

자동차 가격 예측

KNN의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

kaggle open data set, 영국의 중고차량 정보로 구성
정보를 통해 해당 차량의 가격을 예측

company	자동차 제조 회사, 총 5 개	fueltype	해당 차량의 연료유형
model	해당 차량의 제품명 (모델)	tax	해당 차량의 세금
year	해당 차량의 제조년도	mpg	해당 차량의 연비
transmission	해당 차량의 변속기	engine size	해당 차량의 엔진 크기
mileage	해당 차량의 연료 마일리지	price	해당 차량의 가격

knn hyperparameter 값에 따른 성능분석

이 레포트에서는 n_neighbors와 p, weights의 변경에 따른 성능결과를 분석합니다.

no	n_neighbors	p	weights	score
1	3	1	uniform (default)	1327.97249
2	5	1	Uniform	1319.74044
3	7	1	Uniform	1330.90698
4	3	2	uniform	1389.90698
5	5	2	uniform	1385.53384
6	7	2	uniform	1403.73599
7	3	1	distance	1307.67707
8	5	1	distance	1283.64394
9	7	1	distance	1276.77937
10	3	2	distance	1364.08459
11	5	2	distance	1344.85731
12	7	2	distance	1340.79720

이 데이터 셋에서는

n_neighbors 값이 커지거나 작아지는 거에 따른 점수의 경향이 보이지 않는다.

p 대체로 p=1 일 때 p=2 보다 성능이 좋은 것으로 보인다.

weights distance로 설정 시 더 좋은 성능을 보인다.

실제로 competition에서 제출한 코드는 no.9이다.

uniform은 거리와 관계없이 단순 평균, distance는 거리를 고려한 가중 평균을 계산한다.

실제로 위의 결과에 따르면 가중 평균을 통한 모델 학습이 더 성능이 좋았다.

이 데이터 셋에서는 p와 weights가 유의미한 변화를 보이는 hyper parameter임을 알 수 있었다.