실습1

`mpg.csv` 자료를 이용하여 다음을 코딩하라.(결측값은 제거할 것) (조건: 반응변수:`mpg01`, 예측 변수:`cylinders`, `weight`, `displacement`, `horsepower`로 지정)

- `mpg`의 값이 중앙값보다 크면 1로, 아니면 0으로 할당하는 이항 변수 자료를 만들어라. 그리고 `Auto` 자료에 `mpg01`이라는 변수명을 추가하라.
- 선형 판별 분석(LDA)을 수행하라. 이 때, 자료를 짝수년을 기준으로 훈련 자료와 시험 자료로 분리하고 전체 오차율을 구하여라.
- 이차 판별 분석(QDA)을 수행하라. 그리고 혼동 행렬(confusion matrix)을 만들고 전체 오차율을 구하여라.
- 다중 로지스틱 회귀 분석을 수행하라. 그리고 시험 자료에서 예측한 값과 실제 값을 비교하고 전체 오차율을 구하여라.
- 최근접이웃방법(KNN)을 수행하라. 그리고 K=1, K=10, K=100일 때, 각 전체 오차율을 구하여라.

실습2

`Default.txt` 자료를 이용하여 다음을 코딩하라. 단, 반응변수는 `default`(파산 여부) 이다.

- summary()와 glm()함수를 이용하여, `income`과 `balance`를 예측변수로 하는 로지스틱 모형을 적합하고 회귀 계수의 표준오차를 알아내라.
- `Default` 데이터와 인덱스를 인풋으로 하고 `income`과 `balance`의 로지스틱 회귀계수를 아웃풋으로 하는 함수 `boot.fn()`을 만들어라.
- `boot()`함수와 위에서 만든 `boot.fn()`을 이용해 `income`과 `balance`의 로지스틱 회귀계수의 표준오차 추정치를 구하라. 결과를 첫번째 문제에서 구한 값과 비교해보자.

실습3

"training.csv", "test.csv" 자료를 이용하여 다음을 코딩하라.
자료는 보험 과 관련된 자료이며, 반응변수는 clm ($Claim = 0$; 청구하지 않음, $Claim = 1$; 청구함), 예측변수는 veh_value, exposure, veh_body, veh_age, gender, area, agecat을 사용한다. 자세한 자료의 설명은 "자료설명.docx"에 있다.

- "training.csv" 자료를 이용해서 고객이 보험청구 여부를 위의 제시된 예측변수들을 이용하여 Classifier를 만들어라.
- "test.csv" 자료를 이용해, 위에서 만든 Classifier의 오분류율을 계산해 보고, 어떤 방법이 가장 잘 맞는지 비교해보자.