final\_project\_p8106

Michelle Lui

4/27/2021

## Import dataset

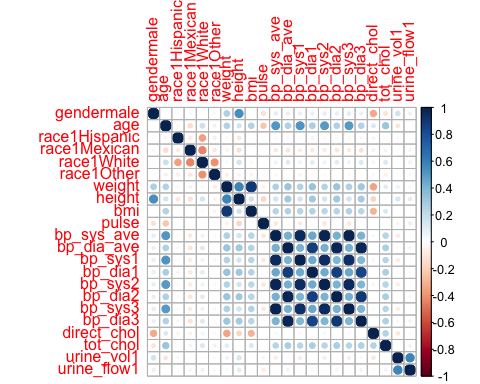
#import dataset  
data("NHANES")  
  
#clean dataset - limit to 2011/12 and only include biological predictors, dropped some biological predictors due to excess missing data  
nhanes = NHANES %>% janitor::clean\_names() %>% filter(survey\_yr == "2011\_12") %>% select(gender, age, race1, weight, height, bmi, pulse:bp\_dia3, direct\_chol:urine\_flow1, diabetes) %>% drop\_na()

# Exploratory Data Analysis

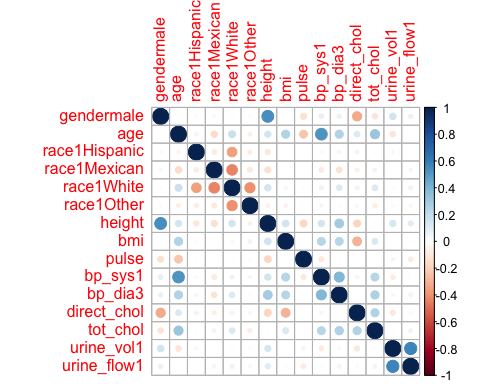
nhanes\_2 =   
 nhanes %>%  
 select(diabetes, everything())  
  
summary(nhanes\_2)

## diabetes gender age race1 weight   
## No :3223 female:1735 Min. : 8.00 Black : 392 Min. : 23.00   
## Yes: 293 male :1781 1st Qu.:25.00 Hispanic: 218 1st Qu.: 62.90   
## Median :42.00 Mexican : 315 Median : 77.10   
## Mean :41.81 White :2319 Mean : 78.25   
## 3rd Qu.:57.00 Other : 272 3rd Qu.: 91.10   
## Max. :80.00 Max. :188.50   
## height bmi pulse bp\_sys\_ave   
## Min. :124.5 Min. :12.90 Min. : 40.00 Min. : 81   
## 1st Qu.:160.3 1st Qu.:22.90 1st Qu.: 64.00 1st Qu.:107   
## Median :167.8 Median :26.60 Median : 72.00 Median :117   
## Mean :167.5 Mean :27.63 Mean : 73.41 Mean :119   
## 3rd Qu.:175.4 3rd Qu.:31.30 3rd Qu.: 82.00 3rd Qu.:128   
## Max. :199.9 Max. :69.00 Max. :136.00 Max. :209   
## bp\_dia\_ave bp\_sys1 bp\_dia1 bp\_sys2   
## Min. : 0.00 Min. : 74.0 Min. : 0.0 Min. : 82.0   
## 1st Qu.: 62.00 1st Qu.:108.0 1st Qu.: 62.0 1st Qu.:108.0   
## Median : 69.00 Median :118.0 Median : 70.0 Median :118.0   
## Mean : 68.44 Mean :119.5 Mean : 69.2 Mean :119.2   
## 3rd Qu.: 77.00 3rd Qu.:128.0 3rd Qu.: 78.0 3rd Qu.:128.0   
## Max. :116.00 Max. :212.0 Max. :110.0 Max. :208.0   
## bp\_dia2 bp\_sys3 bp\_dia3 direct\_chol   
## Min. : 0.00 Min. : 78.0 Min. : 0.00 Min. :0.470   
## 1st Qu.: 62.00 1st Qu.:106.0 1st Qu.: 62.00 1st Qu.:1.110   
## Median : 70.00 Median :116.0 Median : 70.00 Median :1.290   
## Mean : 68.55 Mean :118.7 Mean : 68.33 Mean :1.365   
## 3rd Qu.: 78.00 3rd Qu.:128.0 3rd Qu.: 78.00 3rd Qu.:1.580   
## Max. :116.00 Max. :210.0 Max. :116.00 Max. :4.030   
## tot\_chol urine\_vol1 urine\_flow1   
## Min. : 1.530 Min. : 1 Min. : 0.0110   
## 1st Qu.: 4.110 1st Qu.: 47 1st Qu.: 0.3995   
## Median : 4.780 Median : 88 Median : 0.6800   
## Mean : 4.858 Mean :113 Mean : 0.9642   
## 3rd Qu.: 5.530 3rd Qu.:156 3rd Qu.: 1.2140   
## Max. :10.290 Max. :446 Max. :10.1430

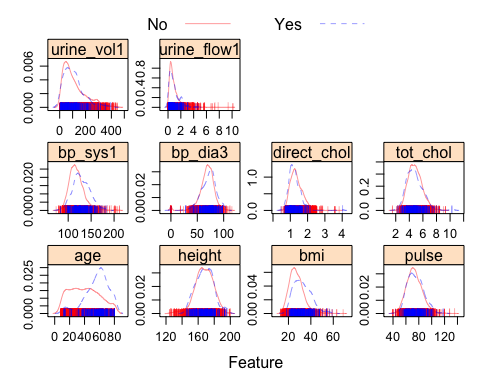
#correlation plots for predictor variables  
corrx = model.matrix(diabetes~., nhanes\_2)[,-1]  
corrplot(cor(corrx))



