출석수업 과제물(평가결과물) 표지(온라인제출용)

**교과목명 : 데이터마이닝**

**학 번 : 202135-368864**

**성 명 : 홍 원 표**

**강 의 실 : 서울 (2023년 5월 14일 Zoom)**

**연 락 처 : 010-5343-4341**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1.

2. 

2-①텍스트, 폰트, 스크린샷이(가) 표시된 사진

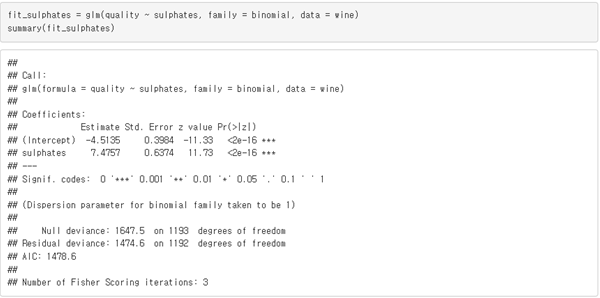
자동 생성된 설명

* 적합된 회귀 모형은 다음과 같다

- alcohol이 1단위 증가할 때마다 오즈비는 1.17배가 증가된다.

- AIC는 1385.5이다.



2-②

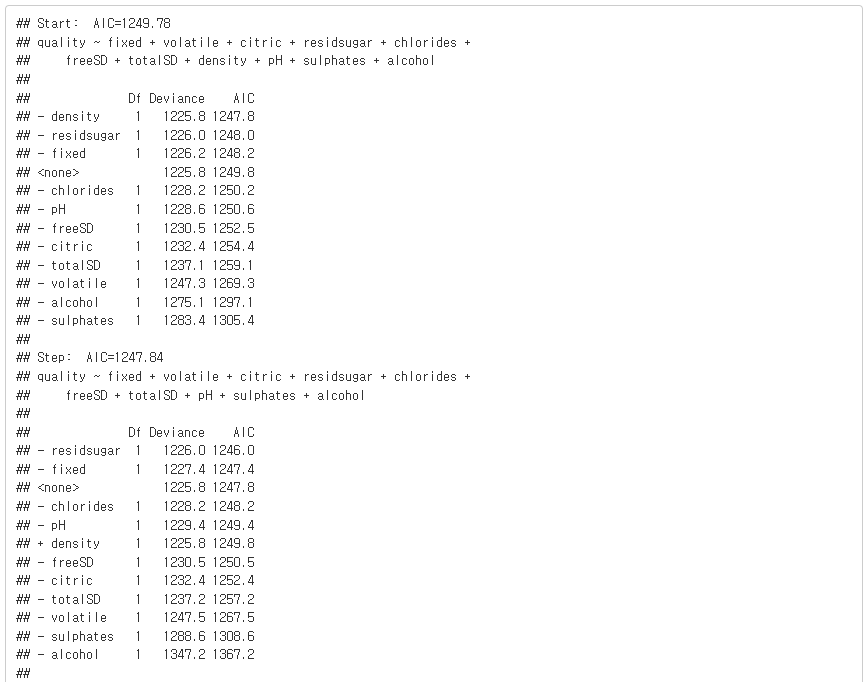
* 적합된 회귀 모형은 다음과 같다

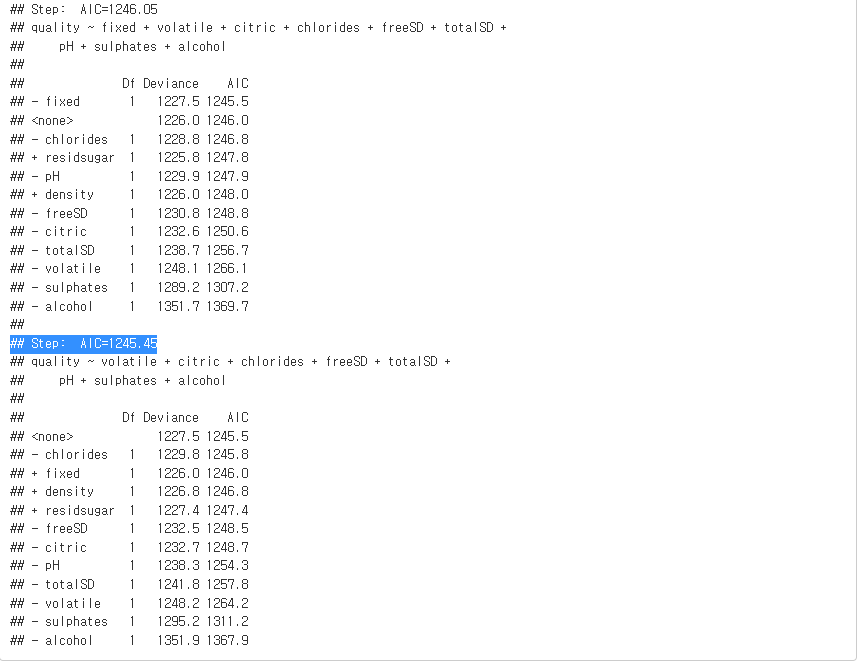
- sulphates이 1단위 증가할 때마다 오즈비는 7.48배가 증가된다.

- AIC는 1478.6이다.

