**9/4 과제: 서포트 벡터 머신 & 나무 모형**

**데이터 설명: Heart Disease UCI (heart.csv)**

age : age in years

sex : (1 = male; 0 = female)

cp : chest pain type

trestbps : resting blood pressure (in mm Hg on admission to the hospital)

chol : serum cholestoral in mg/dl

fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

restecg : resting electrocardiographic results

thalach : maximum heart rate achieved

exang : exercise induced angina (1 = yes; 0 = no)

oldpeak : ST depression induced by exercise relative to rest

slope : the slope of the peak exercise ST segment

ca : number of major vessels (0-3) colored by flourosopy

thal : 3 = normal; 6 = fixed defect; 7 = reversable defect

target : have disease or not (1=yes, 0=no)

1. **SVM(서포트 벡터 머신)을 이용하여 Classification문제를 해결해보자 (10점)**

( √ target변수를 as.factor() 함수를 사용해서 범주형 변수로 변환 시켜줘야 함)

**1.1. heart데이터의 2/3을 학습 데이터셋, 1/3을 테스트 데이터 셋으로 분할 하라. (seed값은 1000으로)**

**1.2. 다양한 커널을 사용해서 SVM모델을 만들어라. (커널: radial base, polynomial, sigmoid, linear 등등). 어떤 모델이 가장 분류를 잘 하는가? 그 판단 근거는 무엇인가?**

* 1. **Tree모델(model-a)을 만들어라. 정확도는?**
  2. **가지치기를 해서 Tree 모델(model-b)을 만들어라. (seed값은 1000) 이때 최적 터미널 노드 수와 정확도는?**
  3. **model-a와 model-b 중 정확도가 더 높은 모델은? 그 이유는 무엇인가?**