

기계학습시스템설계 (ELEC 0504)

과제 03

5장. 정규방정식, 정규화, 표준화

과제 평가 방식

- 매 주 과제당 총점 100점 기준
 - 10점: 코드의 맨 처음 줄은 본인의 학번 및 이름을 주석으로 작성
주석은 # 을 줄 맨 앞에 삽입 ex) # 20221234 홍길동
 - 30점: 코드 가독성, 코드가 이해가능 하도록 작성되었는가?
 - 30점 : 작성한 코드의 실행 가능 여부
 - 30점 : 작성한 코드 실행 후 출력값 일치 여부

**** 주의 사항 ****

- 제출기한: 매주 화요일 자정(밤12시)까지 LMS에 업로드
- 지각 제출시 1시간당 -10점

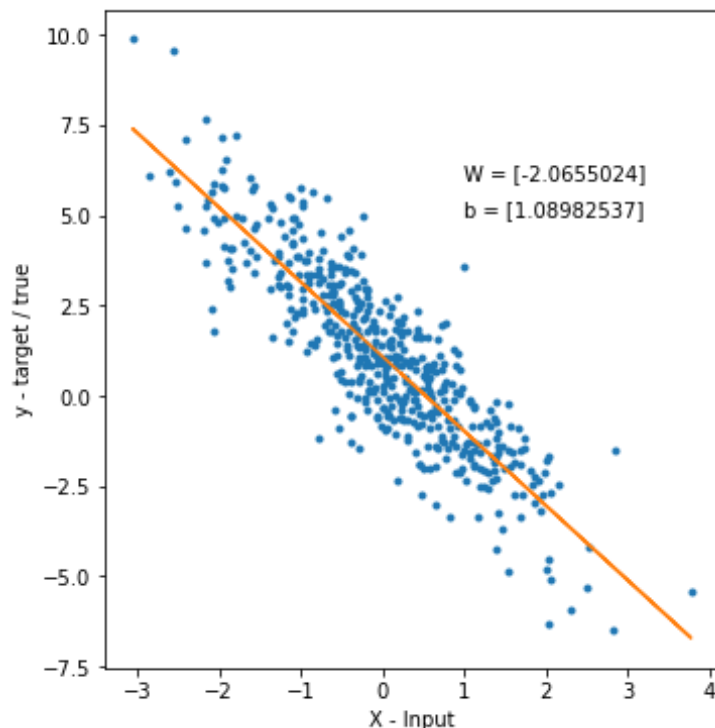
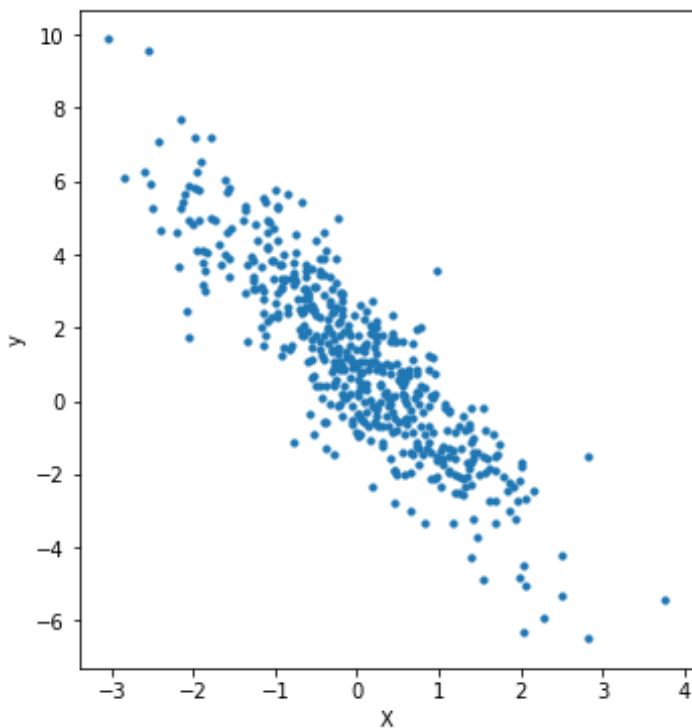
과제 제출 방법

- 평가 환경
 - Python 3.9.18
 - Spyder 5.4.3
- 파일명: 과제번호-문제번호 (ex. 03-1.py, 03-1.png, 03-2.py, 03-2.png)
- 제출방법
 - 코드파일(.py)들과 스크린샷파일(.png)들을 압축하여 1개의 zip파일로 제출

