|  |
| --- |
| >  > algorithmList <- c('rf', 'knn', 'earth', 'ada', 'xgbDART', 'svmRadial', 'svmLinear')  > set.seed(100)  > models <- caretList(class ~ .,  + data=trainData,  + trControl=trainControl,  + metric = "ROC",  + methodList=algorithmList  + )  There were 47 warnings (use warnings() to see them)  > save(models, file = "0712\_10Genes\_7Models.RData")  > results <- resamples(models)  > summary(results)  Call:  summary.resamples(object = results)  Models: rf, knn, earth, ada, xgbDART, svmRadial, svmLinear  Number of resamples: 50  ROC  Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  rf 0.9829282 0.9932837 0.9982509 0.9961302 1.0000000 1 0  knn 0.9635417 0.9852870 0.9976359 0.9917088 0.9996337 1 0  earth 0.9698582 0.9947437 0.9991414 0.9954255 1.0000000 1 0  ada 0.9746981 0.9948286 0.9982270 0.9958125 0.9998553 1 0  xgbDART 0.9825650 0.9958844 0.9985464 0.9967288 1.0000000 1 0  svmRadial 0.9710402 0.9858895 0.9982639 0.9936760 1.0000000 1 0  svmLinear 0.9704861 0.9930556 0.9985464 0.9945728 1.0000000 1 0  Sens  Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  rf 0.9361702 0.9787234 0.9791667 0.9839450 1 1 0  knn 0.9361702 0.9583333 0.9787234 0.9788652 1 1 0  earth 0.9574468 0.9787234 0.9791667 0.9864716 1 1 0  ada 0.9361702 0.9583333 0.9787234 0.9784574 1 1 0  xgbDART 0.9361702 0.9787234 0.9791667 0.9877482 1 1 0  svmRadial 0.9361702 0.9787234 0.9791667 0.9835106 1 1 0  svmLinear 0.9375000 0.9787234 0.9791667 0.9856383 1 1 0  Spec  Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  rf 0.9166667 0.9722222 0.9725976 0.9801051 1 1 0  knn 0.9444444 0.9722222 0.9729730 0.9801201 1 1 0  earth 0.9444444 0.9722222 0.9722222 0.9778979 1 1 0  ada 0.9444444 0.9722222 0.9725976 0.9801051 1 1 0  xgbDART 0.8648649 0.9722222 0.9722222 0.9762613 1 1 0  svmRadial 0.9166667 0.9722222 0.9729730 0.9784384 1 1 0  svmLinear 0.9166667 0.9722222 0.9722222 0.9795495 1 1 0  > scales <- list(x=list(relation="free"), y=list(relation="free"))  > bwplot(results, scales=scales)  > bwplot(results, layout=c(3,1))  > dotplot(results, metric = "ROC")  > xyplot(results, what = "BlandAltman")  > splom(results)  > rocobj\_models <- roc(models$rf$pred$obs,  + models$rf$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="darkblue", lwd=4,  + print.auc=TRUE,  + print.auc.y = 65  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > ppv\_npv\_rf <- t( coords( smooth( rocobj\_models ),  + x = 0.9,  + input = "sensitivity",  + ret = c("specificity", "sensitivity", "ppv", "npv") ))  Warning message:  In coords.smooth.roc(smooth(rocobj\_models), x = 0.9, input = "sensitivity", :  An upcoming version of pROC will set the 'transpose' argument to FALSE by default. Set transpose = TRUE explicitly to keep the current behavior, or transpose = FALSE to adopt the new one and silence this warning. Type help(coords\_transpose) for additional information.  > rocobj\_models <- roc(models$svmRadial$pred$obs,  + models$svmRadial$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="green", lwd=4,  + print.auc=TRUE,  + print.auc.y = 60,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > t( coords( smooth( rocobj\_models ),  + x = 0.9,  + input = "sensitivity",  + ret = c("specificity", "sensitivity", "ppv", "npv") ))  specificity sensitivity ppv npv  [1,] 99.99995 0.9 99.99255 56.86858  Warning message:  In coords.smooth.roc(smooth(rocobj\_models), x = 0.9, input = "sensitivity", :  An upcoming version of pROC will set the 'transpose' argument to FALSE by default. Set transpose = TRUE explicitly to keep the current behavior, or transpose = FALSE to adopt the new one and silence this warning. Type help(coords\_transpose) for additional information.  > rocobj\_models <- roc(models$svmLinear$pred$obs,  + models$svmLinear$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="red", lwd=4,  + print.auc=TRUE,  + print.auc.y = 55,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > rocobj\_models <- roc(models$xgbDART$pred$obs,  + models$xgbDART$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="black", lwd=4,  + print.auc=TRUE,  + print.auc.y = 50,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > rocobj\_models <- roc(models$earth$pred$obs,  + models$earth$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="yellow", lwd=4,  + print.auc=TRUE,  + print.auc.y = 45,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > rocobj\_models <- roc(models$knn$pred$obs,  + models$knn$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="pink", lwd=4,  + print.auc=TRUE,  + print.auc.y = 40,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > rocobj\_models <- roc(models$ada$pred$obs,  + models$ada$pred$tumor,  + ci=TRUE,  + plot=TRUE,  + legacy.axes=TRUE, percent=TRUE,  + xlab="False Positive Percentage",  + ylab="True Postive Percentage",  + col="brown", lwd=4,  + print.auc=TRUE,  + print.auc.y = 35,  + add = TRUE  + )  Setting levels: control = normal, case = tumor  Setting direction: controls < cases  > legend("bottomright",  + legend=c( "rf", "svmRadial", "svmLinear", "xgbDART", "MARS", "knn", "ada" ),  + col=c( "darkblue", "green", "red", "black", "yellow", "pink", "brown" ),  + lwd=4  + )  > predict\_rf <- predict(models$rf, testData)  > conf\_rf <- confusionMatrix(reference = testData$class, data = predict\_rf, mode='everything', positive='tumor')  > predict\_knn <- predict(models$knn, testData)  > conf\_knn <- confusionMatrix(reference = testData$class, data = predict\_knn, mode='everything', positive='tumor')  > predict\_mars <- predict(models$earth, testData)  > conf\_mars <- confusionMatrix(reference = testData$class, data = predict\_mars, mode='everything', positive='tumor')  > predict\_ada <- predict(models$ada, testData)  > conf\_ada <- confusionMatrix(reference = testData$class, data = predict\_ada, mode='everything', positive='tumor')  > predict\_xgbDART <- predict(models$xgbDART, testData)  > conf\_xgbDART <- confusionMatrix(reference = testData$class, data = predict\_xgbDART, mode='everything', positive='tumor')  > predict\_svmRadial <- predict(models$svmRadial, testData)  > conf\_svmRadial <- confusionMatrix(reference = testData$class, data = predict\_svmRadial, mode='everything', positive='tumor')  > predict\_svmLinear <- predict(models$svmLinear, testData)  > conf\_svmLinear <- confusionMatrix(reference = testData$class, data = predict\_svmLinear, mode='everything', positive='tumor')  > conf\_rf  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 195 5  tumor 7 149    Accuracy : 0.9663  95% CI : (0.9419, 0.9825)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.9314    Mcnemar's Test P-Value : 0.7728    Sensitivity : 0.9675  Specificity : 0.9653  Pos Pred Value : 0.9551  Neg Pred Value : 0.9750  Precision : 0.9551  Recall : 0.9675  F1 : 0.9613  Prevalence : 0.4326  Detection Rate : 0.4185  Detection Prevalence : 0.4382  Balanced Accuracy : 0.9664    'Positive' Class : tumor    > conf\_knn  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 192 2  tumor 10 152    Accuracy : 0.9663  95% CI : (0.9419, 0.9825)  No Information Rate : 0.5674  P-Value [Acc > NIR] : < 2e-16    Kappa : 0.9318    Mcnemar's Test P-Value : 0.04331    Sensitivity : 0.9870  Specificity : 0.9505  Pos Pred Value : 0.9383  Neg Pred Value : 0.9897  Precision : 0.9383  Recall : 0.9870  F1 : 0.9620  Prevalence : 0.4326  Detection Rate : 0.4270  Detection Prevalence : 0.4551  Balanced Accuracy : 0.9688    'Positive' Class : tumor    > conf\_mars  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 194 2  tumor 8 152    Accuracy : 0.9719  95% CI : (0.9489, 0.9864)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.943    Mcnemar's Test P-Value : 0.1138    Sensitivity : 0.9870  Specificity : 0.9604  Pos Pred Value : 0.9500  Neg Pred Value : 0.9898  Precision : 0.9500  Recall : 0.9870  F1 : 0.9682  Prevalence : 0.4326  Detection Rate : 0.4270  Detection Prevalence : 0.4494  Balanced Accuracy : 0.9737    'Positive' Class : tumor    > conf\_ada  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 192 3  