

Course 2 Section 4.6 - REGRESSION TREES

Jiaying Wu

17/10/2020

Give it a go!

Continue to develop your skills and understanding of regression trees by making your way through this exercise.

For this exercise use the following data set.

```
library(tibble)
d <- tibble(x=c(1.5, 2.8, 4.1, 2.0, 5.2), y=c(6.2, 4.8, 5.1, 7.1, 3.7))
d
```

```
## # A tibble: 5 x 2
##       x     y
##   <dbl> <dbl>
## 1  1.5   6.2
## 2  2.8   4.8
## 3  4.1   5.1
## 4  2     7.1
## 5  5.2   3.7
```

Q1. Sketch the data on a piece of paper. How many possible splits are there in this data?

```
p <- ncol(d)
n <- nrow(d)
(p-1)*(n-1)
```

```
## [1] 4
```

Q2. Use the function provided in this step to compute the ANOVA criterion for each split.

```
SST <- sum((d$y - mean(d$y))^2)
SSG <- NA

for(i in 1:(n-1)){
  SSG1 <- sum((d$y[1:i] - mean(d$y[1:i]))^2)
  SSG2 <- sum((d$y[-c(1:i)] - mean(d$y[-c(1:i)]))^2)
  SSG[i] <- SSG1 + SSG2
  message("i=", i, ", SSG1=", SSG1, ", SSG2=", SSG2, ", SSG1+SSG2=", SSG[i], ", ANOVA=", SST-SSG[i])
}
SST - SSG
```

```
## [1] 0.840500000 0.048000000 0.001333333 3.528000000
```