

- 제출일 : 10월 2일 월요일 오후 11시 59분까지
- 보고서 형식의 파일 (.pdf 또는 .html), 사용한 code가 적힌 파일 (.R)을  
**bdi.bd.test@gmail.com** 로 제출합니다.

(R 파일은 채점자의 컴퓨터에서 똑같이 실행할 수 있어야 하며, 어떤 코드인지에 대한 주석이 필요합니다.)

- 파일명, 메일제목은 이름으로 해주세요.

### 문제 1

Suppose we have a data set with five predictors,  $X_1 = \text{GPA}$ ,  $X_2 = \text{IQ}$ ,  $X_3 = \text{Gender}$  (1 for Female and 0 for Male),  $X_4 = \text{Interaction between GPA and IQ}$ , and  $X_5 = \text{Interaction between GPA and Gender}$ . The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get  $\hat{\beta}_0 = 50$ ,  $\hat{\beta}_1 = 20$ ,  $\hat{\beta}_2 = 0.07$ ,  $\hat{\beta}_3 = 35$ ,  $\hat{\beta}_4 = 0.01$ ,  $\hat{\beta}_5 = -10$ .

1. Which answer is correct, and why?
  - (a) For a fixed value of IQ and GPA, males earn more on average than females.
  - (b) For a fixed value of IQ and GPA, females earn more on average than males.
  - (c) For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.
  - (d) For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.
2. Predict the salary of a female with IQ of 110 and a GPA of 4.0.
3. True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

## 문제 2

mpg 자료를 이용하여 다음을 코딩하라.(결측값은 제거할 것). (조건: 반응 변수: mpg01, 예측 변수: cylinders, weight, displacement, horsepower로 지정)

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

### 문제 3

강의 홈에 제공된 "data1.xlsx"에 대하여 "ideo\_self" 예측하는 모형을 구축하고 ideo\_self 관측되지 않은 자료("NA"자료)에 대한 예측을 시행하여라. 예측 모형을 구축함에 있어 그 과정을 설명함으로서 제안된 예측 모형의 타당성을 이야기 하여야 한다.

#### 문제 4

강의 홈에 제공된 "data2.xlsx"을 이용하여 의원 들의 군집을 파악하여 보아라. 문제 3과 같이 군집을 찾아가는 과정에 대한 설명을 제시함으로서 분석의 타당성을 이야기 하여야 한다. (단,수업시간에 설명한 것과 다르게 자료의 대각원소는 해당 의원이 단독으로 (대표)발의한 법안의 수이다.)