DS 회귀분석 수업중 팀과제

* 선택 1 또는 선택 2 중 택일, 과제 앞부분에 선택을 명기할 것

1. Reference
   1. <https://cran.r-project.org/web/packages/olsrr/vignettes/variable_selection.html>
   2. https://daviddalpiaz.github.io/appliedstats/variable-selection-and-model-building.html
2. Boston housing data set의 각 속성에 대하여 조사하라.   
   참조: https://archive.ics.uci.edu/ml/datasets/Housing.
3. 위 데이터 셋을 기반으로 주택 가격을 결정하는 선형회귀 모델을 생성하라.
   1. 예측, 설명의 목적으로 각각 모델을 생성한다.
4. 예측모델로써 성능을 측정하라.
   1. 테스트 집합에 대한 RMSE 사용 – CV 필요
5. 설명모델로써 성능을 측정하라.
   1. 테스트 집합이 필요 없음. – 이유는?
   2. 설명모델로써의 성능과 유의미한지를 test 하라.
      * 모델 F-test 의 p-value, R, Adjusted-R 등
   3. 주택 가격에 영향을 미치는 정도에 따라 변수들을 순서대로 나열해 보라.
      * 변수 T-test 의 p-value
   4. (선택 1) 변수 선택 Stepwise-selection을 실행
      * forward / backward 방법에서 먼저 선택 / 나중에 제거 되는 변수 일수록 더 중요한 변수이다. 위의 p-value 와 비교해 볼 것
   5. 각 변수의 회귀계수는 어떠한 의미를 가지는가?
      * 설명력과는 무관함. 단지, 음, 양은 목표변수와 음, 양 비례를 나타냄.
6. (선택 2) 다른 회귀 모델과 비교
   1. 결정나무, random forest, 신경망 등을 사용하여 회귀 모델 생성
   2. 정확도, 변수 중요도 등의 측면에서 비교

- 끝 -