서울의 땅값을 결정하는 요인은?

통계학과 정희정 한지현
허은정 홍승범

[오늘의 부동산 이슈]13년 연속 전국 땅값 1위에 등극한 곳은 어디?

'1분기 토지거래량' 11년만에 최대

강성휘기 ◇조상 땅 찾기, 지난해 12.7만명 이용 역대 최다 0:00

30일 국토부에 따[전국N] 교통좋고 개발호재까지…금천구의 '비상'

만7348명이 약 40· 한국경제TV | 2017.10.13. | 네이버뉴스

난해에 **"땅값 오른다"고 속여 60억 챙긴 기획부동산 업자 검거**

연간 초 (전주=연합뉴스) 임채두 기자 = 16일 오전 전북지방경찰청 기자실에서

뉴스1 | 수법을 설명하고 있다

치단체의 지역현안과 업을 확대 추진하기로

"새만금 주변 땅값 오른다" 투자금 가로챈 50대 구속 SBS 뉴스 | 2015.12.02.

집 옆에 '경찰서'가 생기면 집값이 떨어진다?

헤클

[뉴스 투데이] 장애인시설 생기면 땅값 떨어진다?…설 곳 없는 특수학교

세계일보 | A2면 TOP | 2017.09.14. | 네이버뉴스

목차

변수 탐색 – 변수 정의 및 변환

설명변수

숙박업소 | 유흥업소 | 학군 * 10^6

특수학교 | 대형병원 | 대형마트 0 or 1

법정동별 개수 / 단위면적

개별공시지가

Log()

왜 법정동...?

1. 구 vs 동
세밀한 범위

2. 행정동 vs 법정동
지역 특성 반영

표준지공시지가를 바탕으로
토지의 특성에 따라
가중치를 부여한 종합적 지표
단위면적당 토지의 가격

변수 탐색 - 시각화


학교
강남구
강동구
강북구
강서구
관악구
광진구
구로구
금천구
노원구
도봉구
동대문구
동작구
마포구
서대문구
서초구
성동구
성북구
송파구
양천구
영등포구
용산구
은평구
종로구
중구
중랑구

모델 설정 – train & test data 생성

8 : 2의 비율로 train data 와 test 생성

```
dat <- dat[, -c(1, 9, 10)]
set.seed(321)
randomindex <- sample(1:nrow(dat), size=round(nrow(dat)*0.8), replace=F)
train_x <- dat[randomindex, -1]
test_x <- dat[-randomindex, -1]

train_y <- dat[randomindex, 1]
test_y <- dat[-randomindex, 1]

traindat <- data.frame(train_x, train_y)
testdat <- data.frame(test_x, test_y)
```

모델 설정 - 회귀분석

Backward Selection Model MSE = 0.310195

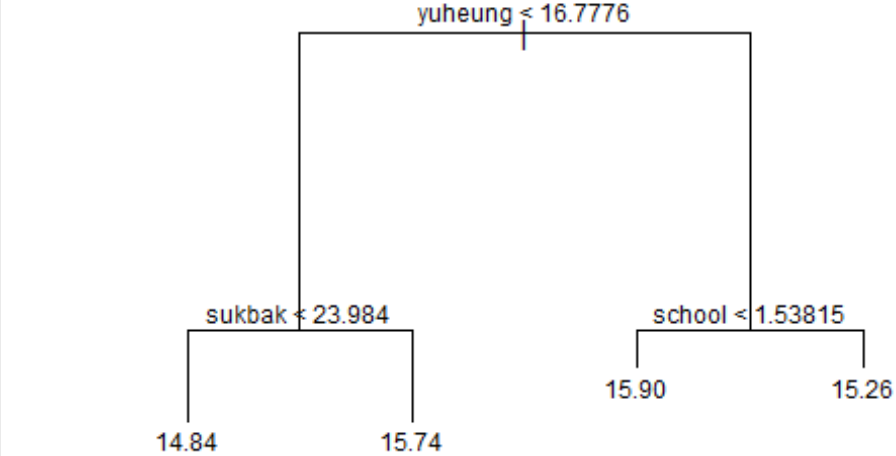
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.0889819  0.0425764 354.398  < 2e-16 ***
yuheung      0.0013949  0.0006898   2.022  0.04388 *
sukbak       0.0026885  0.0011505   2.337  0.01998 *
specialsch  -0.4949437  0.1623085  -3.049  0.00246 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7357 on 370 degrees of freedom
Multiple R-squared:  0.1293, Adjusted R-squared:  0.1223
F-statistic: 18.32 on 3 and 370 DF,  p-value: 4.187e-11

