

Learning Experiments

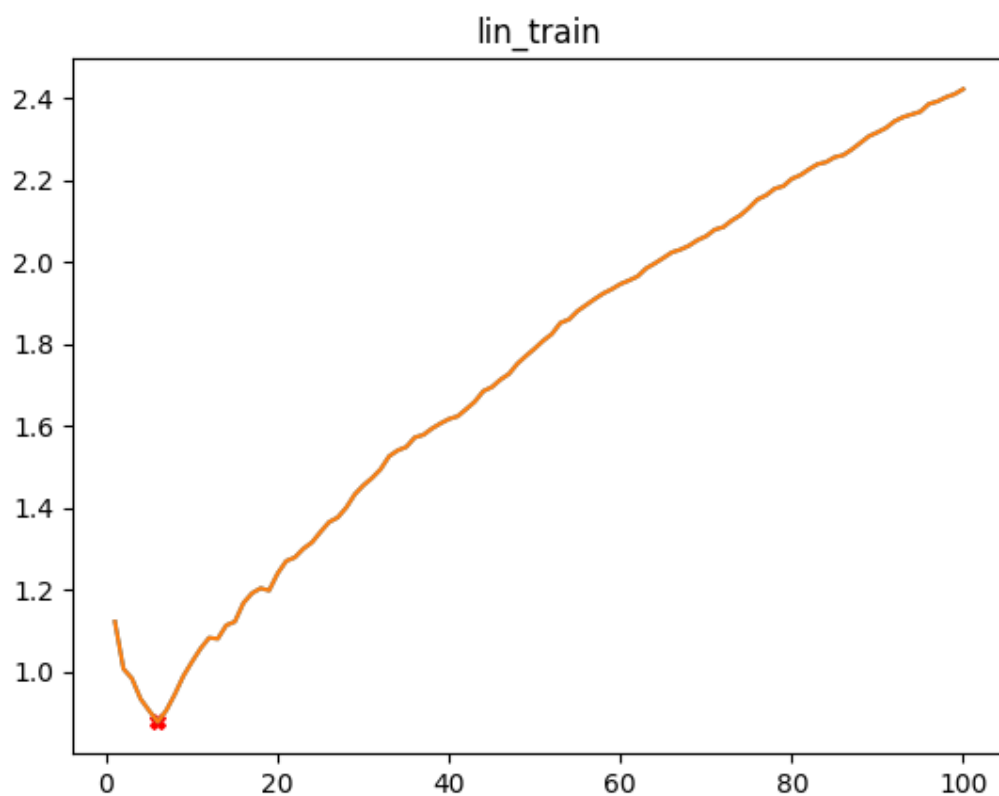
201645825 이승윤

선형과 비선형 데이터 세트를 이용하여 모델을 비교하였다. 학습 세트는 500 개, 테스트 세트는 100 개로 구성되어 있다. 학습 세트로 모델을 학습시키고 테스트 세트로 모델의 성능을 평가하였다. KNN 의 경우 데이터에 따라 가장 적합한 k 값을 찾아보았다.

1. 선형 데이터 세트

학습에는 lin_train.txt 테스트에는 lin_test.txt 를 이용한다.

모델	RMSE
Linear Regression	0.3
KNN (k = 6)	0.88

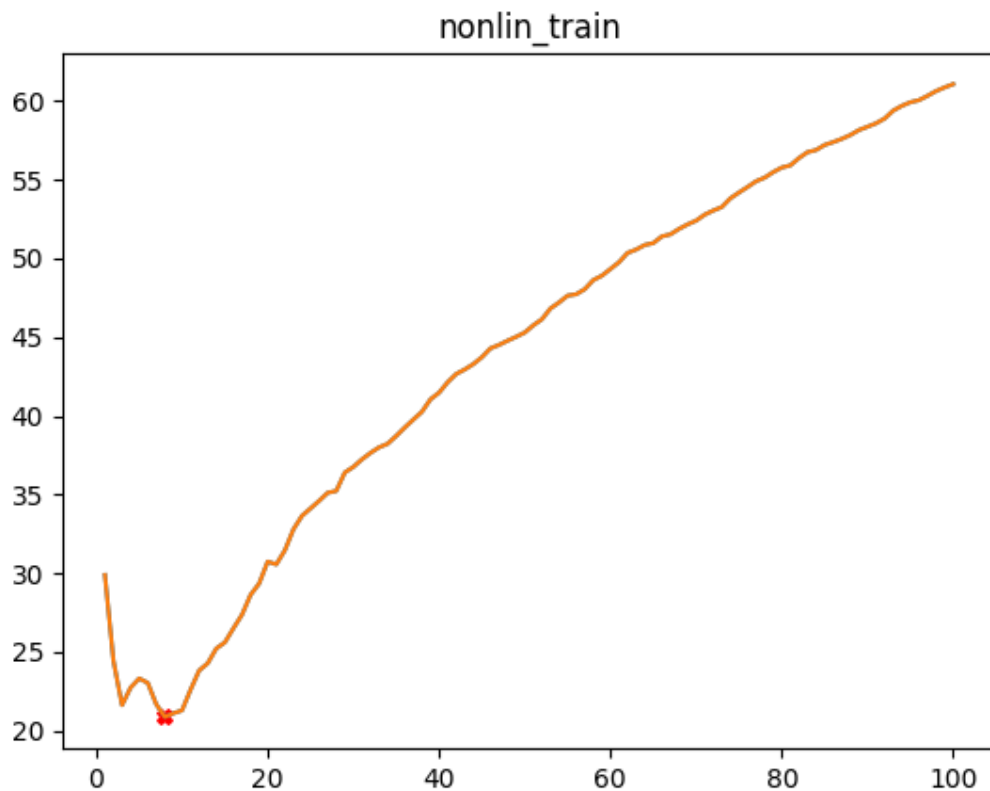


k 값을 찾기 위해서 k=1 부터 k=100 까지 테스트를 시행하였다. k=6 에서 RMSE 가 최소임을 확인할 수 있다.

2. 비선형 데이터 세트

학습에는 nonlin_train.txt 테스트에는 nonlin_test.txt 를 이용한다.

모델	RMSE
Linear Regression	82.33
KNN (k = 8)	20.86



k 값을 찾기 위해서 k=1 부터 k=100 까지 테스트를 시행하였다. k=8 에서 RMSE 가 최소를 가짐을 확인할 수 있다.

선형 데이터에 대해서는 선형 회귀가 KNN 보다 적은 오차를 가져 더 나은 성능을 보여주었고 비선형 데이터에 대해서는 KNN 이 선형 회귀보다 더 적은 오차를 가져 더 나은 성능을 보여주었다. 따라서 데이터가 선형인지 비선형인지 대략적으로 예측 가능한 경우 선형이면 선형 회귀를 비선형이면 KNN 을 사용하는 것이 유리하다. KNN 의 경우 k 값이 증가할수록 성능 개선을 보이다가 특정 값 이후에는 성능 하락을 보였다. 적합한 k 값을 찾는 것이 성능에 중요하다는 점을 알 수 있다.