#remove variables with correlation greater than 0.8  
indexesToDrop = findCorrelation(cor(corrx), cutoff = 0.8)  
corrplot(cor(corrx[,-indexesToDrop]))



#final dataset for analysis using RF and analysis  
nhanes\_3 =   
 nhanes\_2 %>%  
 select(diabetes, age, height, bmi, pulse, bp\_sys1, bp\_dia3, direct\_chol, tot\_chol,  
 urine\_vol1, urine\_flow1, gender, race1)  
  
  
# look at feature plots for continuous predictors (everything but race1 and gender)  
theme1 = transparentTheme(trans =.4)  
trellis.par.set(theme1)  
  
  
featurePlot(x=nhanes\_3[,2:11],  
 y=nhanes\_3$diabetes,  
 scales =list(x=list(relation ="free"),  
 y=list(relation ="free")),  
 plot ="density",pch ="|",  
 auto.key =list(columns =2))



## Create Data Partition

set.seed(2)  
  
train\_rows = createDataPartition(y = nhanes\_3$diabetes,p = 0.7,list = FALSE)  
train = nhanes\_3[train\_rows, ]

## Warning: The `i` argument of ``[`()` can't be a matrix as of tibble 3.0.0.  
## Convert to a vector.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_warnings()` to see where this warning was generated.

test = nhanes\_3[-train\_rows, ]  
  
x = model.matrix(diabetes ~ ., train)[ ,-1]  
y = train$diabetes  
  
x2 = model.matrix(diabetes ~ ., test)[ ,-1]  
y2 = test$diabetes  
  
control1 = trainControl(method = "cv", selectionFunction = "best", sampling = "down")

## Lasso Model

set.seed(2)  
lasso\_fit = train(x, y,   
 method = "glmnet",  
 tuneGrid = expand.grid(alpha = 1, lambda = exp(seq(-20, 20,length=100))),  
 trControl = control1,  
 preProcess=c("center", "scale"),  
 family = "binomial")   
  
#Print the values of alpha and lambda that gave best prediction  
lasso\_fit$bestTune

## alpha lambda  
## 38 1 0.006406097

#Print all of the options examined  
lasso\_fit$results

## alpha lambda Accuracy Kappa AccuracySD KappaSD  
## 1 1 2.061154e-09 0.7316292 0.2097020 0.03069130 0.04648869  
## 2 1 3.087329e-09 0.7316292 0.2097020 0.03069130 0.04648869  
## 3 1 4.624400e-09 0.7316292 0.2097020 0.03069130 0.04648869  
## 4 1 6.926725e-09 0.7316292 0.2097020 0.03069130 0.04648869  
## 5 1 1.037529e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 6 1 1.554078e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 7 1 2.327799e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 8 1 3.486727e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 9 1 5.222645e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 10 1 7.822814e-08 0.7316292 0.2097020 0.03069130 0.04648869  
## 11 1 1.171752e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 12 1 1.755125e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 13 1 2.628939e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 14 1 3.937795e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 15 1 5.898283e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 16 1 8.834829e-07 0.7316292 0.2097020 0.03069130 0.04648869  
## 17 1 1.323338e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 18 1 1.982180e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 19 1 2.969038e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 20 1 4.447216e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 21 1 6.661326e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 22 1 9.977764e-06 0.7316292 0.2097020 0.03069130 0.04648869  
## 23 1 1.494534e-05 0.7316292 0.2097020 0.03069130 0.04648869  
## 24 1 2.238609e-05 0.7316292 0.2097020 0.03069130 0.04648869  
## 25 1 3.353133e-05 0.7316292 0.2097020 0.03069130 0.04648869  
## 26 1 5.022539e-05 0.7316292 0.2097020 0.03069130 0.04648869  
## 27 1 7.523083e-05 0.7316292 0.2097020 0.03069130 0.04648869  
## 28 1 1.126856e-04 0.7316292 0.2097020 0.03069130 0.04648869  
## 29 1 1.687877e-04 0.7316292 0.2097020 0.03069130 0.04648869  
## 30 1 2.528211e-04 0.7316292 0.2097020 0.03069130 0.04648869  
## 31 1 3.786918e-04 0.7316292 0.2097020 0.03069130 0.04648869  
## 32 1 5.672290e-04 0.7312227 0.2093622 0.03089586 0.04713098  
## 33 1 8.496321e-04 0.7308162 0.2107979 0.03156203 0.04394624  
## 34 1 1.272634e-03 0.7320340 0.2118739 0.03043488 0.04329136  
## 35 1 1.906233e-03 0.7320357 0.2119628 0.03101889 0.04444054  
## 36 1 2.855279e-03 0.7320340 0.2114728 0.02786084 0.04070302  
## 37 1 4.276820e-03 0.7304113 0.2101363 0.02952107 0.04323945  
## 38 1 6.406097e-03 0.7320390 0.2160079 0.02794914 0.04695292  
## 39 1 9.595465e-03 0.7235238 0.2088359 0.02496772 0.04389999  
## 40 1 1.437271e-02 0.7198652 0.2181297 0.02588053 0.03343138  
## 41 1 2.152837e-02 0.7174328 0.2185267 0.02293235 0.03639613  
## 42 1 3.224658e-02 0.7105238 0.2071801 0.02161712 0.03689326  
## 43 1 4.830100e-02 0.7048426 0.2016200 0.01713358 0.03999908  
## 44 1 7.234835e-02 0.6971305 0.1907318 0.01528560 0.03501996  
## 45 1 1.083680e-01 0.6861796 0.1863047 0.02413036 0.04322860  
## 46 1 1.623206e-01 0.6715618 0.1752595 0.03056515 0.04497615  
## 47 1 2.431343e-01 0.4185278 0.0000000 0.43040766 0.00000000  
## 48 1 3.641822e-01 0.4185278 0.0000000 0.43040766 0.00000000  
## 49 1 5.454956e-01 0.4185278 0.0000000 0.43040766 0.00000000  
## 50 1 8.170784e-01 0.4185278 0.0000000 0.43040766 0.00000000  
## 51 1 1.223873e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 52 1 1.833195e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 53 1 2.745878e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 54 1 4.112954e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 55 1 6.160647e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 56 1 9.227814e+00 0.4185278 0.0000000 0.43040766 0.00000000  
## 57 1 1.382202e+01 0.4185278 0.0000000 0.43040766 0.00000000  
## 58 1 2.070351e+01 0.4185278 0.0000000 0.43040766 0.00000000  
## 59 1 3.101105e+01 0.4185278 0.0000000 0.43040766 0.00000000  
## 60 1 4.645034e+01 0.4185278 0.0000000 0.43040766 0.00000000  
## 61 1 6.957632e+01 0.4185278 0.0000000 0.43040766 0.00000000  
## 62 1 1.042159e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 63 1 1.561013e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 64 1 2.338186e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 65 1 3.502285e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 66 1 5.245949e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 67 1 7.857720e+02 0.4185278 0.0000000 0.43040766 0.00000000  
## 68 1 1.176980e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 69 1 1.762956e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 70 1 2.640670e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 71 1 3.955365e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 72 1 5.924601e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 73 1 8.874250e+03 0.4185278 0.0000000 0.43040766 0.00000000  
## 74 1 1.329242e+04 0.4185278 0.0000000 0.43040766 0.00000000  
## 75 1 1.991025e+04 0.4185278 0.0000000 0.43040766 0.00000000  
## 76 1 2.982285e+04 0.4185278 0.0000000 0.43040766 0.00000000  
## 77 1 4.467059e+04 0.4185278 0.0000000 0.43040766 0.00000000  
## 78 1 6.691050e+04 0.4185278 0.0000000 0.43040766 0.00000000  
## 79 1 1.002229e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 80 1 1.501203e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 81 1 2.248598e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 82 1 3.368095e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 83 1 5.044950e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 84 1 7.556651e+05 0.4185278 0.0000000 0.43040766 0.00000000  
## 85 1 1.131884e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 86 1 1.695409e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 87 1 2.539492e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 88 1 3.803815e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 89 1 5.697600e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 90 1 8.534232e+06 0.4185278 0.0000000 0.43040766 0.00000000  
## 91 1 1.278312e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 92 1 1.914739e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 93 1 2.868019e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 94 1 4.295904e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 95 1 6.434681e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 96 1 9.638281e+07 0.4185278 0.0000000 0.43040766 0.00000000  
## 97 1 1.443684e+08 0.4185278 0.0000000 0.43040766 0.00000000  
## 98 1 2.162443e+08 0.4185278 0.0000000 0.43040766 0.00000000  
## 99 1 3.239046e+08 0.4185278 0.0000000 0.43040766 0.00000000  
## 100 1 4.851652e+08 0.4185278 0.0000000 0.43040766 0.00000000

