Group Assignment 8

Group 5

2022-11-13

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

8.4 Exercises Problem #12

```
library(randomForest)
```

```
## Warning: 程辑包'randomForest'是用R版本4.2.2 来建造的
```

```
## randomForest 4.7-1.1
```

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

```
library(gbm)
```

```
## Warning: 程辑包'gbm'是用R版本4.2.2 来建造的
```

```
## Loaded gbm 2.1.8.1
```

```
library(MASS)
set.seed(100)

train = sample(1:nrow(Boston),nrow(Boston)/2)
```

```
#boosting
set.seed(100)

boost.boston = gbm(medv~.,Boston[train,],distribution = "gaussian", n.trees=5000,interaction.depth=4)

pred.boost = predict(boost.boston, newdata = Boston[-train,])
```

```
## Using 5000 trees...
```

```
mean((pred.boost-Boston[-train,"medv"])^2)
```

```
## [1] 14.5495
```

```
#bagging
set.seed(100)

bag.boston = randomForest(medv~., data=Boston, subset=train, mtry=13, importance=TRUE)

pred.bag = predict(bag.boston, newdata = Boston[-train,])

mean((pred.bag-Boston[-train,"medv"])^2)
```

```
## [1] 16.37741
```

```
#random forest
set.seed(100)

bag.boston = randomForest(medv~., data=Boston, subset=train, mtry=4, importance=TRUE)

pred.bag = predict(bag.boston, newdata = Boston[-train,])

mean((pred.bag-Boston[-train,"medv"])^2)
```

```
## [1] 14.16742
```

```
#linear regression
set.seed(100)

lm.boston = lm(medv~., data = Boston, subset = train)

pred.lm = predict(lm.boston, newdata = Boston[-train,])

mean((pred.lm-Boston[-train,"medv"])^2)
```

```
## [1] 26.55316
```

We can find that compared to simple methods like linear regression, boosting, random forest and bagging both have greater accuracy, and they test MSE are lower than taht of linear model.