압축파일명: hw03.zip

과제 03-1

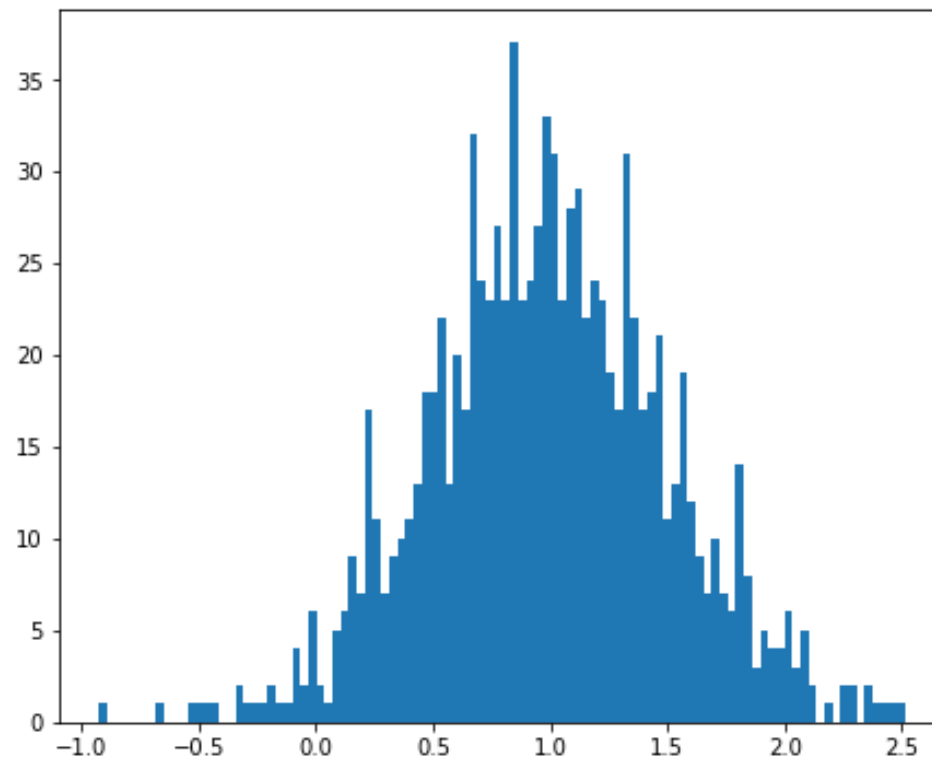
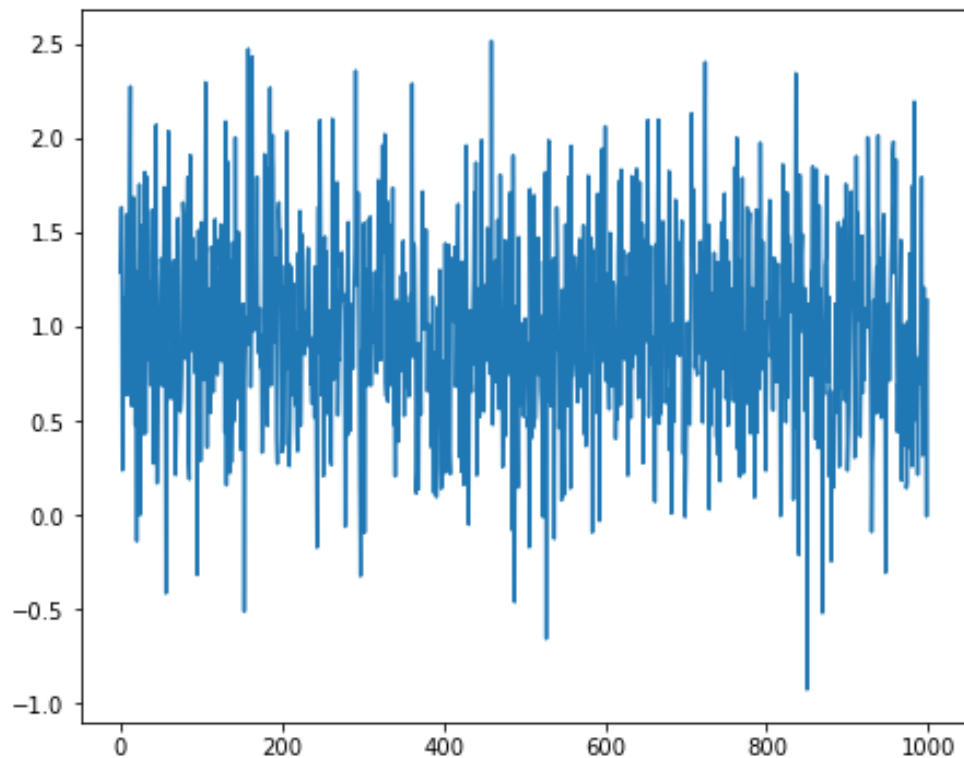
- 정규방정식(Normal Equation)을 이용하여 아래의 조건을 만족하는 선형회귀의 최적의 파라미터(w, b)를 구하고 다음과 같이 plot 하라. (오차는 MSE)
 - $X \sim N(0,1)$ 을 만족하는 500개의 sample 을 매번 다르게 생성
 - $y = -2X + 1 + 1.2 \cdot N(0,1)$



