AI 프로그래밍

#HOMEWORK 12

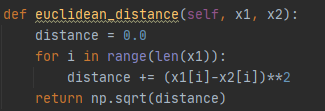


학번: 201724461

이름: 류지환

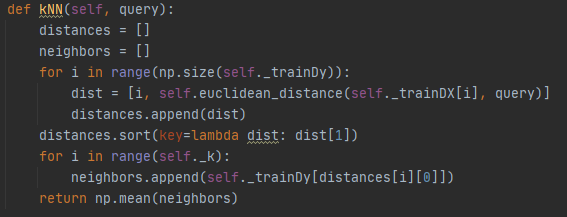
Implementing k-NN

1. Def euclidean\_distance(self, x1, x2)
   1. 함수 개요:



* 1. 주요 고려 사항:
     1. TrainDX(x1)와 query(x2) 사이의 거리를 구하는 함수를 구현.

1. Def kNN(self, query)
   1. 함수 개요:



* 1. 주요 고려 사항:
     1. TrainDX 내 모든 원소와 query 사이의 거리를 distance라 할 때

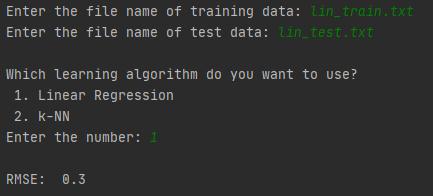
dist[trainDX의 index, distance]를 원소로 하는 distances list를 구현.

구현 시 index의 range는 train data들의 x값의 size가 3인 관계로 size가 1인 data들의 y값인 trainDy의 size로 지정.

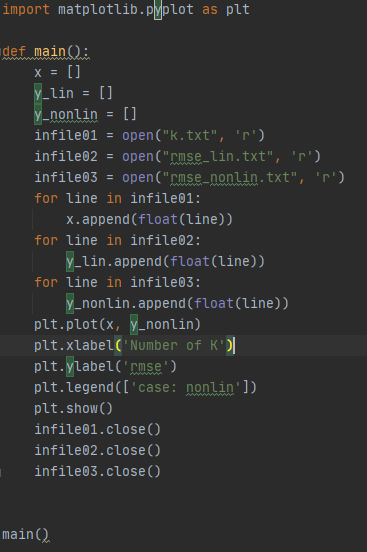
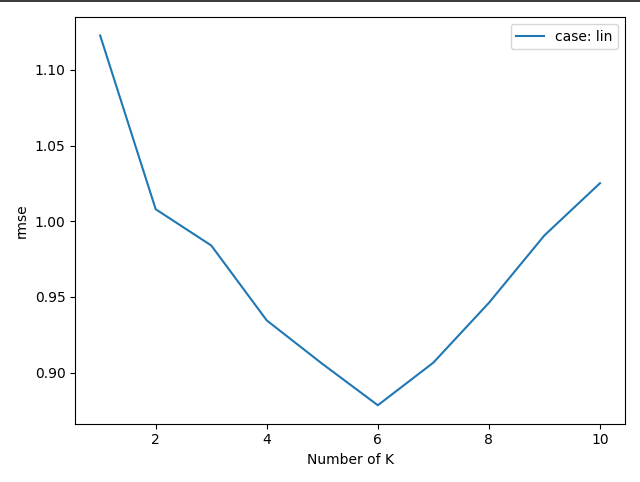
* + 1. 모든 trainDX의 index와 해당 원소와 query 사이의 distance를 저장한 distances list를 distance 값을 기준으로 오름차순으로 정렬.
    2. distances list 내 원소에 저장된 trainDX의 index를 활용해 해당 원소와 대응되는 trainDy 원소에 접근 후 neighbors list에 추가. 이를 distances[0]에서 distances[k-1]까지 반복.
    3. numpy.mean 함수를 활용해 neighbors 원소들의 평균 값을 구한 후 이를 return.

