**Assignment #2 - Regression**

“data.csv”파일은 2000개의 data이다. 각 data는 feature (x)와 label 값(y) 구성된 2차원 data이다. 터이다. Polynomial regression을 이용하여 예측 모델을 만들려고 한다. Data에서 40%는 training data로 하고 나머지 60%는 test data로 한다. Training data 10%는 verification data로 한다. 제공하는 regression.py를 다음의 task를 수행하기 위하여 완성하시오. regression.py에서는 polynomial 매개 변수를 구하는 library를 사용한다. (sklearn.linear\_model의 package에서 LinearRegression을 사용. PyhCharm을 사용할 경우 matplotlib와 sklearn의 package을 설치하시오.)

**#######################################################################**

**[Package 설치 방법]**

Pycharm에서, 왼쪽상단의 File – Settings – Project:projectname- Project Interpreter – 오른쪽의 +버튼 – sklearn검색 – 왼쪽 하단의 Install Package – 설치 완료 후 pycharm 재시동

* Matplotlib.pyplot 또한 같은 방법으로 설치.

####################################################**###################**

제공한 프로그램을 완성하시오. 프로그램은 다음을 수행하다.

1. Polynomial의 order를 1차에서 10차까지 증가시며 각 order에서의 MSE performance을 계산 해보고, MSE가 가장 낮은 최적의 polynomial order을 결정하라.
2. 최적의 Polynomial order이 구해졌으면, 이를 이용하여 Train set을 초록색 실선으로(g) Plot 해보고, 그 그래프 위에 Test data set을 파란색 점으로(b.) Overlay시켜보아라.
3. 최종적으로 구해진 Optimal model과 Test data set을 이용해 MSE Performance를 출력하라.