter\_mean 1 0.000 44.000  
## - perimeter\_worst 1 0.000 44.000  
## - radius\_mean 1 0.000 44.000  
## <none> 0.000 46.000  
## - symmetry\_mean 1 23.136 67.136  
## - concavity\_se 1 23.587 67.587  
## - concave\_points\_mean 1 24.739 68.739  
## - area\_se 1 28.439 72.439  
## - fractal\_dimension\_worst 1 31.461 75.461  
## - compactness\_se 1 31.977 75.977  
## - concave\_points\_se 1 32.711 76.711  
## - symmetry\_se 1 35.155 79.155  
## - fractal\_dimension\_se 1 35.470 79.470  
## - compactness\_mean 1 36.154 80.154  
## - symmetry\_worst 1 36.228 80.228  
## - radius\_worst 1 37.031 81.031  
## - texture\_mean 1 40.765 84.765  
##   
## Step: AIC=44  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + smoothness\_mean +   
## compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + fractal\_dimension\_mean + area\_se + smoothness\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + compactness\_worst +   
## symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - fractal\_dimension\_mean 1 0.000 42.000  
## - smoothness\_mean 1 0.000 42.000  
## - smoothness\_se 1 0.000 42.000  
## - concavity\_mean 1 0.000 42.000  
## - perimeter\_mean 1 0.000 42.000  
## - perimeter\_worst 1 0.000 42.000  
## <none> 0.000 44.000  
## - compactness\_worst 1 18.355 60.355  
## - radius\_mean 1 20.575 62.575  
## - symmetry\_mean 1 23.267 65.267  
## - concavity\_se 1 23.938 65.938  
## - concave\_points\_mean 1 25.335 67.335  
## - area\_se 1 28.560 70.560  
## - fractal\_dimension\_worst 1 31.525 73.525  
## - compactness\_se 1 32.295 74.295  
## - concave\_points\_se 1 33.771 75.771  
## - symmetry\_se 1 35.317 77.317  
## - fractal\_dimension\_se 1 35.482 77.482  
## - symmetry\_worst 1 36.256 78.256  
## - radius\_worst 1 37.031 79.031  
## - compactness\_mean 1 37.440 79.440  
## - texture\_mean 1 42.140 84.140  
##   
## Step: AIC=42  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + smoothness\_mean +   
## compactness\_mean + concavity\_mean + concave\_points\_mean +   
## symmetry\_mean + area\_se + smoothness\_se + compactness\_se +   
## concavity\_se + concave\_points\_se + symmetry\_se + fractal\_dimension\_se +   
## radius\_worst + perimeter\_worst + compactness\_worst + symmetry\_worst +   
## fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - smoothness\_mean 1 0.000 40.000  
## - smoothness\_se 1 0.000 40.000  
## - concavity\_mean 1 0.000 40.000  
## - perimeter\_mean 1 0.000 40.000  
## - perimeter\_worst 1 0.000 40.000  
## <none> 0.000 42.000  
## - compactness\_worst 1 19.408 59.408  
## - radius\_mean 1 20.603 60.603  
## - concave\_points\_mean 1 25.404 65.404  
## - symmetry\_mean 1 26.370 66.370  
## - concavity\_se 1 26.380 66.380  
## - area\_se 1 29.967 69.967  
## - fractal\_dimension\_worst 1 32.004 72.004  
## - compactness\_se 1 32.505 72.505  
## - concave\_points\_se 1 33.882 73.882  
## - symmetry\_se 1 35.439 75.439  
## - fractal\_dimension\_se 1 36.176 76.176  
## - symmetry\_worst 1 36.796 76.796  
## - radius\_worst 1 37.234 77.234  
## - compactness\_mean 1 39.671 79.671  
## - texture\_mean 1 42.329 82.329  
##   
## Step: AIC=40  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + compactness\_mean +   
## concavity\_mean + concave\_points\_mean + symmetry\_mean + area\_se +   
## smoothness\_se + compactness\_se + concavity\_se + concave\_points\_se +   
## symmetry\_se + fractal\_dimension\_se + radius\_worst + perimeter\_worst +   
## compactness\_worst + symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## - smoothness\_se 1 0.000 38.000  
## - concavity\_mean 1 0.000 38.000  
## - perimeter\_worst 1 0.000 38.000  
## <none> 0.000 40.000  
## - compactness\_worst 1 21.007 59.007  
## - perimeter\_mean 1 21.594 59.594  
## - radius\_mean 1 26.202 64.202  
## - concavity\_se 1 26.386 64.386  
## - symmetry\_mean 1 27.731 65.731  
## - compactness\_se 1 33.258 71.258  
## - fractal\_dimension\_worst 1 33.536 71.536  
## - concave\_points\_se 1 33.946 71.946  
## - symmetry\_se 1 36.586 74.586  
## - fractal\_dimension\_se 1 36.826 74.826  
## - concave\_points\_mean 1 36.994 74.994  
## - radius\_worst 1 38.359 76.359  
## - symmetry\_worst 1 38.385 76.385  
## - compactness\_mean 1 39.777 77.777  
## - area\_se 1 41.245 79.245  
## - texture\_mean 1 42.635 80.635  
##   
## Step: AIC=38  
## diagnosis ~ radius\_mean + texture\_mean + perimeter\_mean + compactness\_mean +   
## concavity\_mean + concave\_points\_mean + symmetry\_mean + area\_se +   
## compactness\_se + concavity\_se + concave\_points\_se + symmetry\_se +   
## fractal\_dimension\_se + radius\_worst + perimeter\_worst + compactness\_worst +   
## symmetry\_worst + fractal\_dimension\_worst  
##   
## Df Deviance AIC  
## <none> 0.00 38.00  
## - perimeter\_mean 1 22.05 58.05  
## - compactness\_worst 1 23.08 59.08  
## - concavity\_mean 1 25.78 61.78  
## - radius\_mean 1 26.20 62.20  
## - concavity\_se 1 28.18 64.18  
## - symmetry\_mean 1 28.24 64.24  
## - compactness\_se 1 33.27 69.27  
## - concave\_points\_se 1 34.41 70.41  
## - fractal\_dimension\_worst 1 34.77 70.77  
## - symmetry\_se 1 36.59 72.59  
## - concave\_points\_mean 1 37.00 73.00  
## - fractal\_dimension\_se 1 38.01 74.01  
## - symmetry\_worst 1 38.94 74.94  
## - radius\_worst 1 39.52 75.52  
## - compactness\_mean 1 41.26 77.26  
## - area\_se 1 42.74 78.74  
## - texture\_mean 1 44.61 80.61  
## - perimeter\_worst 1 1081.31 1117.31