Multiple R-squared:  0.1383, Adjusted R-squared:  0.1242
F-statistic: 9.815 on 6 and 367 DF,  p-value: 4.837e-10
```

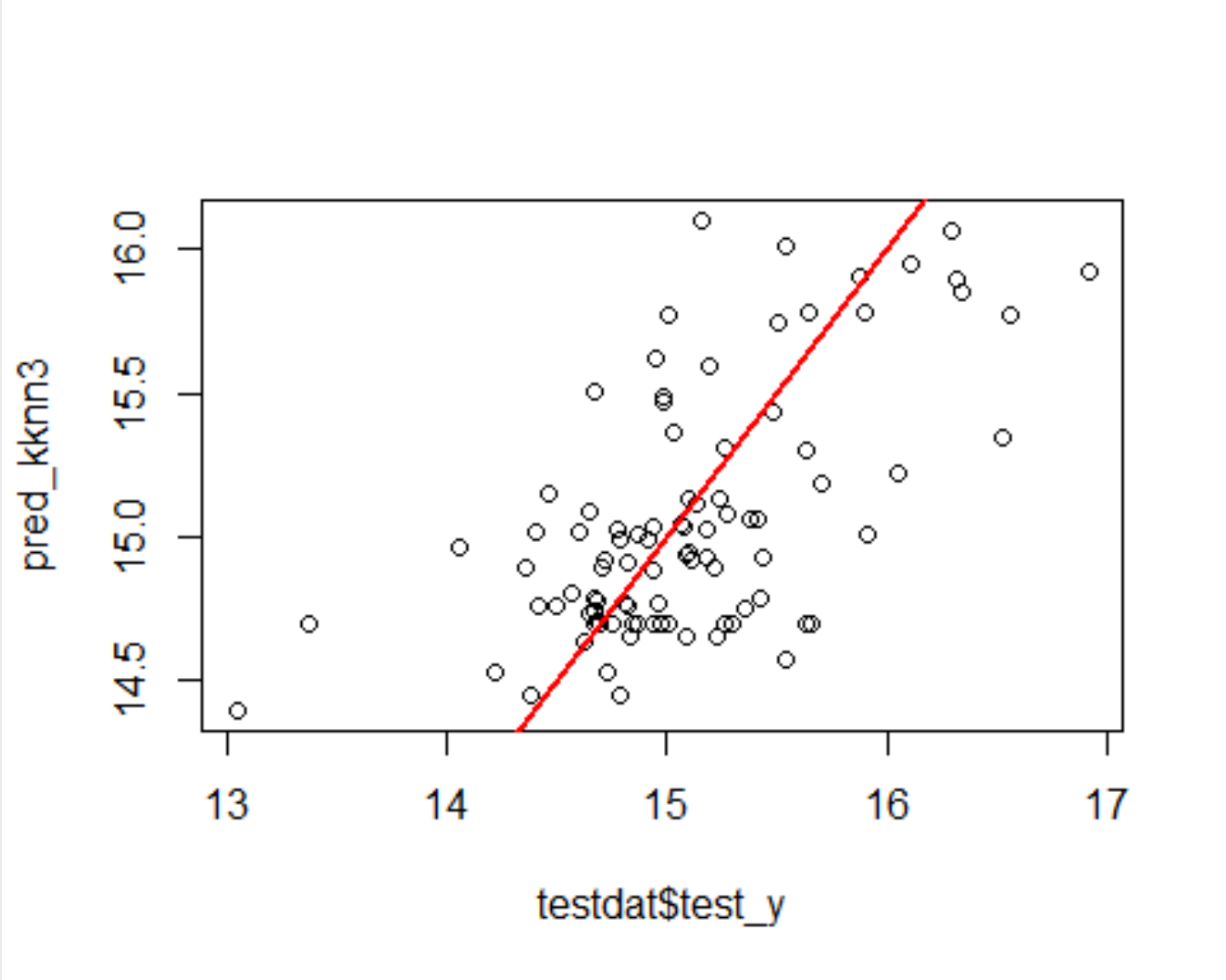
모델 설정 – 회귀나무 by tree

MSE = 0.301669



