**과제2. 분류**

**R Script 맨 위에 본인 이름 기재**

사용패키지: rpart, pROC, randomForest

A 대학교의 입학 담당자인 안씨는 입학 지원자의 과목별 성적외에 다른 요소와 함께 합격/불합격을 분류하는 분석을 수행하고자 한다.

데이터는 탭으로 분리된 “ClassData.txt” 파일로부터 header와 함께 읽어온다.

PassFail 변수가 타겟 변수이므로, factor로 변환한다.

(1) 데이터 분리 (10점)

set.seed(1809)를 설정하고 sample함수를 사용해 훈련 데이터와 테스트 데이터를 7:3으로 나눈다.

### tip: sample함수 사용법 ###############

set.seed(1809)

Ind <- sample(2, n, replace=T, prob=(0.7,0.3))

#####################################

(2) rpart

set.seed(1809)를 설정하고 rpart를 사용해 분류를 수행하고, 훈련된 모델의 confusion matrix를 생성과 정확도를 구한다. (3점)

테스트 세트에 대한 confusion matrix를 생성하고 정확도를 구한다. (3점)

예측데이터로 AUC를 구한다. (4점)

### tip: 훈련모델로 예측한 결과 혼동행렬 구하기 #########

table(predict(mdl\_rp, newdata=tt, type="class"), tt$PassFail)

################################################

(3) randomForest (10점)

set.seed(1809)를 설정하고 ntree=2001로 하여 분류를 수행하고, 훈련된 모델의 confusion matrix를 생성과 정확도를 구한다. (3점)

테스트 세트에 대한 confusion matrix를 생성하고 정확도를 구한다. (3점)

예측데이터로 AUC를 구한다. (4점)