## Using Entropy-Based Feature Selection Algorithms

library(FSelectorRcpp)  
x <- information\_gain(diagnosis ~ ., train\_set)  
x %>% arrange(desc(importance)) %>%  
 kable()

|  |  |
| --- | --- |
| attributes | importance |
| perimeter\_worst | 0.4850561 |
| area\_worst | 0.4675581 |
| concave\_points\_worst | 0.4538449 |
| radius\_worst | 0.4478213 |
| concave\_points\_mean | 0.4155797 |
| perimeter\_mean | 0.4087355 |
| area\_mean | 0.3881128 |
| radius\_mean | 0.3814810 |
| area\_se | 0.3664849 |
| concavity\_mean | 0.3499271 |
| concavity\_worst | 0.3458024 |
| radius\_se | 0.2562297 |
| perimeter\_se | 0.2523637 |
| compactness\_worst | 0.2145325 |
| compactness\_mean | 0.2142234 |
| concavity\_se | 0.1483622 |
| concave\_points\_se | 0.1402913 |
| texture\_mean | 0.1265121 |
| texture\_worst | 0.1217746 |
| symmetry\_worst | 0.1008219 |
| smoothness\_worst | 0.0941130 |
| compactness\_se | 0.0691604 |
| symmetry\_mean | 0.0669995 |
| smoothness\_mean | 0.0641805 |
| fractal\_dimension\_worst | 0.0596582 |
| symmetry\_se | 0.0272433 |
| fractal\_dimension\_se | 0.0257642 |
| fractal\_dimension\_mean | 0.0231045 |
| texture\_se | 0.0000000 |
| smoothness\_se | 0.0000000 |