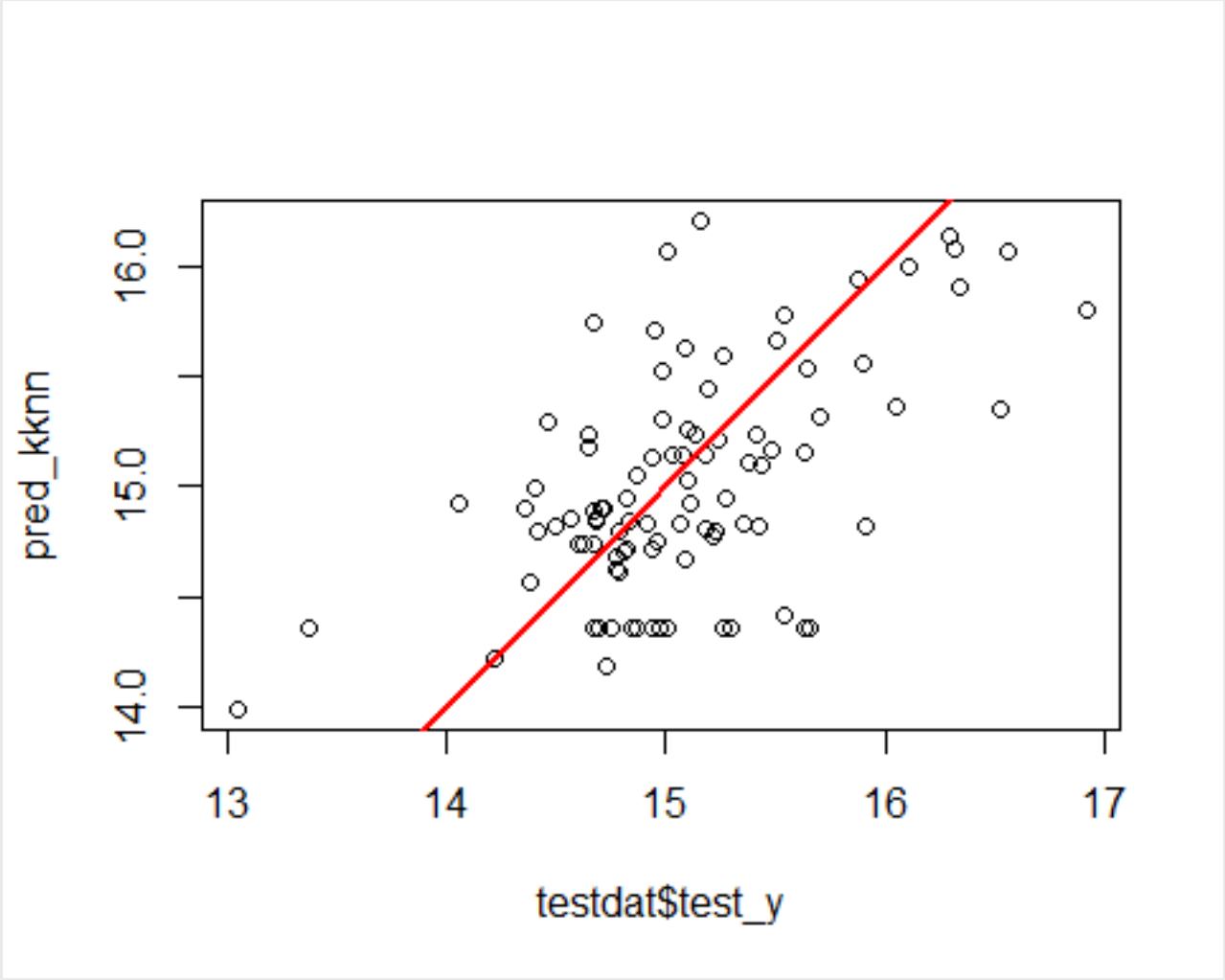
모델 설정 – KNN(unweighted) by kknn

K = 16, MSE = 0.234068



모델 설정 – KNN(weighted) by kknn

K = 29, MSE = 0.225853



모델 강화 –Ridge by glmnet

MSE = 0.306884, lambda = 0.6361991

