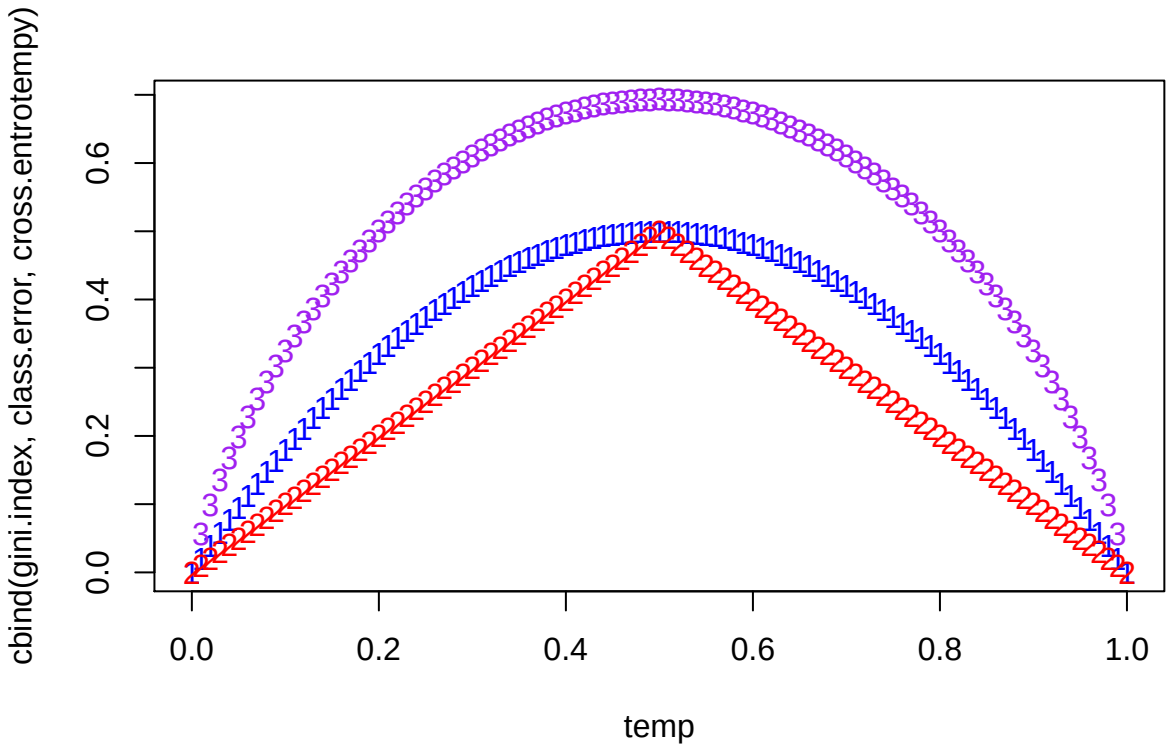
MA679 Hw6

Chaoqun Yin

4/15/2019

8.3

```
temp <- seq(0, 1, 0.01)
gini.index <- 2 * temp * (1 - temp)
class.error <- 1 - pmax(temp, 1 - temp)
cross.entrotempy <- - (temp * log(temp) + (1 - temp) * log(1 - temp))
matplot(temp, cbind(gini.index, class.error, cross.entrotempy), col = c("blue", "red", "purple"))
```



8.5

```
#1. We classify X as Red as it is the most commonly occurring class among the 10 predictions.
#2. With the average probability approach, we classify X as Green as the average of the 10 probabilitie
```

8.7

```
library(MASS)
library(randomForest)
```

```
## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.
```

```r
set.seed(0)
train <- sample(1:nrow(Boston), nrow(Boston) / 2)
Boston.train <- Boston[train, -14]
Boston.test <- Boston[-train, -14]
Y.train <- Boston[train, 14]
Y.test <- Boston[-train, 14]
rf.boston1 <- randomForest(Boston.train, y = Y.train, xtest = Boston.test, ytest = Y.test, mtry = ncol(
rf.boston2 <- randomForest(Boston.train, y = Y.train, xtest = Boston.test, ytest = Y.test, mtry = (ncol
rf.boston3 <- randomForest(Boston.train, y = Y.train, xtest = Boston.test, ytest = Y.test, mtry = sqrt(
plot(1:500, rf.boston1$test$mse, col = "green", type = "l", xlab = "Number of Trees", ylab = "Test MSE"
lines(1:500, rf.boston2$test$mse, col = "red", type = "l")
lines(1:500, rf.boston3$test$mse, col = "blue", type = "l")
legend("topright", c("m = p", "m = p/2", "m = sqrt(p)"), col = c("green", "red", "blue"), cex = 1, lty =
```