Multiple Regression

# 라이브러리 링크

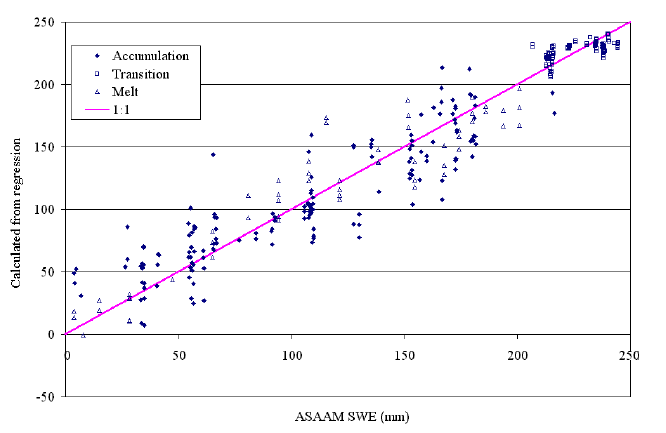
**install :**

<https://scikit-learn.org/stable/install.html>

**github :**

<https://github.com/scikit-learn/scikit-learn/blob/main/sklearn/linear_model/_base.py>

# 기초 설명



통계학에서의 선형 회기중, 둘 이상의 설명 변수(독립 변수, X value ) 에 기반한 선형 회귀 방법.

# 버전 정보

* NumPy >= 1.14.6 (pip install numpy)
* Scipy >= 1.1.0 (pip install scipy)
* Joblib >= 0.11 (pip install joblib
* Threadpoolctl >= 2.0.0 (pip install threadpoolctl)
* pandas >= 1.2.4 (pip install pandas)
* matplotlib == 3.22 (pip install matplotlib)

# 데이터셋 설명 및 출처

* imports-85.data : UCI Machine Learning Repository 에서 제공하는 Automobile Data set.
* 상세 데이터 정보는 imports-85.names 항목 내부에 별첨.
* Cite This Dataset : Automobile. (1987). UCI Machine Learning Repository.

출처 : <https://archive-beta.ics.uci.edu/ml/datasets/automobile>

# 코드 설명

* imports-85.data의 자동차 데이터셋을 활용하여 multiple-regression 형식을 통해 설명변수 [“engine-size”,”highway-mpg”,”city-mpg”] 에 대한 “price” 정보를 학습 및 이를 검증하여 이에대한 수치값 및 learning curve 형식으로 나타내는 코드.
* (별첨 1 : 데이터 내부의 Missing Attribute Values로 인해 NaN 데이터가 발생하여 이를 데이터 내부 “make” 변수 기준으로 그룹화 하여 이를 그룹 내부의 평균수치값으로 대체함. )

# 검증 방법

* multiple\_regression\_ex1.ipynb: 코드 내부에서 데이터셋을 학습용 데이터셋과 검증용 데이터셋으로 나누어 이를 검증함.
* (test\_size = 0.2, random\_state = 0)
* r2\_socre 함수를 이용하여 학습된 모델에서 예측한 값과 실제 데이터값의 유사도를 측정.