

Lab 2

Binary Classification Using Logistic Regression

권용혜 (robotmany@kw.ac.kr)

남건욱 (ngotic@kw.ac.kr)

실습 과제 (1)

- 광운대학교의 남건욱 조교는 만사가 다 귀찮다. 어느 날, 남건욱 조교는 수강생이 67명인 한 과목의 평점을 분류하는 작업을 맡게 됐다. 남건욱 조교는 수강생 34명의 평점을 분류하다가 귀찮음을 느끼며 산책을 나갔다. 사과 나무에 앉아 시간을 허비하던 중 떨어지는 사과를 보며 생각했다.

“지금까지 분류한 수강생 34명의 평점을 토대로 Logistic Regression 기법을 이용하여 나머지 수강생들의 평점을 자동으로 분류해볼 수 있을까?”



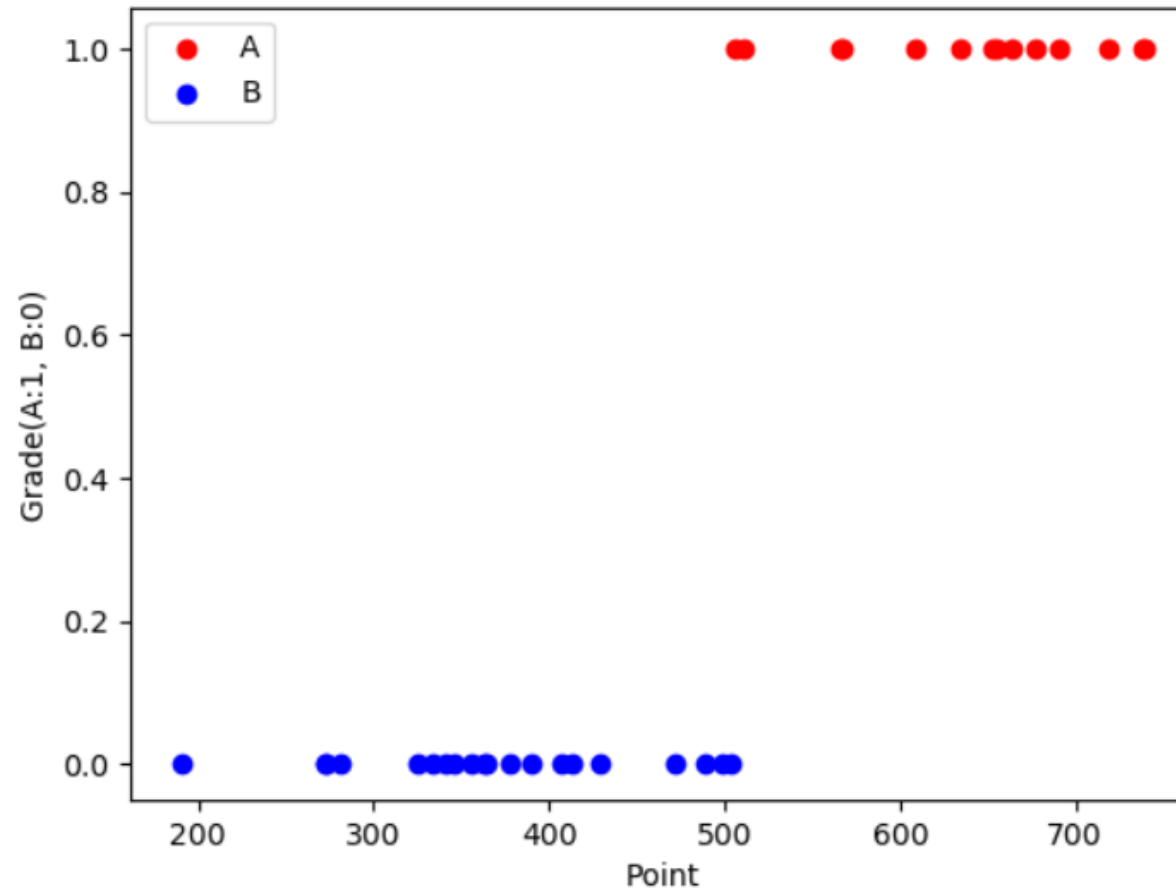
지금까지 분류한 수강생 34명의 평점을 이용하여 Logistic Regression 모델을 생성하고 학습하여, 나머지 수강생 33 명의 평점을 예측해보자!

실습 과제 (2)

- 과제의 평점은 중간고사, 기말고사, 출석, 과제 점수를 합친 **총점**으로 평가
- 평점은 $A(=1)$ 와 $B(=0)$ 로 분류
- 이미 분류한 34명의 평점에 대한 분류 기준은 알 수 없다.

