Classification
실습과제

엑셀파일의 db_score 데이터에 대하여 label attribute를 grade로 정의한다. feature variable은 homework, discussion, midterm 3가지 애트리뷰트만으로 구성하기로 한다. 다음 각각에 대하여 과제를 수행하시오.

(1) db_score 엑셀 파일을 읽어들여 DB 테이블로 구축하되, home, discussion, midterm, grade 만 포함하는 테이블을 구성하라. 단, grade는 A 학점은 그대로 A학점, B와 C 학점은 B학점으로 변환하고, D와 F 학점은 C학점으로 변환하여 DB에 입력하기로 한다. (20점)

(2) classification 알고리즘 중에서 SVM, Logistic Regression 두가지 알고리즘에 대하여 공부하고 알고리즘에 대한 원리 등을 요약해서 작성하고, scikit-learn 라이브러리에서 각각의 함수를 어떻게 사용하는지 설명하라. (20점)

(3) grade B 에 대한 2가지 알고리즘들의 binary classification 성능결과를 얻을 수 있는 python 프로그램을 작성하고 결과를 제시/분석하라. (30점)

(3-1) train_test_split() 를 활용하여, 2:1로 데이터를 나누었을때의 성능

(3-2) K-fold cross validation 방법으로 데이터를 5 그룹으로 나누어서 실행했을 때의 성능

- 성능은 accuracy 등 4가지 measure 사용할 것)

(4) grade A, B, C 에 대한 2가지 알고리즘들의 multi-class classification 성능결과를 얻을 수 있는 python 프로그램을 작성하고 결과를 제시/분석하라. (30점)

(4-1) train_test_split() 를 활용하여, 2:1로 데이터를 나누었을때의 성능

(4-2) K-fold cross validation 방법으로 데이터를 5 그룹으로 나누어서 실행했을 때의 성능

- 성능은 accuracy 등 4가지 measure 사용할 것)

제출물

- 보고서: 개발 단계별 설명
- 소스코드: .py 는 반드시 제출할 것.
- 실행화면 캡쳐

- 중간고사가 없기 때문에 과제 점수가 중요합니다.