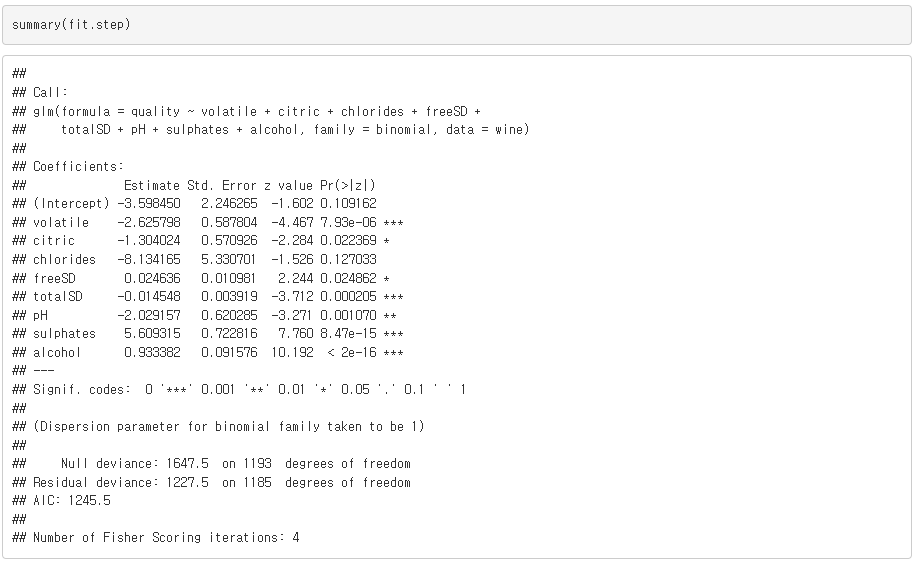


2-③





* 단계적 선택법에 의해 AIC가 제일 낮은 모형을 선택하였다.



* 단계적 선택법으로 입력변수들을 선택한 결과 density, residsugar, fixed 입력변수가 제외되고 나머지 변수들이 선택되었고 회귀식은 아래와 같다.

- AIC값은 1245.5이다.

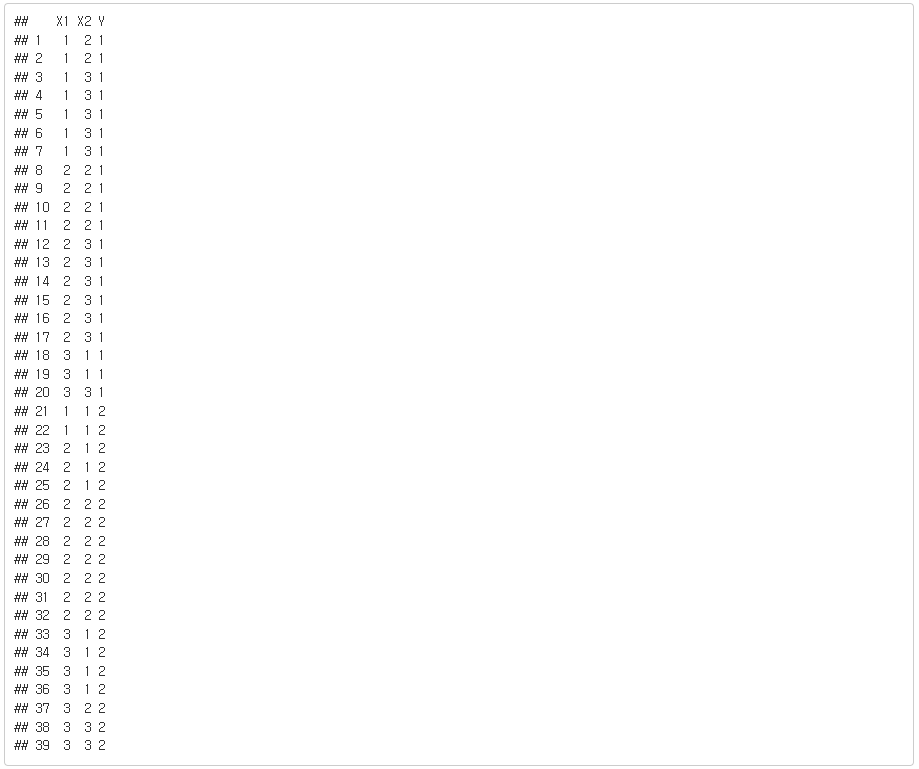


* 적합한 모형에서 얻은 예측값과 관측값으로 만든 정오분류표를 확인해보면 예측정확도는 75.1%, 오분류율은 24.9%이고 민감도는 76.7%, 특이도는 73.2%이다.

2-①②③

* ①②③ 3가지 모형의 적합도를 AIC값으로 비교해 보면 ③번 모형이 1245.5로 가장 낮은 값을 갖는것으로 보아 가장 잘 적합된 모형이다.
* 예측정확도도 확인해 보면 ①번 모형의 예측정확도는 71.2%이고 ②번 모형은 64.5% ③번 모형은 75.1%로 ③번 모형의 예측정확도가 가장 높게 나타나는 것을 알 수 있다.

3-(1)



3-(2)

텍스트, 폰트, 번호, 스크린샷이(가) 표시된 사진

자동 생성된 설명

최초분할은 X2=3과 X2=1,2인 경우로 분할이 되고 최초 분할 후 지니지수는 아래와 같다.

X2=3일 경우는

X2=1,2일 경우는

가중평균은

3-(3)

3-(2)에서 구한 의사결정 나무에서 X=3인 경우에는 Y=1인 경우는 85.71%를 차지하고 Y=2인 경우는 14.29%를 차지하고 있다. 그리고 X2=1,2인 경우에는 Y=1인 경우는 32%에 해당하고 Y=2인 경우에는 68%를 차지하고 있다.

3-(4)

root인 경우이 지니지수는 였기 때문에 한번분할 후 가중평균은 0.3669이기 때문에 불순도가 0.1328 정도 감소하는 효과가 발생했다.

* 2023년 5월 14일 데이터마이닝 출석수업 과제물 끝 -