Polynomial Regression

# 라이브러리 링크

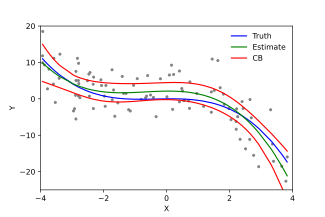
**install :**

<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.

설명변수(x value)들과 반응변수(y value) 사이의 관계가 비선형 적인 설정으로 선형 회가를 확장하는 표준적인 방법으로.

Y = b + ax + e 와 같은 표준적 선형 모델을

Y = B + ax + aX^2 …… ax^n + e 와같은 다항식 함수로 대체하는 것이다.

# 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

* 내부에서 Sin 그래프를 그리고 이에 대해 약간의 랜덤한 수치의 잡음 데이터를 섞어Sin 그래프와 유사한 경향성을 나타내는 1000개의 점들에 대한 샘플 데이터를 만들어 사용.

# 코드 설명

* Polynomial\_regression\_ex1.ipynb: Sin 그래프와 유사한 경향성을 보이는 1000개의 샘플 데이터셋을 만든후 , 이를 학습용 train 셋과 검증용 val 셋으로 나눈뒤 이를 각각 1차항 (degree 1), 3차항 (degree 3) 모델에 각각 학습시킨후 이에 대해 각 모델의 경향성 그래프를 나타냄과 동시에 각 모델의 learning curves를 나타낸다.
* 이후 이 학습 모델들을 사용하여 val셋을 통해 추측한 예측값과 실제 val셋의 값사이의 적합도를 r2\_socre 함수를 통해 표현한다.

# 검증 방법

* 코드 내부에서 데이터셋을 학습용 데이터셋과 검증용 데이터셋으로 나누어 이를 검증함.

(test\_size = 0.2, random\_state = 0)

* 추가적으로 r2\_socre 함수를 사용하여 실제값과 예측값 사이의 적합도를 평가.