

기계학습 6주차 과제

당뇨병 예상문제

컴퓨터공학과
17011584 정재경

개요

주어진 데이터셋은 사람들의 몸상태 정보에 대한 정보가 나타나있어, 각 정보에 따른 당뇨병 발병 여부를 나타낸다. 제공된 정보에 따라서 당뇨병 발병 여부를 예측하는 문제로, 분류문제에 해당한다. 이번 과제는 Quadric Discriminant Analysis 방식을 사용해서 해결해보았다.

실험

이번 과제에서 변경을 해보며 결과값의 차이 추이를 본 하이퍼파라미터는 solver, shrinkage 이다.

Reg_Param	Score
-10	0.65367
-5	0.65367
-1	0.65367
-0.5	0.65367
0	0.76623
0.5	0.70562
1	0.63203
5	0.65367
10	0.65367

결론

이 데이터셋은 reg_param=0 (default) 일 때의 성능이 가장 좋았다. QDA의 경우에는 하이퍼파라미터를 변경할 여지가 많이 없어서 데이터 전처리를 여러 방식으로 해보았지만, default일 때가 가장 점수가 좋았다. 이 문제는 QDA가 LDA보다 전반적으로 점수가 높았다.