

은하계 종류 예측

Logistic Regression 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

3 가지 종류의 은하 데이터 셋 (Spiral, Edge, Smooth) 이 제공됩니다.

학습 데이터는 총 6,000 개, 평가 데이터는 1,500 개입니다.

Logistic Regression - hyperparameter 값에 따른 성능분석

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

no	solver	penalty	C	score
1	sag	l2(d)	1.0	0.79600
2	sag	l2	200.0	0.79666
3	sag	l2	300.0	0.79666
4	sag	none	default	0.79666
5	saga	l1	1.0	0.79600
6	saga	l1	200.0	0.79600
7	saga	l1	300.0	0.79600
8	saga	l2	200.0	0.79600
9	saga	l2	300.0	0.79600
10	newton-cg	l2	1.0	0.79666
11	newton-cg	l2	200.0	0.79666
12	newton-cg	l2	300.0	0.79666

이 데이터 셋에서는

solver

"sag"의 경우 penalty 값이 l2 와 none 만 사용 가능하다. 성능면에서는 penalty 값과 관계없이 C 값이 큰 경우 성능이 좋은 것으로 보인다.

"saga"의 경우 "sag"보다 대체적으로 성능이 떨어지는 것으로 보인다.

"newton-cg"의 경우 C 값이 변화되어도 동일한 성능을 보였다.

default 값인 "lbfgs" 또한 max_iter 값을 200, 500, 1000 으로 주고 계산해 보았으나 최저점인 0.79600 이 도출되었다.

penalty 크게 영향을 주지 않는 파라미터로 관찰된다.

C 크게 영향을 주지 않는 파라미터로 관찰된다.

분석에 따르면 solver 가 성능에 영향을 주는 것으로 보이지만 큰 차이가 없는 것으로 보아 추가적인 전처리가 필요할 것으로 생각된다.