Homework #4 Due: Apr. 30, 2020

2021310001 이성우

Implement the regularized logistic regression and apply it to the wine example. Use different values for the regularization parameter and see how the values of the coefficient changes with respect to . To deal with multiple classes, use the one-vs-all method.

코드:

텍스트이(가) 표시된 사진

자동 생성된 설명

wine데이터를 불러와 70% 대 30%로 train과 test data set을 나누었고, 본 연구에서는 validation data set은 제외하였습니다.

데이터의 첫열은 목표값인 y이고 2열부터 14열의 13개의 열은 feature data로 x로 설정하였습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

모든 데이터를 standardization을 진행해 주었습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

Training에 사용되는 Parameters는 다음과 같습니다.

parameter 의 경우 0에서 10간격으로 300까지 변경하면서 결과값이 어떻게 변하는지를 관찰하였습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

관찰할 람다의 수 만큼 전체 training을 반복하였습니다.

한번의 training에서는 아래와 같은 식의 update rule을 적용하였습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

3 class의 w값을 한번에 관찰하기 위해 평균을 내어주었습니다.

True data와 비교하여 다르게 예측하였으면 misclassification value에 1을 추가하여 ratio를 구했습니다.

결과분석

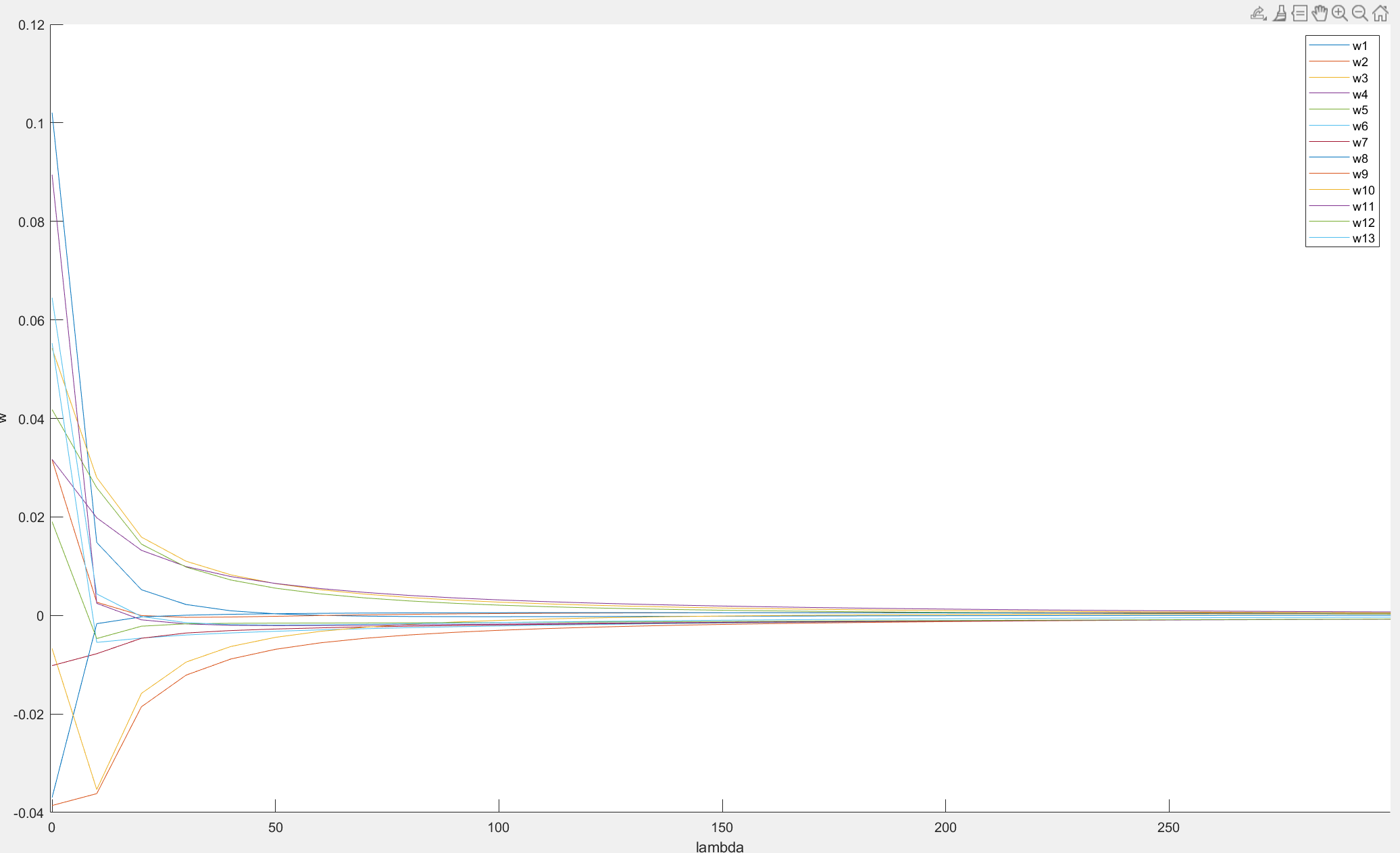
본연구의 train과 test accuracy는 다음과 같습니다.