```
7 x 1 sparse Matrix of class "dgCMatrix"
                             1
(Intercept) 15.126566124
yuheung      0.001003639
sukbak       0.001757917
school      -0.007855151
specialsch  -0.274356188
hospital    -0.105817143
store        0.053358195
```

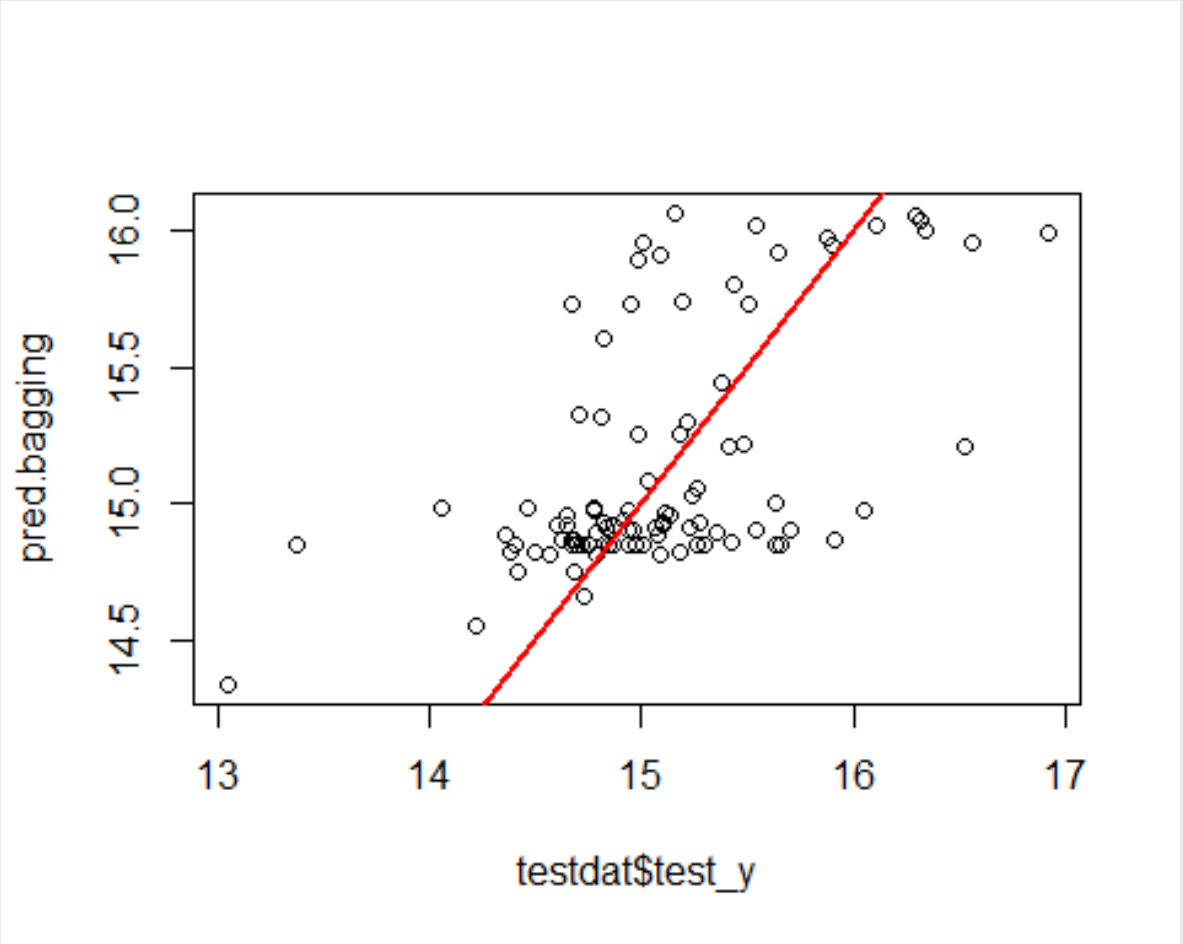
모델 강화 –Lasso by glmnet

MSE = 0.302142, lambda = 0.01248885