과제 03-2

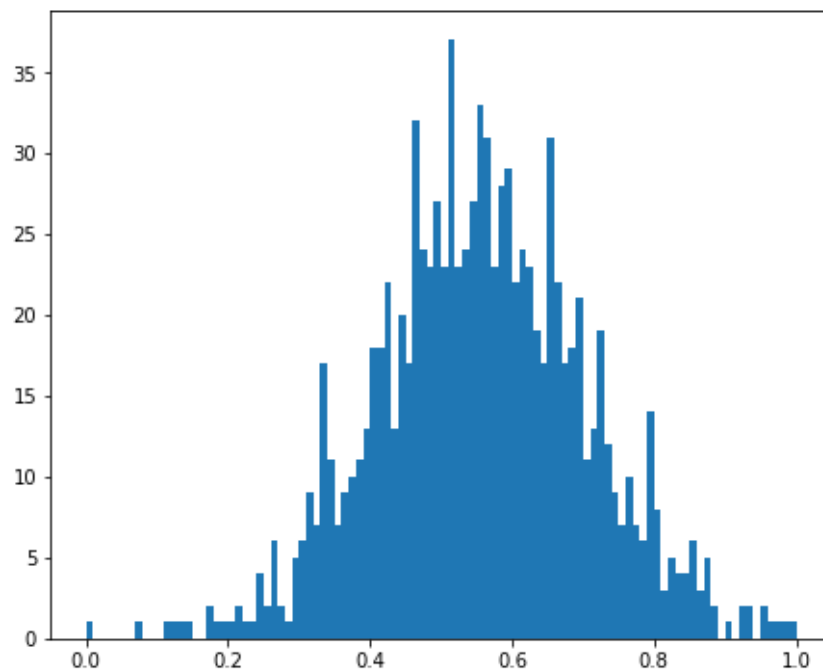
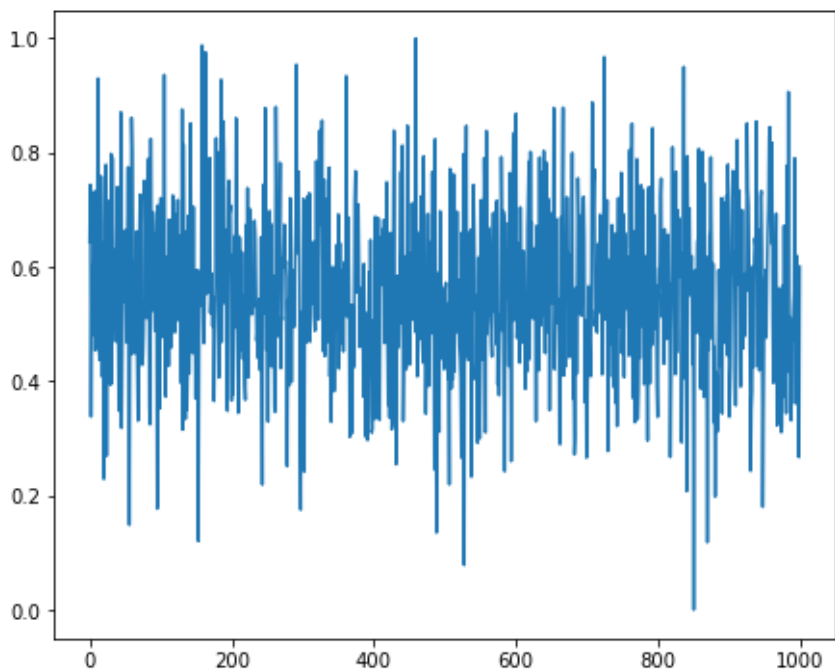
- 주어진 데이터(03-2_dataset.csv)를 불러와 아래의 작업을 수행하라.

