

당뇨병 환자 예측

Quadratic Discriminant Analysis 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

8 가지 조건을 바탕으로 당뇨병 보유 여부를 알아보는 데이터 셋입니다.

Pregnancies	임신 횟수
Glucose	글루코오스 (탄수화물 화합물)
BloodPressure	혈압
SkinThickness	피부 두께
Insulin	인슐린 수치
BMI	BMI 지수
DiabetesPedigreeFunction	가족력
Age	나이
Diabetes	당뇨병보유 여부

Quadratic Discriminant Analysis - hyperparameter 값에 따른 성능분석

이 레포트에서는 reg_param 의 변경에 따른 성능결과를 분석합니다.

no	reg_param	score
1	-10.0	0.65367
2	-5.0	0.65367
3	-1.0	0.65367
4	-0.5	0.65367
5	-0.3	0.65367
6	-0.1	0.51082
7	0 (default)	0.77056
8	0.1	0.77056
9	0.3	0.75757
10	0.5	0.70562
11	1.0	0.63203
12	5.0	0.65367
13	10.0	0. 65367

이 데이터 셋에서는 유일하게 변경가능한

reg_param default 값인 0 에서 가장 좋은 성능을 보였다. 값이 양수인 경우 커질수록 성능은 대체로 나빠지는 경향을 보였고, 값이 음수인 경우 작아져도 성능의 차이를 보이지 않았다.

이 데이터에서 좋은 성능을 보인 모델은 파라미터 변화가 아닌 train 세트에서 임신 횟수 데이터를 제외시키고 학습시킨 모델이었다.