

Assignment #2

머신러닝 이론과 실전

Due: September 13, 2021

1. Assume that a training dataset and a test dataset are available. Modify your program in Assignment #1 to do followings.
 - a. Prompt the user to enter the training/test data file name.
 - b. Use a data file named 'boston_tr.csv' for the training and 'boston_tst.csv' as the test data.
 - c. Fit a regression model on the dependent variable 'medv' with other variables as independent variables.
 - d. Calculate the predicted R2, MAE, MAPE, RMSE.
 - e. The output file for classification generated by the program must look like below. (The numbers are fictitious).

```
Coefficients
-----
Constant:  5.312
Beta1:    1.345
Beta2:    .236
Beta3:   -.439
Beta4:    .457

Model Summary
-----
R-square = .5689
MSE = .234

Prediction Performance
-----
Predictive R-square = .5689
MAE = 0.03
MAPE = 1.3
RMSE = .234
```

2. Use a Python package ('sklearn' or 'statmodels' packages can be used) to run a linear regression. Calculate the values of predicted R2, MAE, MAPE, RMSE.
3. Compare the values of #1 and #3.

과제로 제출할 보고서의 내용에는 다음과 같은 것이 포함되어야 합니다.

1. 파이썬 코드 내용 (코드의 내용을 과제의 보고서에 그대로 포함시키시기 바랍니다. 또한 파이썬코드는 파일로도 별도로 제출합니다. 제가 코드를 돌려보기 위해 코드파일도 필요합니다.)
2. output 파일의 내용 (파이썬코드를 실행시키면 생성되는 파일의 내용을 그대로 보고서에 포함시키시기 바랍니다. (제가 코드를 돌려보면 output 파일이 생성될것이기 때문에 여러분은 output 파일은 별도로 제출하지 않습니다.)