>> DS2 Ensemble

Call:

summary.resamples(object = results)

Models: svmRadial, rf, J48, AdaBoost.M1

Number of resamples: 25

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

svmRadial 0.8993576 0.9079229 0.9186296 0.9166104 0.9229122 0.9399142 0

rf 0.9036403 0.9143469 0.9206009 0.9203810 0.9252137 0.9549356 0

J48 0.8586724 0.8650964 0.8736617 0.8774816 0.8803419 0.9102564 0

AdaBoost.M1 0.8907923 0.9059829 0.9143469 0.9172966 0.9250535 0.9592275 0

Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

svmRadial 0.7987198 0.8158737 0.8372434 0.8332203 0.8458152 0.8798283 0

rf 0.8073238 0.8287056 0.8412017 0.8407618 0.8504274 0.9098712 0

J48 0.7173383 0.7302137 0.7472967 0.7549602 0.7606838 0.8205128 0

AdaBoost.M1 0.7815656 0.8119658 0.8287056 0.8345925 0.8501133 0.9184549 0

> dotplot(results)

>

>

> # correlation between results

> modelCor(results)

svmRadial rf J48 AdaBoost.M1

svmRadial 1.0000000 0.7023120 0.3602258 0.6893022

rf 0.7023120 1.0000000 0.4682178 0.7157660

J48 0.3602258 0.4682178 1.0000000 0.5101970

AdaBoost.M1 0.6893022 0.7157660 0.5101970 1.0000000

> splom(results)

>

> #individual prediction of each method on data

> p <- as.data.frame(predict(models, newdata=head(Exdata[,predictors])))

> print(p)

svmRadial rf J48 AdaBoost.M1

1 0.985880420 0.842 0.93815988 0.6667966

2 0.002746197 0.110 0.07395143 0.4384586

3 0.527787886 0.512 0.86666667 0.4658201

4 0.991822334 0.962 0.93815988 0.6575101

5 0.999304927 0.966 0.93815988 0.6417618

6 0.873493516 0.748 0.93815988 0.6058318

>

>

> # stack using random forest

> set.seed(3233)

> stack.rf <- caretStack(models, method="rf", metric="Accuracy", trControl=control)

> print(stack.rf)

A rf ensemble of 2 base models: svmRadial, rf, J48, AdaBoost.M1

Ensemble results:

Random Forest

11680 samples

4 predictor

2 classes: 'neg', 'pos'

No pre-processing

Resampling: Cross-Validated (5 fold, repeated 5 times)

Summary of sample sizes: 9344, 9344, 9344, 9344, 9344, 9344, ...

Resampling results across tuning parameters:

mtry Accuracy Kappa

2 0.9232021 0.8464041

3 0.9226712 0.8453425

4 0.9219863 0.8439726

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was mtry = 2.

>

> # validation

> library("caTools")

Attaching package: ‘caTools’

The following object is masked from ‘package:RWeka’:

LogitBoost

Warning message:

package ‘caTools’ was built under R version 3.4.4

> model\_preds <- lapply(models, predict, newdata=Exdata[,predictors], type="prob")

> #model\_preds <- lapply(model\_preds, function(x) x[,"M"])

> model\_preds <- data.frame(model\_preds)

>

> model\_preds$ensemble <- predict(stack.rf,newdata=Exdata[,predictors],type="prob")

> #CF <- coef(stack.rf$ens\_model$finalModel)[,outcomeName]

> CF <- coef(stack.rf$models$svmRadial)[,-1]

> CF

NULL

> colAUC(model\_preds, Exdata[,outcomeName])

svmRadial.neg svmRadial.pos rf.neg rf.pos J48.neg J48.pos

neg vs. pos 0.962282 0.962282 0.9678676 0.9678676 0.8789339 0.8789339

AdaBoost.M1.neg AdaBoost.M1.pos ensemble

neg vs. pos 0.9532835 0.9532835 0.9658188

> CF

NULL

>

>

> colAUC(model\_preds$ensemble,Exdata[,outcomeName])

[,1]

neg vs. pos 0.9658188