## Recursive Feature Elimination (RFE)

ctrl <- rfeControl(functions = rfFuncs,  
 method = "repeatedcv",  
 repeats = 5,  
 verbose = FALSE)  
  
lmProfile <- rfe(diagnosis ~ .,   
 data = train\_set,  
 rfeControl = ctrl)  
  
lmProfile

##   
## Recursive feature selection  
##   
## Outer resampling method: Cross-Validated (10 fold, repeated 5 times)   
##   
## Resampling performance over subset size:  
##   
## Variables Accuracy Kappa AccuracySD KappaSD Selected  
## 4 0.9080 0.8044 0.03563 0.07491   
## 8 0.9397 0.8715 0.03133 0.06639   
## 16 0.9482 0.8900 0.03104 0.06576 \*  
## 30 0.9402 0.8725 0.03295 0.07047   
##   
## The top 5 variables (out of 16):  
## perimeter\_worst, concave\_points\_worst, area\_worst, radius\_worst, concave\_points\_mean

lmProfile$optVariables

## [1] "perimeter\_worst" "concave\_points\_worst" "area\_worst"   
## [4] "radius\_worst" "concave\_points\_mean" "area\_se"   
## [7] "texture\_worst" "concavity\_worst" "texture\_mean"   
## [10] "concavity\_mean" "area\_mean" "radius\_se"   
## [13] "smoothness\_worst" "perimeter\_mean" "perimeter\_se"   
## [16] "radius\_mean"

var

## function (x, y = NULL, na.rm = FALSE, use)   
## {  
## if (missing(use))   
## use <- if (na.rm)   
## "na.or.complete"  
## else "everything"  
## na.method <- pmatch(use, c("all.obs", "complete.obs", "pairwise.complete.obs",   
## "everything", "na.or.complete"))  
## if (is.na(na.method))   
## stop("invalid 'use' argument")  
## if (is.data.frame(x))   
## x <- as.matrix(x)  
## else stopifnot(is.atomic(x))  
## if (is.data.frame(y))   
## y <- as.matrix(y)  
## else stopifnot(is.atomic(y))  
## .Call(C\_cov, x, y, na.method, FALSE)  
## }  
## <bytecode: 0x0000000029098d10>  
## <environment: namespace:stats>

## Model

cv\_fold <- createFolds(train\_set$diagnosis, k = 5)  
  
train\_ctrl <- trainControl(method = "cv",  
 number = 5,  
 summaryFunction = twoClassSummary,  
 classProbs = TRUE,  
 allowParallel=T,  
 index = cv\_fold,  
 verboseIter = FALSE,  
 savePredictions = TRUE,  
 search = "grid")  
glm\_grid <- expand.grid(  
 alpha = 0:1,  
 lambda = seq(0.0001, 1, length = 10)  
 )

full\_model <- train(  
 diagnosis~.,  
 data = train\_set,  
 method = "glmnet",  
 metric = "ROC",  
 trControl = train\_ctrl,  
 tuneGrid = glm\_grid  
)  
  
full\_model

## glmnet   
##   
## 398 samples  
## 30 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results across tuning parameters:  
##   
## alpha lambda ROC Sens Spec   
## 0 0.0001 0.9883920 0.9867969 0.9127354  
## 0 0.1112 0.9882908 0.9898425 0.8897304  
## 0 0.2223 0.9872652 0.9888325 0.8633518  
## 0 0.3334 0.9865297 0.9898477 0.8501558  
## 0 0.4445 0.9860878 0.9898477 0.8369869  
## 0 0.5556 0.9857958 0.9898477 0.8238315  
## 0 0.6667 0.9854119 0.9898477 0.8156076  
## 0 0.7778 0.9850109 0.9908629 0.8024522  
## 0 0.8889 0.9847022 0.9908629 0.7975207  
## 0 1.0000 0.9843605 0.9908629 0.7942284  
## 1 0.0001 0.9751883 0.9644411 0.8996206  
## 1 0.1112 0.9744598 0.9888325 0.7744615  
## 1 0.2223 0.9658903 0.9959391 0.6462133  
## 1 0.3334 0.9650977 1.0000000 0.1712505  
## 1 0.4445 0.5000000 1.0000000 0.0000000  
## 1 0.5556 0.5000000 1.0000000 0.0000000  
## 1 0.6667 0.5000000 1.0000000 0.0000000  
## 1 0.7778 0.5000000 1.0000000 0.0000000  
## 1 0.8889 0.5000000 1.0000000 0.0000000  
## 1 1.0000 0.5000000 1.0000000 0.0000000  
##   
## ROC was used to select the optimal model using the largest value.  
## The final values used for the model were alpha = 0 and lambda = 1e-04.

