DS 06주차 과제  
결정나무의 가지치기

* 이과제에서는 사전가지치기와 사후가지치기를 적용하여 결정나무를 생성하고
* 두 가지치기 방법에 대하여 성능 비교를 한다.
* Titanic data 를 사용 (이외 데이터 사용가능)
* 목표 변수는 Survived이며 아래와 같은 예측 변수를 사용할 것
  + Pclass
  + Sex
  + Age
  + Fare

1. EDA
   1. 데이터에 대하여 간단한 데이터 탐색과정 – EDA 를 실행하라.
2. 데이터 전처리
   1. 결측치(missing data)에 대하여 조사하고 결측치를 적절히 채워 넣을 것.
3. 사전 가치치기 - Pre-pruning
   1. Maxtree depth를 최적화
4. 사후 가지치기 - Post-pruning
   1. cp를 최적화
5. 위에서 생성된 두 예측 모델(사전, 사후 가지치기)의 성능을 비교하라.
   1. 각 모델의 정확도, 민감도, 특이도를 비교하라.
6. Plot 등을 이용하여 모델이 최적화되었음을 보일 것.
7. 위의 두 모델 중 하나로부터 rule set 을 추출 할 것
   1. <https://stackoverflow.com/questions/36401411/extracting-information-from-the-decision-rules-in-rpart-package/51680932#51680932>
   2. 추출된 rule 중 가장 중요한 rule 은 무엇인가?  
      이 rule 이 적용되는 사례의 비율
   3. 추출된 rule 중 가장 신뢰도가 높은 rule은 무엇인가?  
      이 rule 이 적용되는 경우의 정확도를 참조

- 끝 -