```
7 x 1 sparse Matrix of class "dgCMatrix"
                             1
(Intercept) 15.108904799
yuheung      0.001302628
sukbak       0.002503382
school      -0.008209807
specialsch  -0.418792572
hospital    -0.124585513
store        0.081589662
```

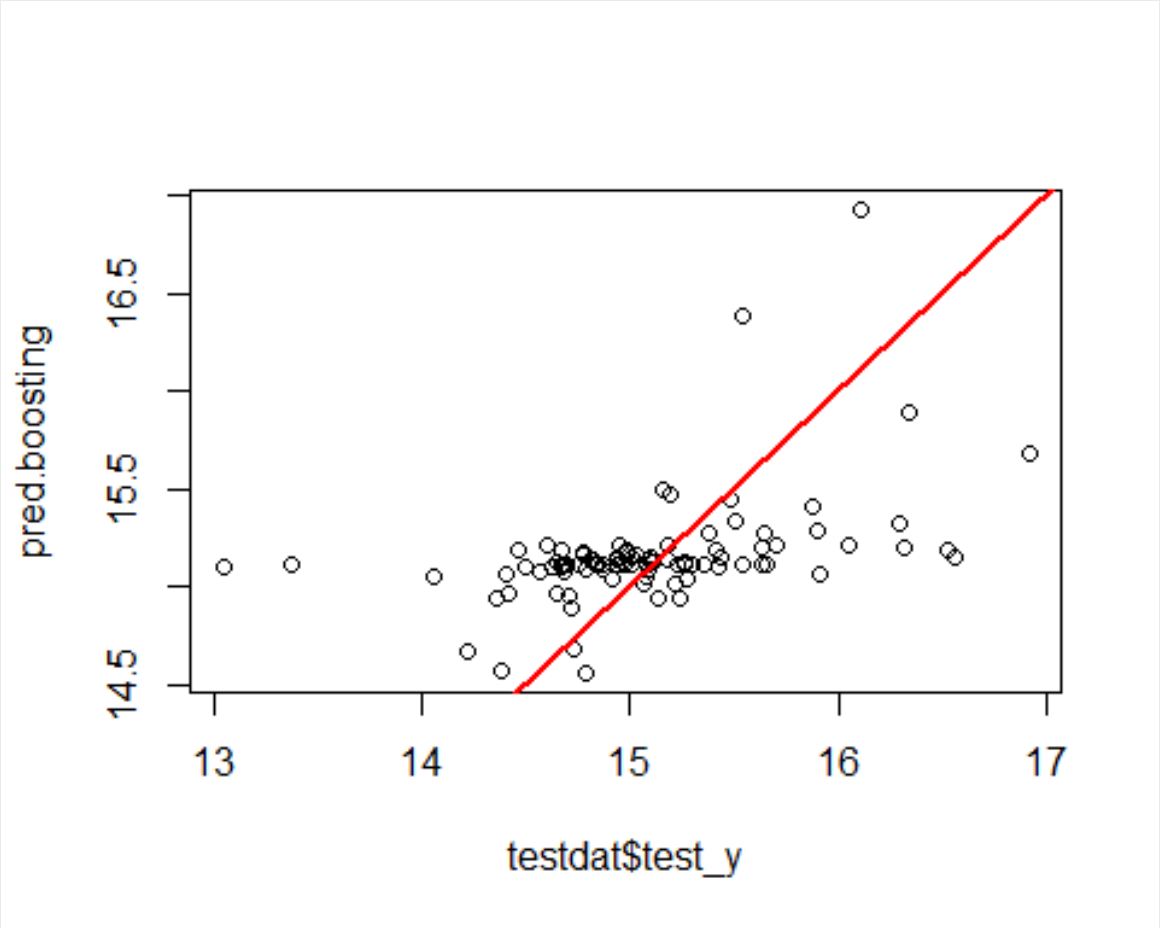
모델 강화 – bagging by rpart

MSE = 0.256919



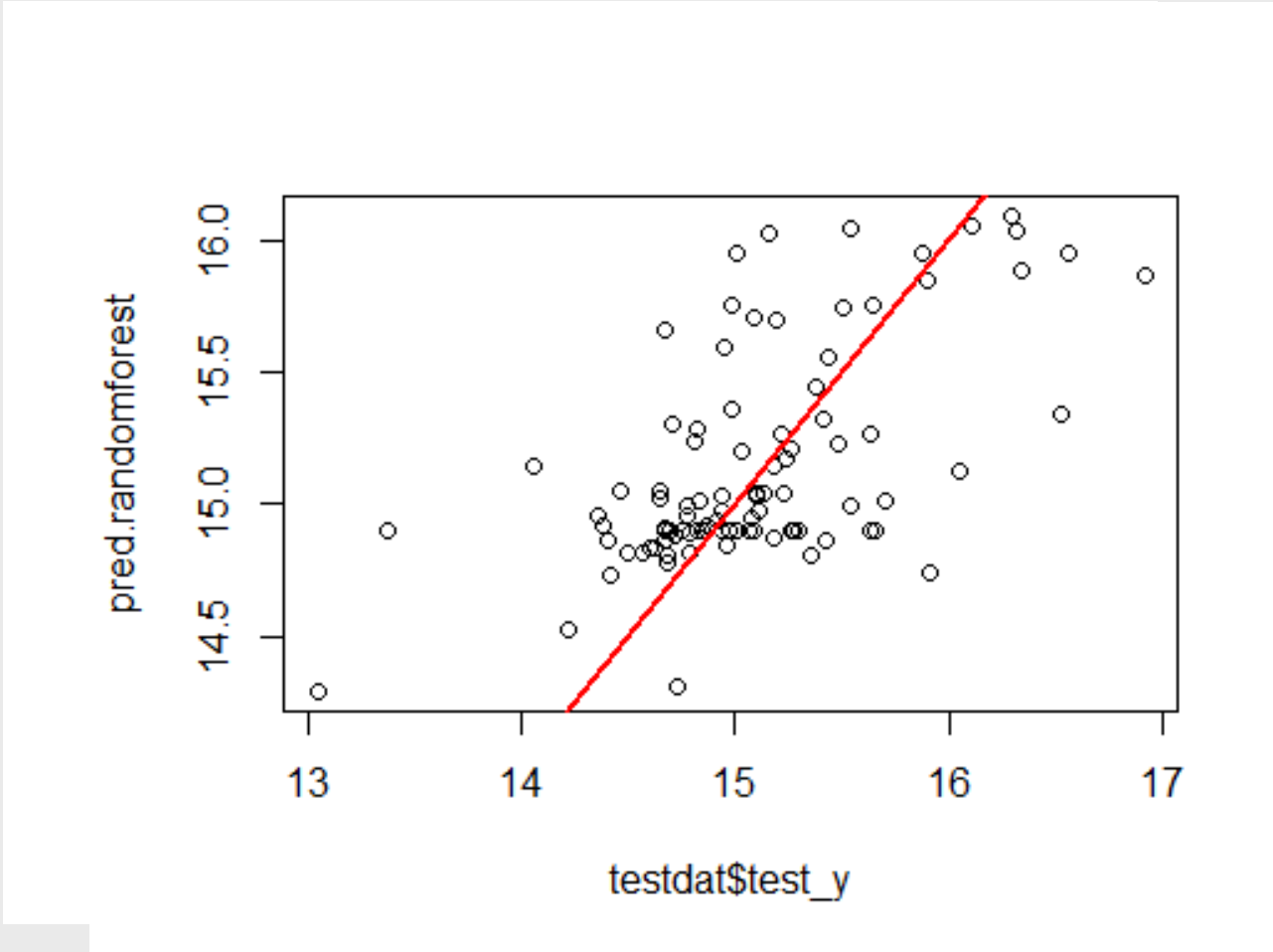
모델 강화 – boosting by mboost

MSE = 0.301867