## Forward model

forward\_model <- train(  
 forward\_select$formula,  
 data = train\_set,  
 method = "glmnet",  
 metric = "ROC",  
 trControl = train\_ctrl,  
 tuneGrid = glm\_grid  
)  
  
forward\_model

## glmnet   
##   
## 398 samples  
## 30 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results across tuning parameters:  
##   
## alpha lambda ROC Sens Spec   
## 0 0.0001 0.9883920 0.9867969 0.9127354  
## 0 0.1112 0.9882908 0.9898425 0.8897304  
## 0 0.2223 0.9872652 0.9888325 0.8633518  
## 0 0.3334 0.9865297 0.9898477 0.8501558  
## 0 0.4445 0.9860878 0.9898477 0.8369869  
## 0 0.5556 0.9857958 0.9898477 0.8238315  
## 0 0.6667 0.9854119 0.9898477 0.8156076  
## 0 0.7778 0.9850109 0.9908629 0.8024522  
## 0 0.8889 0.9847022 0.9908629 0.7975207  
## 0 1.0000 0.9843605 0.9908629 0.7942284  
## 1 0.0001 0.9751883 0.9644411 0.8996206  
## 1 0.1112 0.9744598 0.9888325 0.7744615  
## 1 0.2223 0.9658903 0.9959391 0.6462133  
## 1 0.3334 0.9650977 1.0000000 0.1712505  
## 1 0.4445 0.5000000 1.0000000 0.0000000  
## 1 0.5556 0.5000000 1.0000000 0.0000000  
## 1 0.6667 0.5000000 1.0000000 0.0000000  
## 1 0.7778 0.5000000 1.0000000 0.0000000  
## 1 0.8889 0.5000000 1.0000000 0.0000000  
## 1 1.0000 0.5000000 1.0000000 0.0000000  
##   
## ROC was used to select the optimal model using the largest value.  
## The final values used for the model were alpha = 0 and lambda = 1e-04.

## Fit model with variables selected from backward selection

back\_model <- train(  
 back\_select$formula,  
 data = train\_set,  
 method = "glmnet",  
 metric = "ROC",  
 trControl = train\_ctrl,  
 tuneGrid = glm\_grid  
)  
  
back\_model

## glmnet   
##   
## 398 samples  
## 18 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results across tuning parameters:  
##   
## alpha lambda ROC Sens Spec   
## 0 0.0001 0.9864075 0.9746141 0.8847175  
## 0 0.1112 0.9844619 0.9766446 0.8583796  
## 0 0.2223 0.9826679 0.9796903 0.8336946  
## 0 0.3334 0.9813837 0.9817259 0.8172063  
## 0 0.4445 0.9804572 0.9817259 0.8040509  
## 0 0.5556 0.9797062 0.9817259 0.7794066  
## 0 0.6667 0.9790299 0.9827411 0.7728492  
## 0 0.7778 0.9785379 0.9837563 0.7629725  
## 0 0.8889 0.9778948 0.9847716 0.7547622  
## 0 1.0000 0.9774693 0.9847716 0.7481913  
## 1 0.0001 0.9745458 0.9512431 0.9145102  
## 1 0.1112 0.9723028 0.9857764 0.7447907  
## 1 0.2223 0.9679694 0.9949239 0.5837014  
## 1 0.3334 0.9647793 1.0000000 0.1317437  
## 1 0.4445 0.5000000 1.0000000 0.0000000  
## 1 0.5556 0.5000000 1.0000000 0.0000000  
## 1 0.6667 0.5000000 1.0000000 0.0000000  
## 1 0.7778 0.5000000 1.0000000 0.0000000  
## 1 0.8889 0.5000000 1.0000000 0.0000000  
## 1 1.0000 0.5000000 1.0000000 0.0000000  
##   
## ROC was used to select the optimal model using the largest value.  
## The final values used for the model were alpha = 0 and lambda = 1e-04.

