

# Homework 02

[SCS4049] Machine Learning and Data Science

Due: 09/25 (일), 23:59

아래의 질문들에 대한 jupyter notebook (.ipynb) 파일을 제출하세요. 반드시 제출하기 전에 초기화 후 전체 셀을 실행해서 문제가 없는지 확인하세요. 주석이나 마크다운은 자유롭게 활용해도 됩니다.

1. 주어진 데이터에 대해 linear regression 모델을 batch gradient descent 방법으로 학습하는 프로그램을 작성하세요.

- 주어진 data.csv 파일의 첫번째 열이 출력, 두번째부터 여섯번째 열이 입력입니다.
- 모델은 아래와 같이 bias를 포함합니다.

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 \quad (1)$$

- 초기 parameter는 다음과 같이 설정합니다.

$$\theta = [0.05, 0.20, -0.10, 0.20, 0.20, 0.30]^T \quad (2)$$

- Learning rate는  $\eta = 10^{-7}$ , 최대 epoch은 20회로 설정합니다.
- 아래 그림과 같이 (1) epoch별로 SSE cost를 나타낸 plot 그래프와 (2) true output  $y$  대비 prediction  $\hat{y}$ 를 그린 scatter 그래프를 최종적으로 출력합니다.

