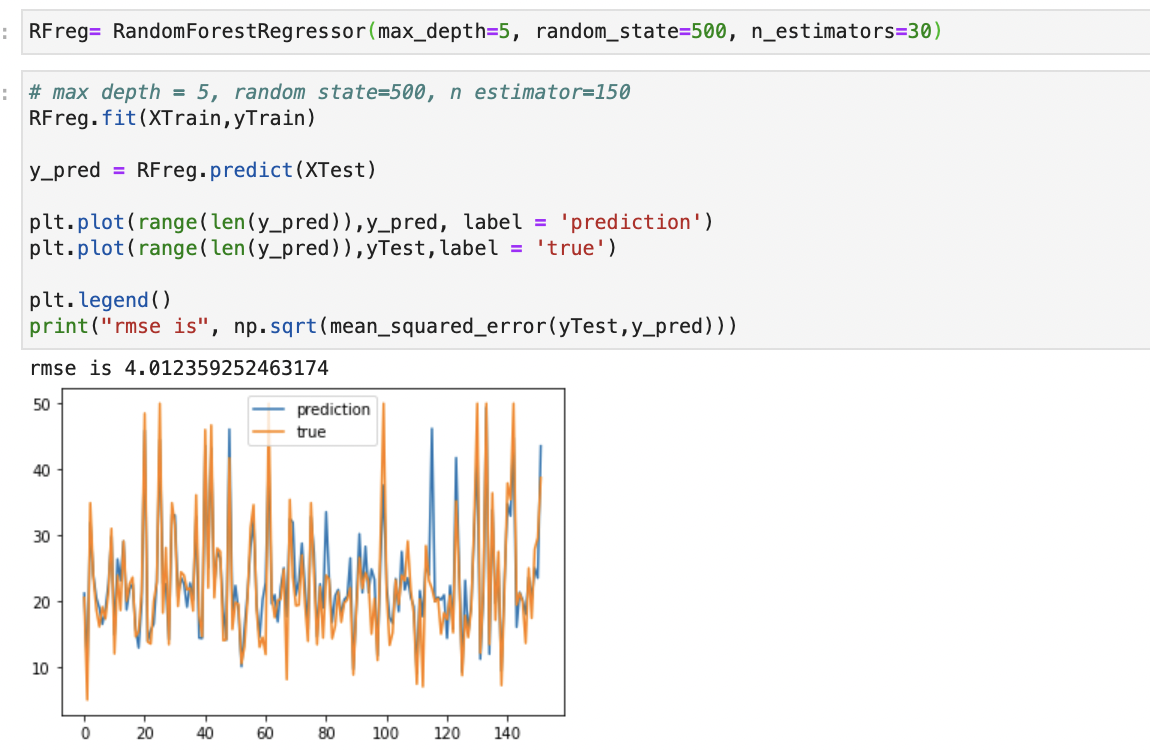
**Random Forest.**

20170396

김문정

1. Cross validation 을 하기 전에 임의로 hyperparameter를 설정하여 boston data set 을 randomforest regression 시켜보았습니다.



Max\_dept = 5, n\_estimators=30, max\_feature = default( same with number of feature) 로 하였을 때 rmse 는 대략 4.01 정도가 나왔습니다.

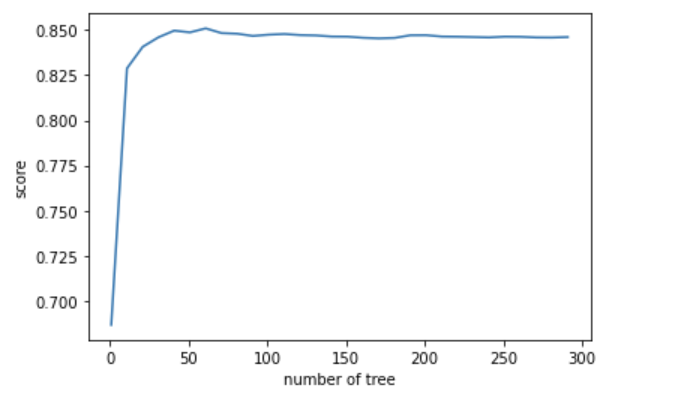
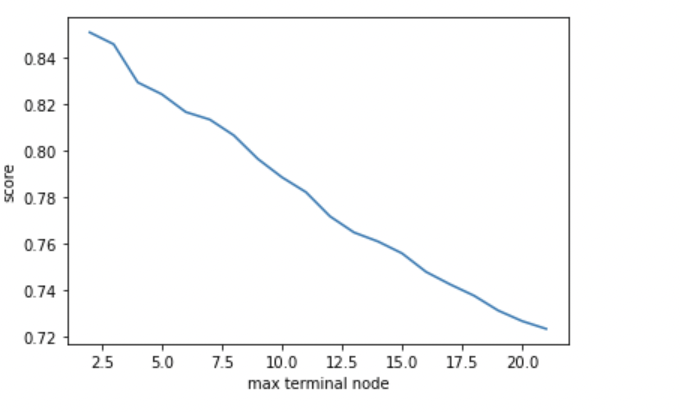
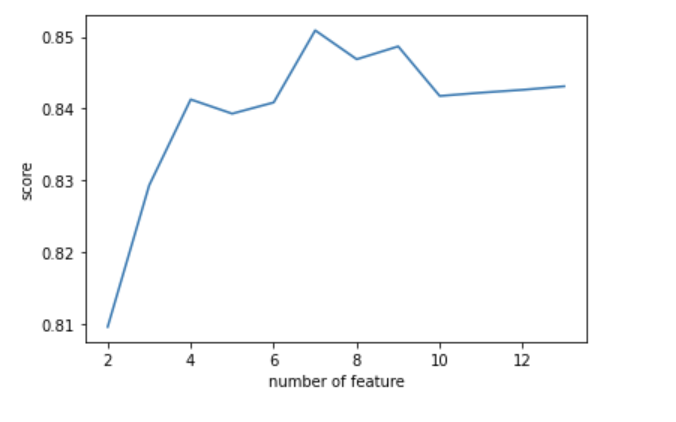
1. Cross validation



Hyperparameter 를 number of tree 를 1부터 301 까지 10간격으로 총 30개의 grid, maximum feauture 를 2부터 14까지 총 12개의 grid, maximum terminal node 수를 2부터 22까지 총 1간격으로 총 20개의 Grid로 나누어 cross validation 을 진행하였습니다.