Experiments

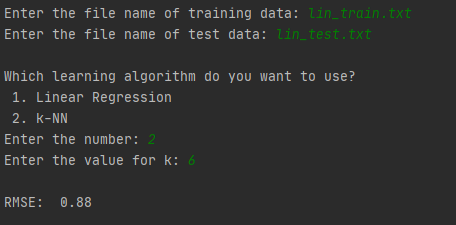
1. Linear test
   1. RMSE of linear regression with linear test data: 0.3



* 1. The best value of k for k-NN for linear data: 6
     1. 1~10 사이의 k-NN의 RMSE 데이터를 모아 k 값을 x축, RMSE 값을 y축으로 하는 그래프를 matplotlib.pyplot을 활용해 생성 후 최소값을 가지는 k와 RMSE 값을 구함.
     2. 참고자료:

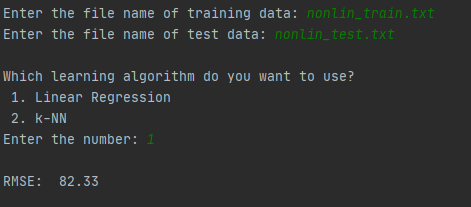
 plot/main.py,  (k, RMSE)그래프

* 1. RMSE of k-NN with linear test data: 0.88

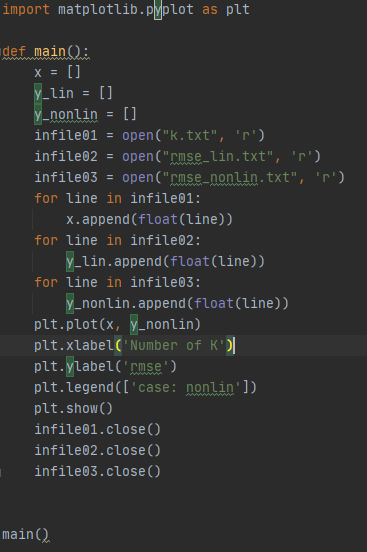
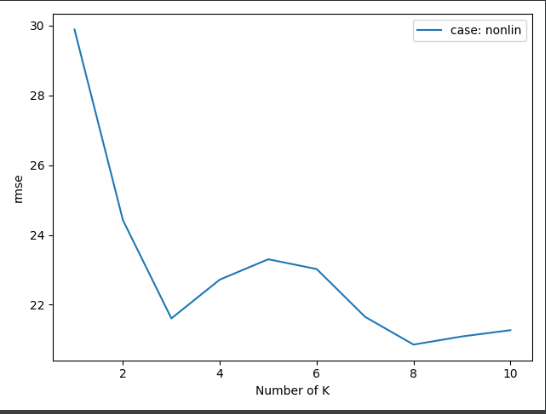


* 1. 평가:
     1. Linear data의 경우 linear regression의 RMSE가 k-NN의 RMSE보다 작으므로 linear regression 모델이 k-NN에 비해 상대적으로 정밀하게 측정됨.

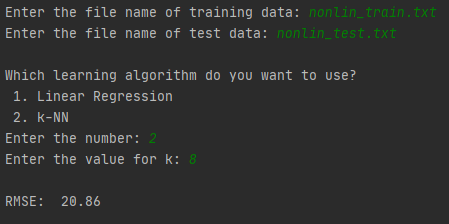
1. Non\_linear test
   1. RMSE of linear regression with non\_linear test data: 82.33



* 1. The best value of k for k-NN for non\_linear data: 8
     1. 1~10 사이의 k-NN의 RMSE 데이터를 모아 k 값을 x축, RMSE 값을 y축으로 하는 그래프를 matplotlib.pyplot을 활용해 생성 후 최소값을 가지는 k와 RMSE 값을 구함.
     2. 참고자료:

 plot/main.py,  (k, RMSE)그래프

* 1. RMSE of k-NN with linear test data: 20.86



* 1. 평가:
     1. Non\_linear data의 경우 k-NN의 RMSE가 linear regression의 RMSE보다 작으므로 k-NN 모델이 linear regression에 비해 상대적으로 정밀하게 측정됨.