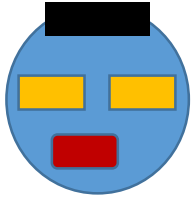
실습 과제 (3)

평점 분류를 마친 34명의 수강생의 총점 및 평점



https://github.com/developer0hye/ML_Lab/blob/master/Lab02-LogisticRegression/visualize_dataset.py

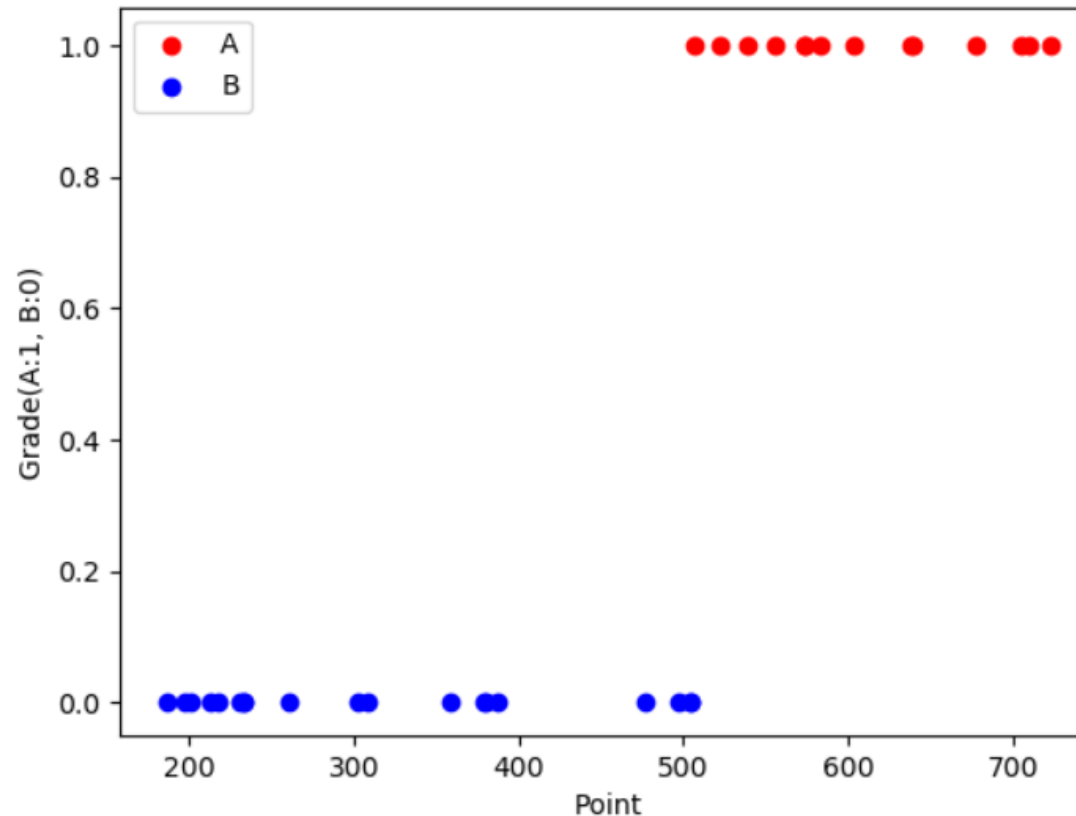
실습 과제 (4)



남건욱 조교

나머지 33 명의 수강생들의 평점도 34명의 수강생들의 정보를 이용하여 생성한 logistic regression 모델로 잘 예측됐으면 좋겠다!

나머지 33명의 수강생들의 총점 및 실제 평점



실습 과제 (5)

■ Materials

- train.txt
 - 남건욱 조교가 평점을 분류한 수강생 34명의 총점/평점이 나열된 파일
- test.txt
 - 남건욱 조교가 평점을 분류하지 않은 나머지 수강생 33명의 총점/평점이 나열된 파일
- https://github.com/developer0hye/ML_Lab/tree/master/Lab02-LogisticRegression
 - Materials 및 스켈레톤 코드(logistic_regression.cpp) 참조

■ Conditions

- C++ 이용
- Gradient Descent 기법 이용하여 logistic regression 모델의 파라미터 학습
 - Cost function 은 BCE(Binary Cross Entropy)

실습 과제 (6)

■ Criteria

- train.txt 파일의 총점/평점으로 학습된 logistic regression 모델을 이용하여 test.txt 파일의 총점을 입력 받아 평점을 예측
- 예측된 평점과 실제 평점이 동일한 경우 정답으로 판단
- 정답의 수를 기준으로 성능 평가

실습 과제 (7)

- Logistic Regression Model
 - https://github.com/developer0hye/ML_Lab/blob/master/Lab02-LogisticRegression/figures/costfunction%26gradientdescent.png
 - 위 링크에서 사진 다운로드 후 수식 참고하여 구현

실습 과제 (8)

- 결국 남건욱 조교는 귀찮음을 극복하고 성실히 평점을 분류했다.