# Model coefficients  
coef(lasso\_fit$finalModel, s = lasso\_fit$bestTune$lambda)

## 16 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) -0.04598234  
## age 1.13689677  
## height -0.03059534  
## bmi 0.69433892  
## pulse 0.41099385  
## bp\_sys1 0.09701002  
## bp\_dia3 0.02860917  
## direct\_chol -0.31025394  
## tot\_chol -0.07709907  
## urine\_vol1 0.09126462  
## urine\_flow1 0.11977272  
## gendermale 0.11819933  
## race1Hispanic .   
## race1Mexican -0.12418383  
## race1White -0.31766043  
## race1Other .

# Make predictions  
lasso\_pred = lasso\_fit %>% predict(x2) %>% as.numeric()  
lasso\_pred\_p = ifelse(lasso\_pred-1 > 0.5,1,0)  
  
test\_outcome\_lasso = (as.numeric(y2)-1)  
  
misclasserror\_lasso = mean(lasso\_pred\_p != test\_outcome\_lasso, na.rm=T)  
print(paste('Accuracy Model 1', 1-misclasserror\_lasso))

## [1] "Accuracy Model 1 0.682811016144349"

## Random Forest

mtry\_vals = c(ncol(train)-1, sqrt(ncol(train)-1), 0.5\*ncol(train)-1)  
  
mtry\_grid = expand.grid(.mtry=mtry\_vals)  
  
set.seed(2)  
rf\_fit = train(diabetes ~.,   
 data = train,   
 method="rf",   
 trControl = control1,   
 metric="Accuracy",   
 tuneGrid=mtry\_grid,   
 ntree=100)  
  
rf\_fit$results

## mtry Accuracy Kappa AccuracySD KappaSD  
## 1 3.464102 0.8031067 0.3411511 0.03894405 0.05986997  
## 2 5.500000 0.7981824 0.3334355 0.02040004 0.04224013  
## 3 12.000000 0.7994266 0.3212905 0.01503463 0.03661672

rf\_fit$bestTune

## mtry  
## 1 3.464102

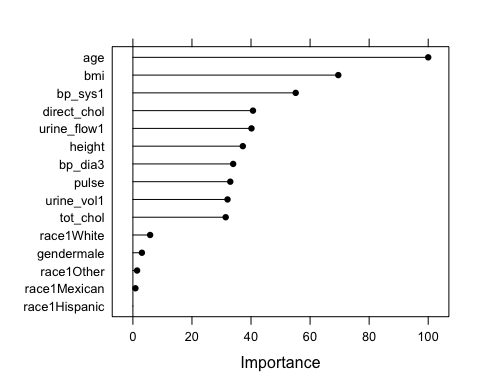
rf\_fit$finalModel

##   
## Call:  
## randomForest(x = x, y = y, ntree = 100, mtry = param$mtry)   
## Type of random forest: classification  
## Number of trees: 100  
## No. of variables tried at each split: 3  
##   
## OOB estimate of error rate: 16.02%  
## Confusion matrix:  
## No Yes class.error  
## No 166 40 0.1941748  
## Yes 26 180 0.1262136