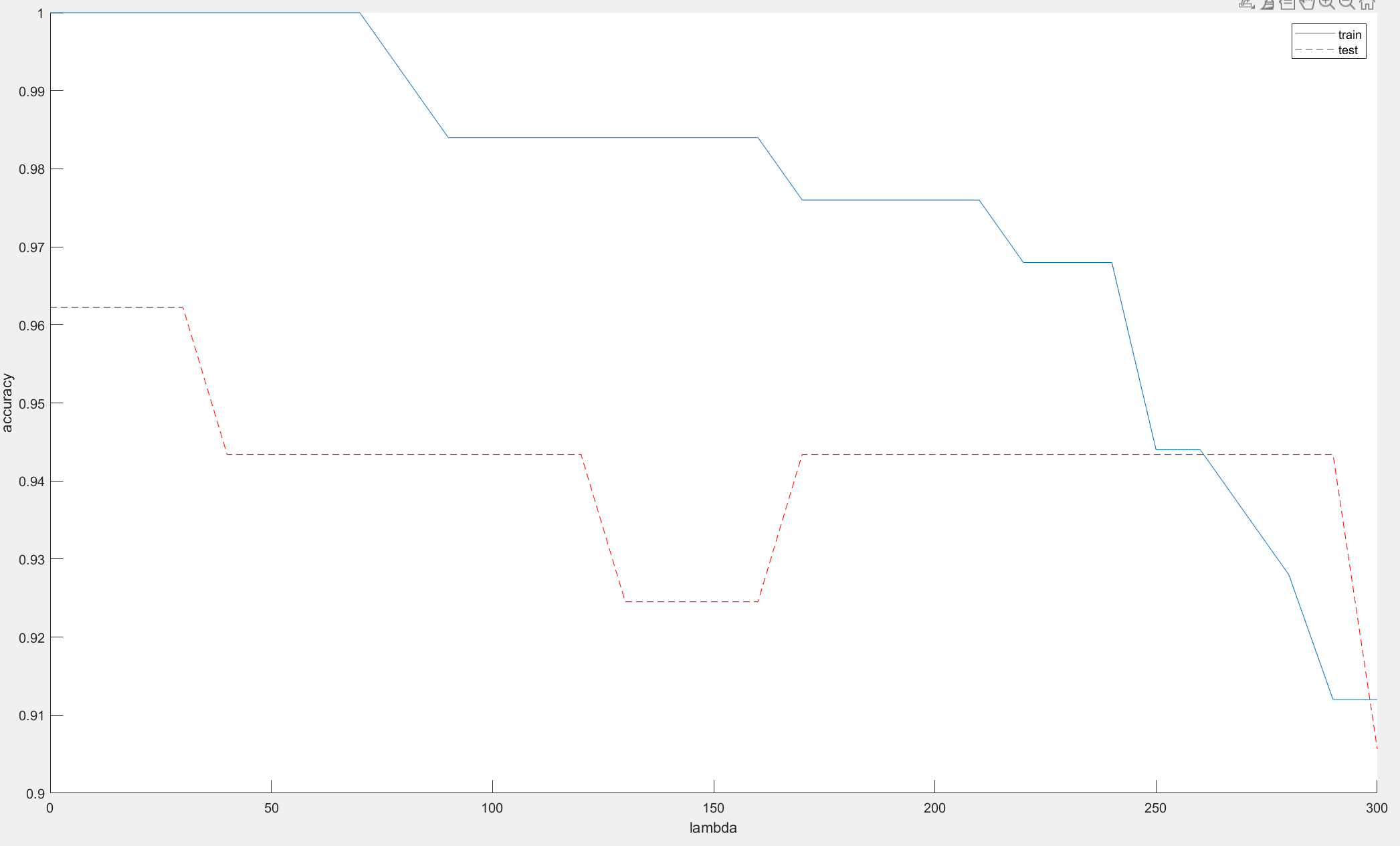
테이블이(가) 표시된 사진

자동 생성된 설명 테이블이(가) 표시된 사진

자동 생성된 설명

다음은 시각화 자료입니다.





lambda값을 점차 늘려가며 training을 반복하였을 때 꾸준히 accuracy가 떨어지는 underfitting현상을 보여주었습니다. lambda값이 0에 가까웠을때는, data가 매우 적긴 하였지만 train\_accuracy는 100%를 보여주었고, test\_accuracy또한 96%를 넘는 정확도를 보여주었습니다.

이런식으로 hyperparameter tuning을 진행하면서 모델의 성능을 최대화 할 수 있는 hyperparameter를 찾을 수 있습니다.