tumor 10 151    Accuracy : 0.9635  95% CI : (0.9384, 0.9804)  No Information Rate : 0.5674  P-Value [Acc > NIR] : < 2e-16    Kappa : 0.926    Mcnemar's Test P-Value : 0.09609    Sensitivity : 0.9805  Specificity : 0.9505  Pos Pred Value : 0.9379  Neg Pred Value : 0.9846  Precision : 0.9379  Recall : 0.9805  F1 : 0.9587  Prevalence : 0.4326  Detection Rate : 0.4242  Detection Prevalence : 0.4522  Balanced Accuracy : 0.9655    'Positive' Class : tumor    > conf\_xgbDART  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 195 4  tumor 7 150    Accuracy : 0.9691  95% CI : (0.9454, 0.9845)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.9372    Mcnemar's Test P-Value : 0.5465    Sensitivity : 0.9740  Specificity : 0.9653  Pos Pred Value : 0.9554  Neg Pred Value : 0.9799  Precision : 0.9554  Recall : 0.9740  F1 : 0.9646  Prevalence : 0.4326  Detection Rate : 0.4213  Detection Prevalence : 0.4410  Balanced Accuracy : 0.9697    'Positive' Class : tumor    > conf\_svmLinear  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 195 4  tumor 7 150    Accuracy : 0.9691  95% CI : (0.9454, 0.9845)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.9372    Mcnemar's Test P-Value : 0.5465    Sensitivity : 0.9740  Specificity : 0.9653  Pos Pred Value : 0.9554  Neg Pred Value : 0.9799  Precision : 0.9554  Recall : 0.9740  F1 : 0.9646  Prevalence : 0.4326  Detection Rate : 0.4213  Detection Prevalence : 0.4410  Balanced Accuracy : 0.9697    'Positive' Class : tumor    > conf\_svmRadial  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 193 5  tumor 9 149    Accuracy : 0.9607  95% CI : (0.9349, 0.9783)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.9201    Mcnemar's Test P-Value : 0.4227    Sensitivity : 0.9675  Specificity : 0.9554  Pos Pred Value : 0.9430  Neg Pred Value : 0.9747  Precision : 0.9430  Recall : 0.9675  F1 : 0.9551  Prevalence : 0.4326  Detection Rate : 0.4185  Detection Prevalence : 0.4438  Balanced Accuracy : 0.9615    'Positive' Class : tumor    >  >  > set.seed(100)  > stackControl <- trainControl(method="repeatedcv",  + number=20,  + repeats=5,  + savePredictions=TRUE,  + classProbs=TRUE  + )  > stack.glm <- caretStack(models, method="glm", metric="Accuracy", trControl=stackControl)  > varImp( models$svmLinear )  ROC curve variable importance  Importance  hsa.miR.21.5p 100.00  hsa.miR.183.5p 94.07  hsa.miR.96.5p 82.59  hsa.miR.182.5p 81.02  hsa.miR.99a.5p 67.35  hsa.miR.204.5p 66.56  hsa.let.7c.5p 63.33  hsa.miR.139.5p 62.17  hsa.miR.181b.5p\_2 61.21  hsa.miR.181b.5p 61.09  hsa.miR.101.5p 0.00  > print(stack.glm)  A glm ensemble of 2 base models: rf, knn, earth, ada, xgbDART, svmRadial, svmLinear  Ensemble results:  Generalized Linear Model  4175 samples  7 predictor  2 classes: 'normal', 'tumor'  No pre-processing  Resampling: Cross-Validated (20 fold, repeated 5 times)  Summary of sample sizes: 3967, 3966, 3965, 3967, 3967, 3965, ...  Resampling results:  Accuracy Kappa  0.9840936 0.967603  > testData.t <- testData  > testData.t$class <- ifelse(testData.t$class=="tumor", "normal", "tumor")  > testData.t$class <- as.factor(testData.t$class)  > stack\_predicteds <- predict(stack.glm, newdata=testData.t)  > summary(testData.t$class)  normal tumor  154 202  > head(stack\_predicteds)  [1] normal normal normal normal normal normal  Levels: normal tumor  > confusionMatrix(reference = testData.t$class, data = stack\_predicteds, mode='everything', positive='tumor')  Confusion Matrix and Statistics  Reference  Prediction normal tumor  normal 151 6  tumor 3 196    Accuracy : 0.9747  95% CI : (0.9526, 0.9884)  No Information Rate : 0.5674  P-Value [Acc > NIR] : <2e-16    Kappa : 0.9486    Mcnemar's Test P-Value : 0.505    Sensitivity : 0.9703  Specificity : 0.9805  Pos Pred Value : 0.9849  Neg Pred Value : 0.9618  Precision : 0.9849  Recall : 0.9703  F1 : 0.9776  Prevalence : 0.5674  Detection Rate : 0.5506  Detection Prevalence : 0.5590  Balanced Accuracy : 0.9754    'Positive' Class : tumor |
|  |
| |  | | --- | | > | |