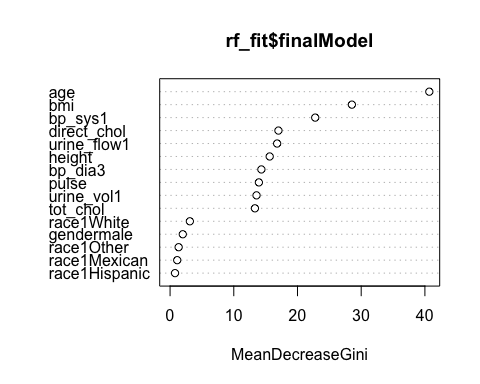
varImp(rf\_fit)

## rf variable importance  
##   
## Overall  
## age 100.0000  
## bmi 69.5471  
## bp\_sys1 55.1292  
## direct\_chol 40.6630  
## urine\_flow1 40.1548  
## height 37.2218  
## bp\_dia3 33.9554  
## pulse 32.9752  
## urine\_vol1 32.0458  
## tot\_chol 31.4214  
## race1White 5.8331  
## gendermale 3.0511  
## race1Other 1.4182  
## race1Mexican 0.8581  
## race1Hispanic 0.0000

plot(varImp(rf\_fit))



varImpPlot(rf\_fit$finalModel)



rf\_pred = predict(rf\_fit, test) %>% as.numeric()  
rf\_pred\_p = ifelse(rf\_pred-1 > 0.5,1,0)  
  
test\_outcome\_rf = (as.numeric(test$diabetes)-1)  
  
misclasserror\_rf = mean(rf\_pred\_p != test\_outcome\_rf, na.rm=T)  
print(paste('Accuracy Model 2', 1-misclasserror\_rf))

## [1] "Accuracy Model 2 0.804368471035138"

## Model Comparisons

resamp = resamples(list(lasso = lasso\_fit, rf = rf\_fit))   
summary(resamp)

##   
## Call:  
## summary.resamples(object = resamp)  
##   
## Models: lasso, rf   
## Number of resamples: 10   
##   
## Accuracy   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## lasso 0.6869919 0.7154472 0.7302261 0.7320390 0.7515492 0.7764228 0  
## rf 0.7439024 0.7786733 0.8008130 0.8031067 0.8401760 0.8536585 0  
##   
## Kappa   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## lasso 0.1460028 0.1816105 0.2281826 0.2160079 0.2448435 0.2893161 0  
## rf 0.2792530 0.2945411 0.3216341 0.3411511 0.3837602 0.4595386 0

## Final dataset after selecting variables from RF model

We selected variables from our random forest model since it had the higher accuracy compared to the lasso model

#Final dataset after variable selection  
nhanes\_4 =   
 nhanes\_3 %>%  
 select(diabetes, age, bmi, bp\_sys1, direct\_chol, urine\_flow1, height, bp\_dia3, pulse, urine\_vol1,  
 tot\_chol)  
  
#training and testing  
  
train\_rows2 = createDataPartition(y = nhanes\_4$diabetes,p = 0.7,list = FALSE)  
train2 = nhanes\_4[train\_rows, ]  
test2 = nhanes\_4[-train\_rows, ]

summary(nhanes\_4)

## diabetes age bmi bp\_sys1 direct\_chol   
## No :3223 Min. : 8.00 Min. :12.90 Min. : 74.0 Min. :0.470   
## Yes: 293 1st Qu.:25.00 1st Qu.:22.90 1st Qu.:108.0 1st Qu.:1.110   
## Median :42.00 Median :26.60 Median :118.0 Median :1.290   
## Mean :41.81 Mean :27.63 Mean :119.5 Mean :1.365   
## 3rd Qu.:57.00 3rd Qu.:31.30 3rd Qu.:128.0 3rd Qu.:1.580   
## Max. :80.00 Max. :69.00 Max. :212.0 Max. :4.030   
## urine\_flow1 height bp\_dia3 pulse   
## Min. : 0.0110 Min. :124.5 Min. : 0.00 Min. : 40.00   
## 1st Qu.: 0.3995 1st Qu.:160.3 1st Qu.: 62.00 1st Qu.: 64.00   
## Median : 0.6800 Median :167.8 Median : 70.00 Median : 72.00   
## Mean : 0.9642 Mean :167.5 Mean : 68.33 Mean : 73.41   
## 3rd Qu.: 1.2140 3rd Qu.:175.4 3rd Qu.: 78.00 3rd Qu.: 82.00   
## Max. :10.1430 Max. :199.9 Max. :116.00 Max. :136.00   
## urine\_vol1 tot\_chol   
## Min. : 1 Min. : 1.530   
## 1st Qu.: 47 1st Qu.: 4.110   
## Median : 88 Median : 4.780   
## Mean :113 Mean : 4.858   
## 3rd Qu.:156 3rd Qu.: 5.530   
## Max. :446 Max. :10.290

## SVC Analysis

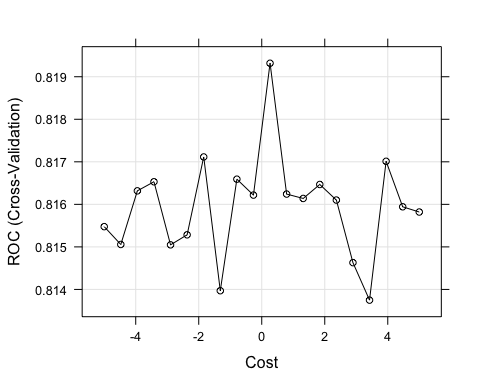
ctrl = trainControl(method = "cv", sampling = "up",summaryFunction = twoClassSummary,  
 classProbs = TRUE)  
  
