Question3

Yufei Yin

3.

(a)

```
library(gbm)
set.seed(452)
```

```
################################
### Import and process data ###
################################

### Import and clean the air quality data
data("airquality")
AQ.raw = na.omit(airquality[,1:4])
AQ = AQ.raw
```

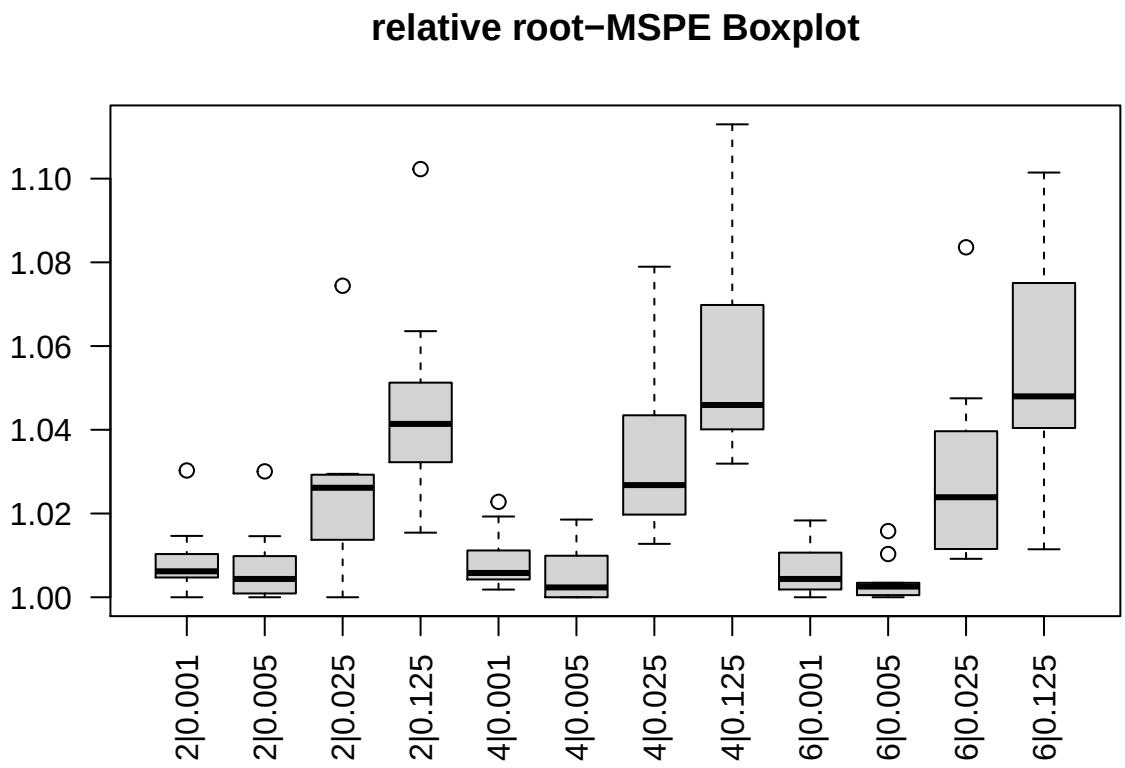
```
V=5
R=2
n2 = nrow(AQ)
# Create the folds and save in a matrix
folds = matrix(NA, nrow=n2, ncol=R)
for(r in 1:R){
        folds[,r]=floor((sample.int(n2)-1)*V/n2) + 1
}

shr = c(.001,.005,.025,.125)
dep = c(2,4,6)
trees = 10000

NS = length(shr)
ND = length(dep)
gb.cv = matrix(NA, nrow=ND*NS, ncol=V*R)
opt.tree = matrix(NA, nrow=ND*NS, ncol=V*R)

qq = 1
for(r in 1:R){
  for(v in 1:V){
    pro.train = AQ[folds[,r]!=v,]
    pro.test = AQ[folds[,r]==v,]
    counter=1
    for(d in dep){
      for(s in shr){
        pro.gbm <- gbm(data=pro.train, Ozone~., distribution="gaussian",
                       n.trees=trees, interaction.depth=d, shrinkage=s,
```

```
                        bag.fraction=0.8)
        treenum = min(trees, 2*gbm.perf(pro.gbm, method="OOB", plot.it=FALSE))
        opt.tree[counter,qq] = treenum
        preds = predict(pro.gbm, newdata=pro.test, n.trees=treenum)
        gb.cv[counter,qq] = mean((preds - AQ$Ozone)^2)
        counter=counter+1
      }
    }
    qq = qq+1
  }
}

parms = expand.grid(shr,dep)
row.names(gb.cv) = paste(parms[,2], parms[,1], sep="|")

# mean root-MSPE
(mean.cv = apply(sqrt(gb.cv), 1, mean))
```

```
##  2|0.001  2|0.005  2|0.025  2|0.125  4|0.001  4|0.005  4|0.025  4|0.125
## 42.32242 42.27408 43.08109 43.92213 42.31898 42.16811 43.35583 44.44169
##  6|0.001  6|0.005  6|0.025  6|0.125
## 42.24067 42.11750 43.26447 44.36027
```

(b)

```
# relative root-MSPE boxplots
min.cv = apply(gb.cv, 2, min)
boxplot(sqrt(t(gb.cv)/min.cv), use.cols=TRUE, las=2,
        main= "relative root-MSPE Boxplot")
```



(c)

According to the relative root-MSPE boxplot, i prefer the combination of $\lambda = 0.005$ and $d = 6$.