모델 강화 – random forest by rpart

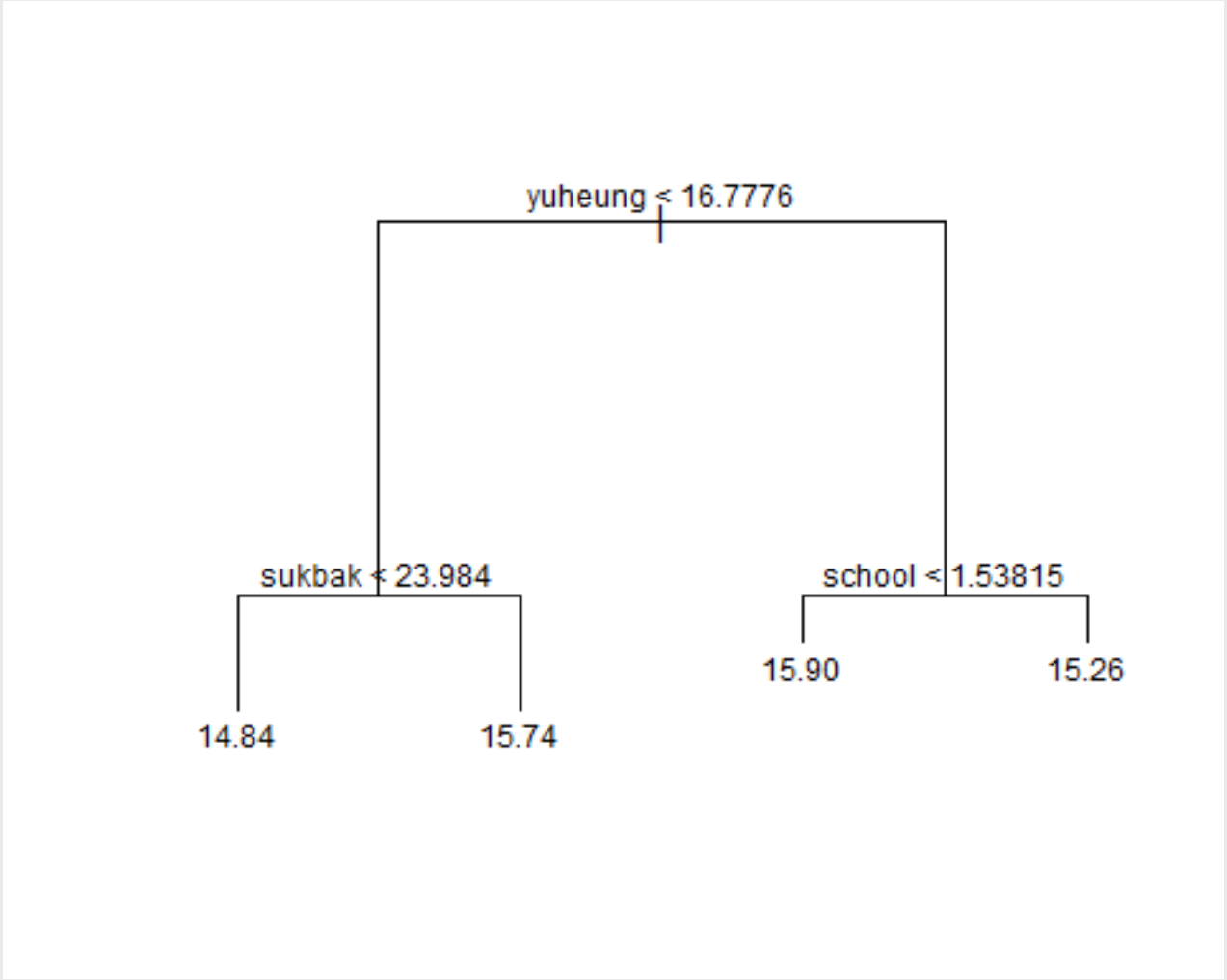
MSE = 0.238214



한계점

- 분포 변화에 대한 반응
- 적은 데이터 수

결론


yuheung < 16.7776
sukbak < 23.984
school < 1.53815
14.84
15.74
15.90
15.26