#Analysis using linear kernal  
# kernlab  
set.seed(2)  
svml.fit = train(diabetes ~ . ,   
 data = nhanes\_4[train\_rows,],   
 method = "svmLinear",  
 preProcess = c("center", "scale"),  
 tuneGrid = data.frame(C = exp(seq(-5,5,len=20))),  
 trControl = ctrl)

## Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not  
## in the result set. ROC will be used instead.

svml.fit #C = 6.309808

## Support Vector Machines with Linear Kernel   
##   
## 2463 samples  
## 10 predictor  
## 2 classes: 'No', 'Yes'   
##   
## Pre-processing: centered (10), scaled (10)   
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2216, 2217, 2216, 2217, 2216, 2217, ...   
## Addtional sampling using up-sampling prior to pre-processing  
##   
## Resampling results across tuning parameters:  
##   
## C ROC Sens Spec   
## 6.737947e-03 0.8154760 0.7350442 0.7764286  
## 1.140522e-02 0.8150567 0.7368279 0.7616667  
## 1.930544e-02 0.8163175 0.7315044 0.7619048  
## 3.267802e-02 0.8165309 0.7332763 0.7666667  
## 5.531358e-02 0.8150480 0.7301770 0.7566667  
## 9.362844e-02 0.8152839 0.7332704 0.7811905  
## 1.584834e-01 0.8171141 0.7323874 0.7809524  
## 2.682625e-01 0.8139699 0.7346077 0.7666667  
## 4.540837e-01 0.8165915 0.7332822 0.7757143  
## 7.686205e-01 0.8162178 0.7319430 0.7811905  
## 1.301032e+00 0.8193168 0.7337266 0.7616667  
## 2.202237e+00 0.8162417 0.7332802 0.7809524  
## 3.727693e+00 0.8161388 0.7292901 0.7664286  
## 6.309808e+00 0.8164673 0.7283992 0.7616667  
## 1.068051e+01 0.8161010 0.7323854 0.7711905  
## 1.807874e+01 0.8146298 0.7297325 0.7569048  
## 3.060161e+01 0.8137485 0.7328279 0.7616667  
## 5.179887e+01 0.8170119 0.7323933 0.7761905  
## 8.767916e+01 0.8159427 0.7288417 0.7664286  
## 1.484132e+02 0.8158197 0.7279607 0.7714286  
##   
## ROC was used to select the optimal model using the largest value.  
## The final value used for the model was C = 1.301032.

plot(svml.fit, highlight = TRUE, xTrans = log)



#Analysis using radial kernal  
svmr.grid = expand.grid(C = exp(seq(-4,6,len=10)),  
 sigma = exp(seq(-6,1,len=10)))  
  
# tunes over both cost and sigma  
set.seed(2)   
svmr.fit = train(diabetes ~ . , nhanes\_4,  
 subset = train\_rows2,  
 method = "svmRadialSigma",  
 preProcess = c("center", "scale"),  
 tuneGrid = svmr.grid,  
 trControl = ctrl)

## Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not  
## in the result set. ROC will be used instead.

## line search fails -1.607978 0.1638672 1.323618e-05 -3.902699e-06 -3.182967e-08 1.304802e-08 -4.722257e-13

## Warning in method$predict(modelFit = modelFit, newdata = newdata, submodels =  
## param): kernlab class prediction calculations failed; returning NAs

## Warning in method$prob(modelFit = modelFit, newdata = newdata, submodels =  
## param): kernlab class probability calculations failed; returning NAs

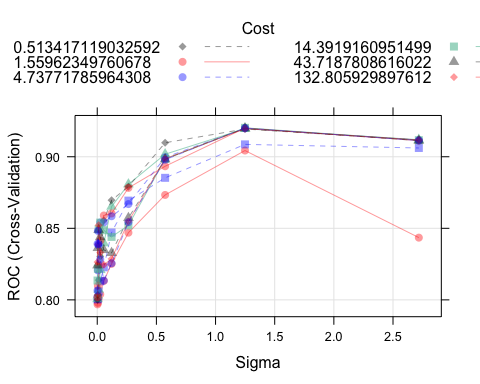
## Warning in data.frame(..., check.names = FALSE): row names were found from a  
## short variable and have been discarded

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :  
## There were missing values in resampled performance measures.