## Fit model with variables selected from backward selection

back\_model <- train(  
 back\_select$formula,  
 data = train\_set,  
 method = "glmnet",  
 metric = "ROC",  
 trControl = train\_ctrl,  
 tuneGrid = glm\_grid  
)  
  
back\_model

## glmnet   
##   
## 398 samples  
## 18 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results across tuning parameters:  
##   
## alpha lambda ROC Sens Spec   
## 0 0.0001 0.9864075 0.9746141 0.8847175  
## 0 0.1112 0.9844619 0.9766446 0.8583796  
## 0 0.2223 0.9826679 0.9796903 0.8336946  
## 0 0.3334 0.9813837 0.9817259 0.8172063  
## 0 0.4445 0.9804572 0.9817259 0.8040509  
## 0 0.5556 0.9797062 0.9817259 0.7794066  
## 0 0.6667 0.9790299 0.9827411 0.7728492  
## 0 0.7778 0.9785379 0.9837563 0.7629725  
## 0 0.8889 0.9778948 0.9847716 0.7547622  
## 0 1.0000 0.9774693 0.9847716 0.7481913  
## 1 0.0001 0.9745458 0.9512431 0.9145102  
## 1 0.1112 0.9723028 0.9857764 0.7447907  
## 1 0.2223 0.9679694 0.9949239 0.5837014  
## 1 0.3334 0.9647793 1.0000000 0.1317437  
## 1 0.4445 0.5000000 1.0000000 0.0000000  
## 1 0.5556 0.5000000 1.0000000 0.0000000  
## 1 0.6667 0.5000000 1.0000000 0.0000000  
## 1 0.7778 0.5000000 1.0000000 0.0000000  
## 1 0.8889 0.5000000 1.0000000 0.0000000  
## 1 1.0000 0.5000000 1.0000000 0.0000000  
##   
## ROC was used to select the optimal model using the largest value.  
## The final values used for the model were alpha = 0 and lambda = 1e-04.

## Fit model with variables selected from entropy

setDT(x)  
#selector predictors with importance of more than 0.05  
predictors <- x[importance > 0.05, attributes]  
  
entropy\_predctors <- train\_set[, ..predictors]  
entropy\_y <- train\_set$diagnosis  
entropy\_model <- train(  
 entropy\_predctors,  
 entropy\_y,  
 method = "glm",  
 metric = "ROC",  
 trControl = train\_ctrl  
)  
  
entropy\_model

## Generalized Linear Model   
##   
## 398 samples  
## 25 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results:  
##   
## ROC Sens Spec   
## 0.9335298 0.9227701 0.8288714

## Fit model with variables selected Recursive Feature Elimination

recu\_pred <- lmProfile$optVariables  
recursive\_predctors <- train\_set[, ..recu\_pred]  
recursive\_y <- train\_set$diagnosis  
recu\_model <- train(  
 recursive\_predctors,  
 recursive\_y,  
 method = "glm",  
 metric = "ROC",  
 trControl = train\_ctrl  
)  
  
recu\_model

## Generalized Linear Model   
##   
## 398 samples  
## 16 predictor  
## 2 classes: 'B', 'M'   
##   
## No pre-processing  
## Resampling: Cross-Validated (5 fold)   
## Summary of sample sizes: 79, 79, 80, 81, 79   
## Resampling results:  
##   
## ROC Sens Spec   
## 0.9275906 0.9268103 0.8437339

## Full model test accuracy

for\_glm <- predict(full\_model, test\_set, type = "prob")  
  
  
for\_glm1 <- ifelse(for\_glm[, "M"] > 0.5, "M", "B")  
for\_glm1 <- factor(for\_glm1, levels = levels(test\_set$diagnosis))  
  
  
  
confusionMatrix(for\_glm1, test\_set$diagnosis,positive = "M")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction B M  
## B 110 4  
## M 1 56  
##   
## Accuracy : 0.9708   
## 95% CI : (0.9331, 0.9904)  
## No Information Rate : 0.6491   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9351   
##   
## Mcnemar's Test P-Value : 0.3711   
##   
## Sensitivity : 0.9333   
## Specificity : 0.9910   
## Pos Pred Value : 0.9825   
## Neg Pred Value : 0.9649   
## Prevalence : 0.3509   
## Detection Rate : 0.3275   
## Detection Prevalence : 0.3333   
## Balanced Accuracy : 0.9622   
##   
## 'Positive' Class : M   
##