1) (좌) 원본 데이터 plot, (우) 히스토그램 plot (100개의 구간 bins)



과제 03-2

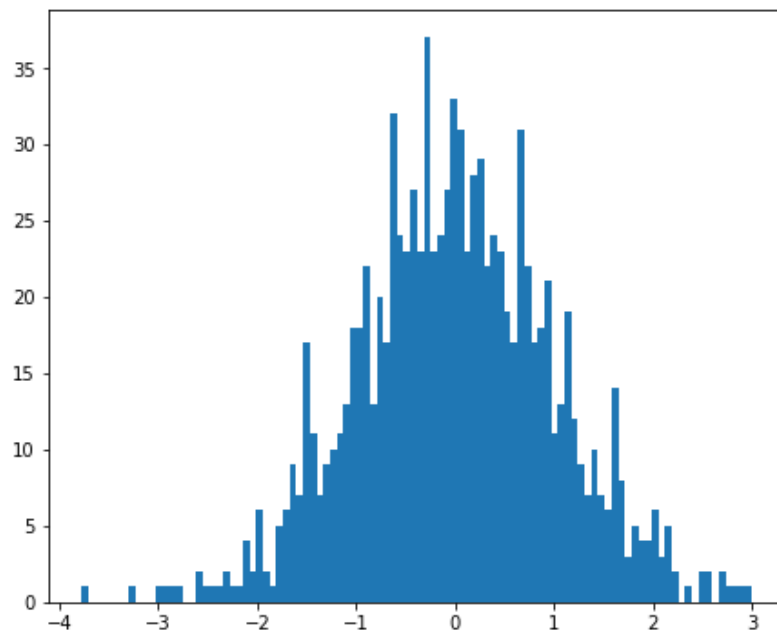
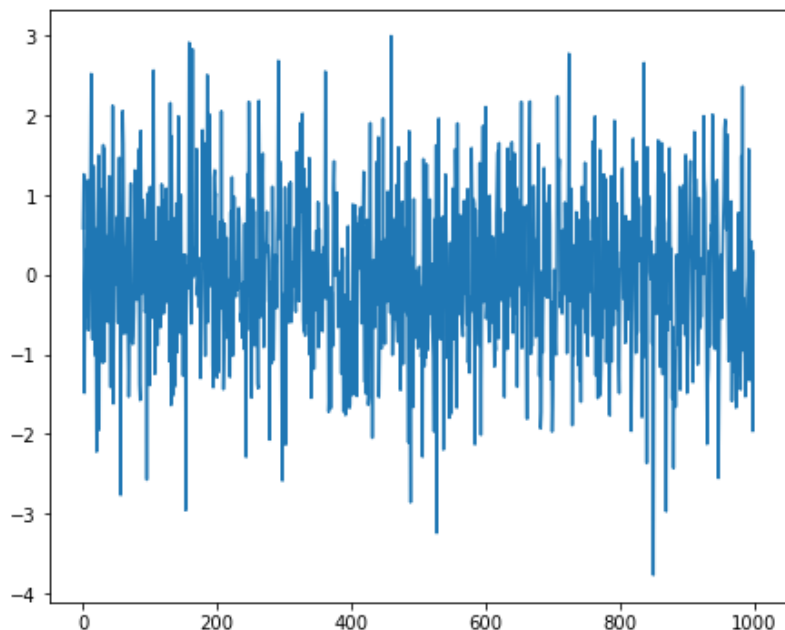
2) 정규화를 하기 위한 함수를 작성하고 결과를 plot 하라. 원본 데이터의 최대값, 최소값을 출력하고 정규화 된 데이터의 최대값, 최소값을 출력하라. 또한 이를 scikit-learn 의 정규화 함수 결과값과 비교하여라.



```
원본 데이터의 최대값 : 2.51344 최소값 : -0.92700  
정규화 된 데이터의 최대값 : 1.00000 최소값 : 0.00000  
scikit-learn 정규화 데이터의 최대값 : 1.0 최소값 : 0.0
```

과제 03-2

3) 표준화를 하기 위한 함수를 작성하고 히스토그램을 plot 하라. 원 데이터의 평균, 표준편차를 출력하고 표준화된 데이터의 평균, 표준편차를 출력하라. 또한 이를 scikit-learn 의 표준화 함수 결과값과 비교하라. (평균 및 표준편차는 소수점 둘째 자리까지만 출력)



```
원본 데이터의 평균값 : 0.99 표준편차 : 0.51
표준화 된 데이터의 평균 : -0.00 표준편차 : 1.00
scikit-learn 표준화 데이터의 평균 : -0.00 표준편차 : 1.00
```

과제 03-2

4) $y = -2X + 1 + 1.2 \cdot N(0,1)$ 를 만족하는 y 데이터를 생성하고, train 80% test 20%로 나누어서 선형회귀 모델을 구현하라. 학습의 결과로 얻은 선형 회귀 모델을 이용하여 train 데이터와 test 데이터를 입력으로 예측치를 계산하라. (붉은 색은 train 데이터, 파란 색은 test 데이터의 예측이 정답과 얼마나 일치하는지 plot, 최종적으로 test데이터의 예측값과 정답값의 MSE 계산)

* 아래 2가지 상황에 계산: 왼쪽은 정규화된 데이터, 오른쪽은 표준화된 데이터 (scikit-learn)의 함수 사용) *