svmr.fit #C = 1.559623, sigma = 1.248849

## Support Vector Machines with Radial Basis Function Kernel   
##   
## 3516 samples  
## 10 predictor  
## 2 classes: 'No', 'Yes'   
##   
## Pre-processing: centered (10), scaled (10)   
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2216, 2217, 2216, 2217, 2216, 2217, ...   
## Addtional sampling using up-sampling prior to pre-processing  
##   
## Resampling results across tuning parameters:  
##   
## C sigma ROC Sens Spec   
## 0.01831564 0.002478752 0.7968681 0.6411131 0.8297619  
## 0.01831564 0.005395326 0.7979282 0.7664975 0.7135714  
## 0.01831564 0.011743628 0.8003822 0.7390265 0.7619048  
## 0.01831564 0.025561533 0.8038525 0.7235241 0.7571429  
## 0.01831564 0.055637998 0.8132386 0.7461259 0.7573810  
## 0.01831564 0.121103332 0.8251826 0.7718210 0.7521429  
## 0.01831564 0.263597138 0.8469285 0.8245408 0.7285714  
## 0.01831564 0.573753421 0.8733136 0.9175949 0.5540476  
## 0.01831564 1.248848869 0.9043390 1.0000000 0.2533333  
## 0.01831564 2.718281828 0.8434755 1.0000000 0.2235714  
## 0.05563800 0.002478752 0.8003009 0.7133451 0.7859524  
## 0.05563800 0.005395326 0.8008798 0.7257424 0.7716667  
## 0.05563800 0.011743628 0.8059285 0.7315025 0.7664286  
## 0.05563800 0.025561533 0.8129502 0.7483402 0.7666667  
## 0.05563800 0.055637998 0.8227581 0.7580905 0.7711905  
## 0.05563800 0.121103332 0.8470433 0.7899941 0.7421429  
## 0.05563800 0.263597138 0.8693454 0.8697542 0.6509524  
## 0.05563800 0.573753421 0.8852698 0.9401632 0.5438095  
## 0.05563800 1.248848869 0.9086323 0.9871485 0.4176190  
## 0.05563800 2.718281828 0.9061319 0.9986706 0.3983333  
## 0.16901332 0.002478752 0.8005587 0.7115713 0.7711905  
## 0.16901332 0.005395326 0.8030630 0.7270678 0.7666667  
## 0.16901332 0.011743628 0.8072256 0.7315044 0.7373810  
## 0.16901332 0.025561533 0.8208879 0.7536696 0.7516667  
## 0.16901332 0.055637998 0.8396007 0.7828948 0.7328571  
## 0.16901332 0.121103332 0.8651461 0.8462675 0.7135714  
## 0.16901332 0.263597138 0.8807577 0.9078446 0.6166667  
## 0.16901332 0.573753421 0.9017591 0.9773963 0.5154762  
## 0.16901332 1.248848869 0.9195559 0.9933550 0.3983333  
## 0.16901332 2.718281828 0.9118899 0.9995575 0.3983333  
## 0.51341712 0.002478752 0.8000697 0.7186470 0.7564286  
## 0.51341712 0.005395326 0.8034351 0.7275241 0.7514286  
## 0.51341712 0.011743628 0.8137549 0.7443559 0.7416667  
## 0.51341712 0.025561533 0.8303839 0.7713864 0.7419048  
## 0.51341712 0.055637998 0.8555613 0.8223363 0.7321429  
## 0.51341712 0.121103332 0.8696345 0.8803756 0.6650000  
## 0.51341712 0.263597138 0.8791077 0.9578997 0.5238095  
## 0.51341712 0.573753421 0.9097887 0.9929125 0.4371429  
## 0.51341712 1.248848869 0.9193704 0.9991150 0.3983333  
## 0.51341712 2.718281828 0.9116166 1.0000000 0.3983333  
## 1.55962350 0.002478752 0.8019812 0.7266372 0.7564286  
## 1.55962350 0.005395326 0.8098279 0.7394907 0.7514286  
## 1.55962350 0.011743628 0.8241250 0.7629597 0.7371429  
## 1.55962350 0.025561533 0.8413272 0.7983992 0.7330952  
## 1.55962350 0.055637998 0.8589392 0.8528968 0.7183333  
## 1.55962350 0.121103332 0.8604182 0.9193569 0.5680952  
## 1.55962350 0.263597138 0.8782436 0.9778446 0.5054762  
## 1.55962350 0.573753421 0.8933542 0.9982262 0.4130952  
## 1.55962350 1.248848869 0.9200823 1.0000000 0.3983333  
## 1.55962350 2.718281828 0.9112518 1.0000000 0.3983333  
## 4.73771786 0.002478752 0.8065146 0.7337286 0.7514286  
## 4.73771786 0.005395326 0.8206244 0.7518938 0.7371429  
## 4.73771786 0.011743628 0.8379888 0.7886667 0.7426190  
## 4.73771786 0.025561533 0.8540714 0.8320787 0.7180952  
## 4.73771786 0.055637998 0.8540518 0.8803697 0.6404762  
## 4.73771786 0.121103332 0.8584313 0.9521396 0.5585714  
## 4.73771786 0.263597138 0.8669299 0.9853766 0.4671429  
## 4.73771786 0.573753421 0.8995615 0.9995575 0.4033333  
## 4.73771786 1.248848869 0.9198416 1.0000000 0.3983333  
## 4.73771786 2.718281828 0.9116353 1.0000000 0.3983333  
## 14.39191610 0.002478752 0.8135069 0.7461377 0.7369048  
## 14.39191610 0.005395326 0.8217556 0.7700728 0.7373810  
## 14.39191610 0.011743628 0.8493059 0.8060898 0.7798942  
## 14.39191610 0.025561533 0.8539867 0.8559961 0.6940476  
## 14.39191610 0.055637998 0.8486510 0.9118230 0.5773810  
## 14.39191610 0.121103332 0.8440102 0.9680983 0.5247619  
## 14.39191610 0.263597138 0.8520654 0.9906942 0.4571429  
## 14.39191610 0.573753421 0.8984934 0.9995575 0.4033333  
## 14.39191610 1.248848869 0.9202123 1.0000000 0.3983333  
## 14.39191610 2.718281828 0.9115886 1.0000000 0.3983333  
## 43.71878086 0.002478752 0.8241693 0.7625310 0.7371429  
## 43.71878086 0.005395326 0.8362837 0.7970875 0.7426190  
## 43.71878086 0.011743628 0.8489714 0.8294199 0.7276190  
## 43.71878086 0.025561533 0.8438893 0.8701750 0.6650000  
## 43.71878086 0.055637998 0.8348766 0.9419469 0.5726190  
## 43.71878086 0.121103332 0.8328075 0.9725270 0.5157143  
## 43.71878086 0.263597138 0.8572641 0.9906903 0.4519048  
## 43.71878086 0.573753421 0.8982504 0.9995575 0.4033333  
## 43.71878086 1.248848869 0.9199254 1.0000000 0.3983333  
## 43.71878086 2.718281828 0.9112274 1.0000000 0.3983333  
## 132.80592990 0.002478752 0.8264273 0.7886745 0.7276190  
## 132.80592990 0.005395326 0.8388226 0.8143658 0.7126190  
## 132.80592990 0.011743628 0.8521603 0.8528889 0.6945238  
## 132.80592990 0.025561533 0.8330281 0.8998604 0.6059524  
## 132.80592990 0.055637998 0.8245095 0.9548083 0.5540476  
## 132.80592990 0.121103332 0.8286380 0.9694277 0.5300000  
## 132.80592990 0.263597138 0.8547695 0.9906922 0.4569048  
## 132.80592990 0.573753421 0.8986067 0.9995575 0.4033333  
## 132.80592990 1.248848869 0.9196614 1.0000000 0.3983333  
## 132.80592990 2.718281828 0.9116876 1.0000000 0.3983333  
## 403.42879349 0.002478752 0.8391130 0.8046214 0.7373810  
## 403.42879349 0.005395326 0.8482195 0.8298663 0.7230952  
## 403.42879349 0.011743628 0.8391176 0.8653097 0.6885714  
## 403.42879349 0.025561533 0.8282799 0.9246608 0.5821429  
## 403.42879349 0.055637998 0.8135166 0.9587965 0.5492857  
## 403.42879349 0.121103332 0.8255903 0.9694317 0.5395238  
## 403.42879349 0.263597138 0.8542765 0.9902517 0.4619048  
## 403.42879349 0.573753421 0.8979099 0.9986686 0.4130952  
## 403.42879349 1.248848869 0.9200789 1.0000000 0.3983333  
## 403.42879349 2.718281828 0.9113170 1.0000000 0.3983333  
##   
## ROC was used to select the optimal model using the largest value.  
## The final values used for the model were sigma = 1.248849 and C = 14.39192.

