

수면시간에 따른 우울증 예측

Logistic Regression 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

다트머스 대학교 학생들을 대상으로 한 설문조사에 대한 응답 데이터를 전처리를 진행한 데이터이다.

1. 어제 밤에 몇시간을 잤는가? 2. 어젯밤 수면에 대해 몇 점을 줄 것인가?

sleep_time_mean	평균 수면시간	sleep_quality_mean	평균 수면의 질
sleep_time_std	수면시간의 표준편차	sleep_quality_std	수면 질의 표준편차
sleep_time_min	최소 수면시간	sleep_quality_min	수면 질의 최소값
sleep_time_25	하위 25%에 해당하는 수면시간	sleep_quality_25	하위 25%에 해당하는 수면의 질
sleep_time_50	하위 50%에 해당하는 수면시간	sleep_quality_50	하위 50%에 해당하는 수면의 질
sleep_time_75	하위 75%에 해당하는 수면시간	sleep_quality_75	하위 75%에 해당하는 수면의 질
sleep_time_max	최대 수면시간	sleep_quality_max	수면 질의 최대값

Logistic Regression - hyperparameter 값에 따른 성능분석

이 레포트에서는 C 와 solver, class_weight, max_ite 의 변경에 따른 성능결과를 분석합니다.

no	class_weight	solver	C	max_iter	score
1	none	lbfgs	1.0	100	0.60869
2	balanced	lbfgs	1.0	100	0.69565
3	balanced	lbfgs	200	500	0.60869
4	balanced	lbfgs	1.0	500	0.69565
5	balanced	sag	1.0	100	0.60869
6	balanced	liblinear	1.0	100	0.60869
7	balanced	liblinear	200	100	0.56521
8	balanced	liblinear	100	100	0.60869
9	balanced	liblinear	10	100	0.52173
10	balanced	liblinear	0.1	100	0.47826
11	none	newton-cg	1.0	100	0.60869
12	balanced	newton-cg	1.0	100	0.69565

이 데이터 셋에서는

C 값이 1 보다 작아지거나 100 보다 커지면 성능이 낮아지는 것을 알 수있다.

class_weight none 인 경우보다 balanced 로 설정되었을 때 더 좋은 성능을 보인다. 이는 주어진 trainY 의 값들이 0 에 치우쳐있음 때문임을 표를 통해 알 수 있었다.

solver sag 의 경우에 다른것들 보다 낮은 성능을 보였고 나머지 solver 들은 비슷한 성능을 보였다.

여러 파라미터들을 바꾸어보았지만 더 나은 성능을 보이는 경우가 무엇인지 확실하게 파악하는 것이 불가능했다.