|  |
| --- |
| > setwd("E://courses//jh-dataexplore//ws\_machine")  > d <- read.csv("pml-training.csv")  >  > colD <- read.table("corColumn.txt")  >  > d.sub <- subset(d, select=c(colD$V1))  >  > # give the subset data meaningful column names  > names(d.sub) <- colD$V2  > library(caret)  Loading required package: lattice  Loading required package: ggplot2  > inTrain <- createDataPartition(y=d.sub$classe,p=0.5,list=FALSE)  > training <- d.sub[inTrain,]  > testing <- d.sub[-inTrain,]  > Sys.time()  [1] "2014-07-21 14:36:38 EST"  > mod.gbm <- train(classe ~ ., method="gbm", data=training, verbose=FALSE)  Loading required package: gbm  Loading required package: survival  Loading required package: splines  Attaching package: ‘survival’  The following object is masked from ‘package:caret’:  cluster  Loading required package: parallel  Loaded gbm 2.1  Loading required package: plyr  Warning messages:  1: package ‘gbm’ was built under R version 3.1.1  2: package ‘e1071’ was built under R version 3.1.1  > Sys.time()  [1] "2014-07-21 14:49:20 EST"  > mod.rf <- train(classe ~ ., method="rf", data=training, trControl = trainControl(method="cv"), number=3)  Loading required package: randomForest  randomForest 4.6-10  Type rfNews() to see new features/changes/bug fixes.  Warning message:  package ‘randomForest’ was built under R version 3.1.1  > Sys.time()  [1] "2014-07-21 14:58:56 EST"  >  >  >  > pred.gbm <- predict(mod.gbm, testing)  > pred.rf <- predict(mod.rf, testing)  > qplot(pred.gbm, colour=classe, fill=classe, data=testing) |
| |  | | --- | | table.gbm <- table(pred.gbm, testing$classe)  > table.rf <- table(pred.rf, testing$classe)  >  > prop.table(table.rf,2)    pred.rf A B C D E  A 0.9982078853 0.0084299262 0.0000000000 0.0000000000 0.0022185247  B 0.0017921147 0.9826132771 0.0105201636 0.0000000000 0.0000000000  C 0.0000000000 0.0084299262 0.9818819404 0.0304726368 0.0016638935  D 0.0000000000 0.0000000000 0.0075978960 0.9682835821 0.0033277870  E 0.0000000000 0.0005268704 0.0000000000 0.0012437811 0.9927897948 | |  | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | diag.gbm <- diag(prop.table(table.gbm,2))  > diag.rf <- diag(prop.table(table.rf,2))  >  diag.comb <- cbind(diag.gbm, diag.rf)  > #par(mfrow=c(1,1))  > with(testing, plot(diag.comb, pch=19, cex=1, col="blue"))  > abline(0,1,lwd=2, col='red')    inTrain <- createDataPartition(y=d.sub$classe,p=0.7,list=FALSE)  > training <- d.sub[inTrain,]  > testing <- d.sub[-inTrain,]  > Sys.time()  [1] "2014-07-21 15:09:34 EST"  > mod.rf7 <- train(classe ~ ., method="rf", data=training, trControl = trainControl(method="cv"), number=3)  > Sys.time()  [1] "2014-07-21 15:23:21 EST"  >  > pred.rf7 <- predict(mod.rf7, testing)  > table.rf7 <- table(pred.rf7, testing$classe)  > diag.rf7 <- diag(prop.table(table.rf7, 2))  >  > diag.rfcomb <- cbind(diag.rf, diag.rf7)    t <- read.csv("pml-testing.csv")  > t.sub <- subset(t, select=c(colD$V1))  >  > # give the subset data meaningful column names  > names(t.sub) <- colD$V2  >  > t.sub$number <- c(1:20)   |  | | --- | | > for (i in 1:20) { print(paste(i,predict(mod.rf7, t.sub[t.sub$number==i,]), sep=" " )) }  [1] "1 B"  [1] "2 A"  [1] "3 B"  [1] "4 A"  [1] "5 A"  [1] "6 E"  [1] "7 D"  [1] "8 B"  [1] "9 A"  [1] "10 A"  [1] "11 B"  [1] "12 C"  [1] "13 B"  [1] "14 A"  [1] "15 E"  [1] "16 E"  [1] "17 A"  [1] "18 B"  [1] "19 B"  [1] "20 B" | |  | | |  | | --- | | > | | | | |
| |  | | --- | | > | |