plot(svmr.fit, highlight = TRUE)



#evaluate svml model  
svml.pred = predict(svml.fit, newdata = test2, type = "prob")[,2]  
test.pred <- rep("No", length(svml.pred))  
test.pred[svml.pred>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 681 28  
## Yes 285 59  
##   
## Accuracy : 0.7028   
## 95% CI : (0.6741, 0.7302)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1635   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.67816   
## Specificity : 0.70497   
## Pos Pred Value : 0.17151   
## Neg Pred Value : 0.96051   
## Prevalence : 0.08262   
## Detection Rate : 0.05603   
## Detection Prevalence : 0.32669   
## Balanced Accuracy : 0.69156   
##   
## 'Positive' Class : Yes   
##

#evaluate svmr model  
svmr.pred = predict(svmr.fit, newdata = test2, type = "prob")[,2]  
test.pred <- rep("No", length(svmr.pred))  
test.pred[svmr.pred>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 966 12  
## Yes 0 75  
##   
## Accuracy : 0.9886   
## 95% CI : (0.9802, 0.9941)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9198   
##   
## Mcnemar's Test P-Value : 0.001496   
##   
## Sensitivity : 0.86207   
## Specificity : 1.00000   
## Pos Pred Value : 1.00000   
## Neg Pred Value : 0.98773   
## Prevalence : 0.08262   
## Detection Rate : 0.07123   
## Detection Prevalence : 0.07123   
## Balanced Accuracy : 0.93103   
##   
## 'Positive' Class : Yes   
##

## QDA

#qda  
set.seed(2)  
  
ctrl1 = trainControl(method = "cv",sampling = "up",  
 summaryFunction = twoClassSummary,  
 classProbs = TRUE)  
  
model.qda = train(diabetes~., data = train2,  
 method = "qda",  
 preProcess = c("center","scale"),  
 trControl = ctrl1)

## Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not  
## in the result set. ROC will be used instead.

#evaluate qda model  
qda.pred = predict(model.qda, newdata = test2, type = "prob")[,2]  
test.pred <- rep("No", length(qda.pred))  
test.pred[qda.pred>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 680 16  
## Yes 286 71  
##   
## Accuracy : 0.7132   
## 95% CI : (0.6848, 0.7404)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.2156   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.81609   
## Specificity : 0.70393   
## Pos Pred Value : 0.19888   
## Neg Pred Value : 0.97701   
## Prevalence : 0.08262   
## Detection Rate : 0.06743   
## Detection Prevalence : 0.33903   
## Balanced Accuracy : 0.76001   
##   
## 'Positive' Class : Yes   
##

## MARS

ctrl2 <- trainControl(method = "cv", sampling = 'up')  
ctrl3 <- trainControl(method = "cv", sampling = 'up',  
 summaryFunction = twoClassSummary,  
 classProbs = TRUE)  
  
library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(earth)

## Loading required package: Formula

## Loading required package: plotmo

## Loading required package: plotrix

## Loading required package: TeachingDemos

set.seed(2)  
  
mars\_grid <- expand.grid(degree = 1:5,   
 nprune = 2:20)  
mars.fit <- train(diabetes ~., data = train2,  
 method = "earth",  
 tuneGrid = mars\_grid,  
 trControl = ctrl2)

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

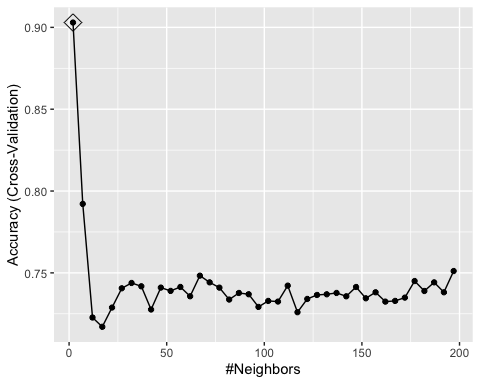
test.pred.prob.mars <- predict(mars.fit, newdata = test2, type = "prob")[,2]  
test.pred <- rep("No", length(test.pred.prob.mars))  
test.pred[test.pred.prob.mars>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 738 33  
## Yes 228 54  
##   
## Accuracy : 0.7521   
## 95% CI : (0.7249, 0.778)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1905   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.62069   
## Specificity : 0.76398   
## Pos Pred Value : 0.19149   
## Neg Pred Value : 0.95720   
## Prevalence : 0.08262   
## Detection Rate : 0.05128   
## Detection Prevalence : 0.26781   
## Balanced Accuracy : 0.69233   
##   
## 'Positive' Class : Yes   
##

