2019150445 신백록

1. 제 9장의 wine data를 이용하여 logistic, SVM, CART 모형을 적합하되 nested cross validation을 적용하여 최선의 모형을 산출하라.

Wine data는 데이터가 178개 밖에 없기에 다음과 같이 nested cv에서 inner\_cv=10을 이용하여 bias를 줄였다. Logistic regression의 inner cv에서는 grid search를 통하여 규제 정도를 초모수로 조정하게 하였고, 약 98.3%의 test acc를 보였고, 0.014의 std로 outer cv간의 acc 차이가 그리 크지 않다는 것을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

SVC를 적용하였을 때는 inner cv에서의 초모수를 선형인지 비선형인지, 그리고 각각에서 에러를 얼마나 허용할 것인지에 대하여 설정하였고, 이에 대한 test acc는 98.3%, std=0.014로 logistic regression과 같은 결과가 나왔다.

텍스트이(가) 표시된 사진

자동 생성된 설명

CART를 적용하였을 때는 Standardization을 하지 않았고, 초모수를 max depth로 설정하였고, 역시 decision tree의 한계 상 전의 모델들보다 test acc가 떨어지고 std는 높은 것을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

추가적으로 PCA를 적용하고 SVM을 실시하였다. PCA의 n components를 초모수로 추가하였다. PCA를 적용하지 않고 SVM을 시행했을 때보다 acc는 약간 줄어들었지만 std 는 0으로 수렴하는 것을 볼 수 있다. 하지만 각각의 CV score을 비교해봤을 떄 차원 축소를 하지 않고 SVM을 실시했을 때랑 모든 CV에서 acc가 같거나 뒤떨어지는 성능을 보이는 것을 볼 수 있다. Final model로 Logistic Regression이나 SVM을 선택할 것이다.

텍스트이(가) 표시된 사진

자동 생성된 설명

