

재배환경 별 작물 종류 예측

KNN 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

환경에 따른 추천 작물을 위한 데이터 셋

평가 데이터 550 개에 대해 총 22 개의 작물 중 하나로 분류

N	토양 내 질소 함량 비율
P	토양 내 인산 함량 비율
K	토양 내 칼륨 함량 배율
temperature	섭씨 온도
humidity	상대 습도 (%)
ph	토양의 ph 값
rainfall	강우량 (mm)

knn hyperparameter 값에 따른 성능분석

이 레포트에서는 n_neighbors 와 p, weights 의 변경에 따른 성능결과를 분석합니다.

no	n_neighbors	p	weights	score
1	3	1	uniform (default)	0.97272
2	5	1	uniform	0.97090
3	7	1	uniform	0.97272
4	3	2	uniform	0.96727
5	5	2	uniform	0.97090
6	7	2	uniform	0.96727
7	3	1	distance	0.97272
8	5	1	distance	0.97090
9	7	1	distance	0.97454

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 uniform 이든 distance 이든 큰 성능차이를 보이지 않는다.

실제로 competition 에서 제출한 코드는 no.9 이다.

처음에는 no.1 코드 값을 사용했으나, weights 값을 변경해본 결과 더 높은 점수가 나왔다.

weights 값이 대체로 큰 성능 차이를 보이지 않았으나 n_neighbors=7, p=1 인 경우에는 weights 값이 distance 일 때 더 좋은 성능을 보였다.

hyper parameter 값의 변화를 통해 성능이 변한다는 것을 알 수 있었다.