Accuracy : 0.7521

## KNN

set.seed(2)  
knn.fit <- train(diabetes ~., data = train2,  
 method = "knn",  
 preProcess = c("center","scale"),  
 tuneGrid = data.frame(k = seq(2,200,by=5)),  
 trControl = ctrl2)  
  
ggplot(knn.fit, highlight = TRUE)



test.pred.prob.knn <- predict(knn.fit, newdata = test2,  
 type = "prob")[,2]  
test.pred <- rep("No", length(test.pred.prob.knn))  
test.pred[test.pred.prob.knn>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 914 37  
## Yes 52 50  
##   
## Accuracy : 0.9155   
## 95% CI : (0.897, 0.9316)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : 0.6157   
##   
## Kappa : 0.483   
##   
## Mcnemar's Test P-Value : 0.1378   
##   
## Sensitivity : 0.57471   
## Specificity : 0.94617   
## Pos Pred Value : 0.49020   
## Neg Pred Value : 0.96109   
## Prevalence : 0.08262   
## Detection Rate : 0.04748   
## Detection Prevalence : 0.09687   
## Balanced Accuracy : 0.76044   
##   
## 'Positive' Class : Yes   
##

Accuracy : 0.9155

##LDA

set.seed(2)  
lda.fit <- train(diabetes ~., data = train2, preProcess = c("center","scale"),  
 method = "lda",  
 trControl = ctrl3)

## Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not  
## in the result set. ROC will be used instead.

test.pred.prob.lda <- predict(lda.fit, newdata = test2, type = "prob")[,2]  
test.pred <- rep("No", length(test.pred.prob.lda))  
test.pred[test.pred.prob.lda>0.5] <- "Yes"  
  
##Creating the confusion matrix:  
confusionMatrix(data = as.factor(test.pred),  
 reference = test2$diabetes,  
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 668 23  
## Yes 298 64  
##   
## Accuracy : 0.6952   
## 95% CI : (0.6664, 0.7229)  
## No Information Rate : 0.9174   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1752   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.73563   
## Specificity : 0.69151   
## Pos Pred Value : 0.17680   
## Neg Pred Value : 0.96671   
## Prevalence : 0.08262   
## Detection Rate : 0.06078   
## Detection Prevalence : 0.34378   
## Balanced Accuracy : 0.71357   
##   
## 'Positive' Class : Yes   
##

Accuracy : 0.6952

##Plotting the ROC curves to determine the best model, printing the AUC.

set.seed(2)  
  
roc.svml <- roc(test2$diabetes, svml.pred)

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

roc.svmr <- roc(test2$diabetes, svmr.pred)

## Setting levels: control = No, case = Yes  
## Setting direction: controls < cases

roc.qda <- roc(test2$diabetes, qda.pred)

## Setting levels: control = No, case = Yes  
## Setting direction: controls < cases

roc.MARS <- roc(test2$diabetes, test.pred.prob.mars)

## Setting levels: control = No, case = Yes  
## Setting direction: controls < cases

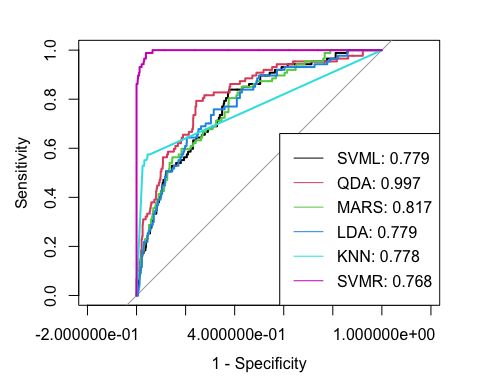
roc.lda <- roc(test2$diabetes, test.pred.prob.lda)

## Setting levels: control = No, case = Yes  
## Setting direction: controls < cases

roc.knn <- roc(test2$diabetes, test.pred.prob.knn)

## Setting levels: control = No, case = Yes  
## Setting direction: controls < cases

auc <- c(roc.svml$auc[1],roc.svmr$auc[1],roc.qda$auc[1], roc.MARS$auc[1], roc.lda$auc[1], roc.knn$auc[1])  
  
plot(roc.svml, legacy.axes = TRUE, col = 1)  
plot(roc.qda, legacy.axes = TRUE, add=TRUE, col = 2)  
plot(roc.MARS, legacy.axes = TRUE, add=TRUE, col = 3)  
plot(roc.lda, legacy.axes = TRUE, add=TRUE, col = 4)  
plot(roc.knn, legacy.axes = TRUE, add=TRUE, col = 5)  
plot(roc.svmr, legacy.axes = TRUE, add=TRUE, col = 6)  
models.used <- c("SVML","QDA","MARS", "LDA", "KNN", "SVMR")   
  
legend("bottomright", legend = paste0(models.used, ": ", round(auc,3)), col = 1:6, lwd = 1)



## Model Comparisons

resamp2 = resamples(list(svml = svml.fit, svmr = svmr.fit, qda = model.qda, mars = mars.fit, knn = knn.fit, lda = lda.fit))

## Warning in resamples.default(list(svml = svml.fit, svmr = svmr.fit, qda =  
## model.qda, : Some performance measures were not computed for each model:  
## Accuracy, Kappa, ROC, Sens, Spec

summary(resamp2)

##   
## Call:  
## summary.resamples(object = resamp2)  
##   
## Models: svml, svmr, qda, mars, knn, lda   
## Number of resamples: 10