2-1) sklearn 에 있는cross\_val\_score 을 이용하여 최고의 parameter 를 찾았습니다.

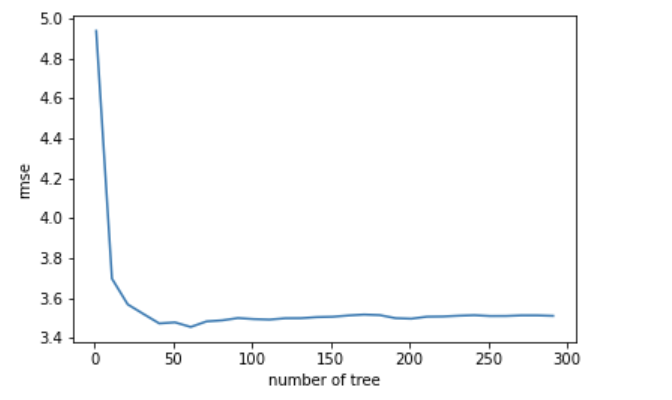
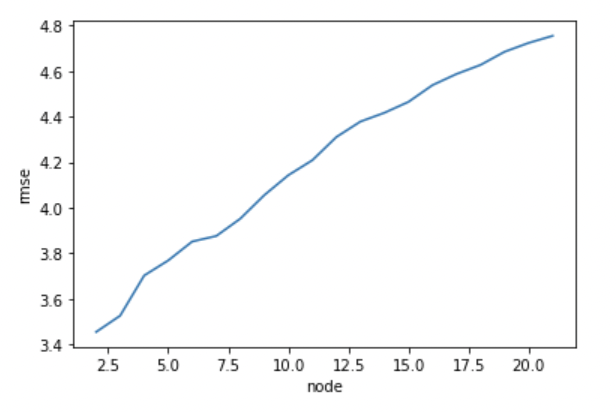
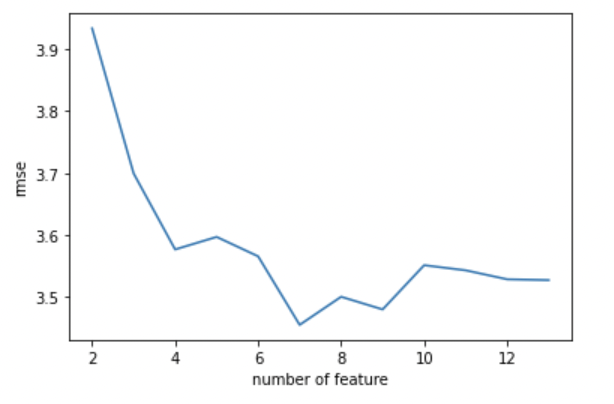


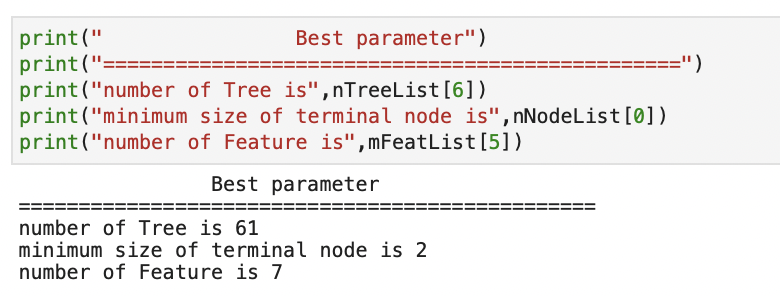
  

Number of tree 가 61일때, minimum size of terminal node 가 2일때 number of feature 가 7일 때 가장 score 이 높은 것을 확인해볼 수 있었습니다.

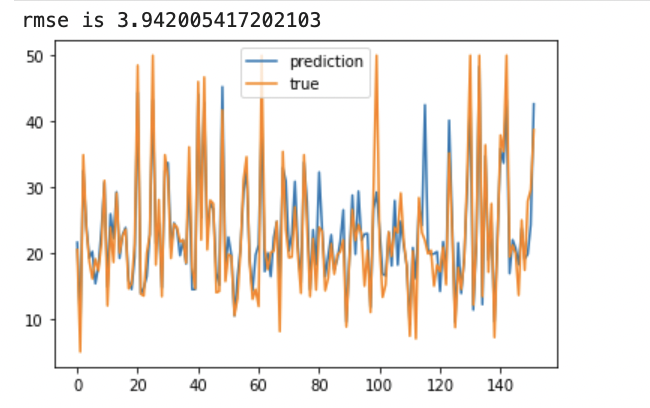
여기서 score 이 정확히 어떤 기준인지 몰라, 다시 rmse 를 기준으로 cross validation 을 실시해보았습니다.

2-2) rmse 



Rmse 를 기준으로 했을 때도 sklearn 의 cross validation score 을 기준으로 했을 때와 똑같은 parameter 가 최적의 값으로 나왔습니다.



이경우 rmse 가 약 3.94 정도로 나왔습니다.