## Forward test accuracy

for\_glm <- predict(forward\_model, test\_set, type = "prob")  
  
  
for\_glm1 <- ifelse(for\_glm[, "M"] > 0.5, "M", "B")  
for\_glm1 <- factor(for\_glm1, levels = levels(test\_set$diagnosis))  
  
  
  
confusionMatrix(for\_glm1, test\_set$diagnosis,positive = "M")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction B M  
## B 110 4  
## M 1 56  
##   
## Accuracy : 0.9708   
## 95% CI : (0.9331, 0.9904)  
## No Information Rate : 0.6491   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9351   
##   
## Mcnemar's Test P-Value : 0.3711   
##   
## Sensitivity : 0.9333   
## Specificity : 0.9910   
## Pos Pred Value : 0.9825   
## Neg Pred Value : 0.9649   
## Prevalence : 0.3509   
## Detection Rate : 0.3275   
## Detection Prevalence : 0.3333   
## Balanced Accuracy : 0.9622   
##   
## 'Positive' Class : M   
##

## Backward test accuracy

for\_glm <- predict(back\_model, test\_set, type = "prob")  
  
  
for\_glm1 <- ifelse(for\_glm[, "M"] > 0.5, "M", "B")  
for\_glm1 <- factor(for\_glm1, levels = levels(test\_set$diagnosis))  
  
  
  
confusionMatrix(for\_glm1, test\_set$diagnosis,positive = "M")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction B M  
## B 110 4  
## M 1 56  
##   
## Accuracy : 0.9708   
## 95% CI : (0.9331, 0.9904)  
## No Information Rate : 0.6491   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9351   
##   
## Mcnemar's Test P-Value : 0.3711   
##   
## Sensitivity : 0.9333   
## Specificity : 0.9910   
## Pos Pred Value : 0.9825   
## Neg Pred Value : 0.9649   
## Prevalence : 0.3509   
## Detection Rate : 0.3275   
## Detection Prevalence : 0.3333   
## Balanced Accuracy : 0.9622   
##   
## 'Positive' Class : M   
##

## entropy method test accuracy

for\_glm <- predict(entropy\_model, test\_set, type = "prob")  
  
  
for\_glm1 <- ifelse(for\_glm[, "M"] > 0.5, "M", "B")  
for\_glm1 <- factor(for\_glm1, levels = levels(test\_set$diagnosis))  
  
  
  
confusionMatrix(for\_glm1, test\_set$diagnosis,positive = "M")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction B M  
## B 106 2  
## M 5 58  
##   
## Accuracy : 0.9591   
## 95% CI : (0.9175, 0.9834)  
## No Information Rate : 0.6491   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9112   
##   
## Mcnemar's Test P-Value : 0.4497   
##   
## Sensitivity : 0.9667   
## Specificity : 0.9550   
## Pos Pred Value : 0.9206   
## Neg Pred Value : 0.9815   
## Prevalence : 0.3509   
## Detection Rate : 0.3392   
## Detection Prevalence : 0.3684   
## Balanced Accuracy : 0.9608   
##   
## 'Positive' Class : M   
##

## Recursive Feature Elimination method test accuracy

for\_glm <- predict(recu\_model, test\_set, type = "prob")  
  
  
for\_glm1 <- ifelse(for\_glm[, "M"] > 0.5, "M", "B")  
for\_glm1 <- factor(for\_glm1, levels = levels(test\_set$diagnosis))  
  
  
  
confusionMatrix(for\_glm1, test\_set$diagnosis,positive = "M")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction B M  
## B 104 0  
## M 7 60  
##   
## Accuracy : 0.9591   
## 95% CI : (0.9175, 0.9834)  
## No Information Rate : 0.6491   
## P-Value [Acc > NIR] : < 2e-16   
##   
## Kappa : 0.9125   
##   
## Mcnemar's Test P-Value : 0.02334   
##   
## Sensitivity : 1.0000   
## Specificity : 0.9369   
## Pos Pred Value : 0.8955   
## Neg Pred Value : 1.0000   
## Prevalence : 0.3509   
## Detection Rate : 0.3509   
## Detection Prevalence : 0.3918   
## Balanced Accuracy : 0.9685   
##   
## 'Positive' Class : M   
##