팀과제: Naïve Bayesian – 자유주제

자유주제: 금주의 학습내용과 관련된 주제로 팀과제를 자유롭게 정한다.   
Data:

1. 팀과제의 목표를 정하라.

독립성이 비교적 큰 데이터 셋과 독립성이 비교적 적은 데이터 셋에서 분류 트리와 나이브 베이지안 모델을 각각 모델링하고, 정확도를 비교해 본다.

1. 세부 목표를 3, 4 항목으로 정하라.
2. 독립성이 작은 데이터 셋 탐색 및 독립성 검증
3. 모든 변수를 독립변수 값으로 하여 분류 트리, 나이브 베이지안 모델 생성해 보기.
4. 일부 변수를 독립변수 값으로 하여 분류 트리, 나이브 베이지안 모델 생성해 보고, (2)의 정확도와 비교해서 고찰하기.
5. 독립성이 큰 데이터 셋 탐색 및 독립성 검증
6. (4)에서 찾은 데이터 셋이 너무 방대한 관계로 종속 변수와 상관관계가 높은데 다른 독립변수들과는 상관관계가 낮은, 즉 독립성이 큰 변수를 선정한다.
7. (5)에서 고른 변수들을 이용해 분류 트리와 나이브 베이지안 모델을 생성하고, 정확도를 비교하고 고찰한다.
8. 독립성이 작은 데이터 셋으로 만든 나이브 베이지안 모델과 얼마나 정확도가 차이나는지 고찰한다.
9. 위에서 기술한 목표 및 세부 목표를 달성하기 위하여 팀과제의 제목을 정하라.   
   여러 데이터 셋에서의 정확도 비교
10. 수행할 팀과제의 내용을 설정하라.   
    필요에 따라 여러 개의 문항들로 나누어 순서대로 기술할 수 있다.   
    1) 독립성이 비교적 작은 데이터셋(iris data set) 이용
    * 1. Iris data set의 전처리(결측치 확인, 깨진 부분 확인, 정규화(한눈에 알아보기 쉽게))
      2. Iris data set의 분포를 히스토그램으로 나타내어 얼마나 정규분포와 일치하는지 정도 확인
      3. Iris data set의 모든 변수를 이용한 분류 트리와 나이브 베이지안 모델 만들어서 정확도 비교.
      4. 모든 변수 대신 독립성 큰 변수들을 이용해 나이브 베이지안 모델 만든 후 정확도 비교 및 고찰

2) 독립성이 비교적 큰 데이터 셋(attrition data set)이용

i. attrition data set의 전처리(결측치 확인, 깨진 부분 확인, 정규화(한눈에 알아보기 쉽게))

ii. attrition data set의 분포를 히스토그램으로 나타내어 얼마나 정규분포와 일치하는지 정도 확인

iii. attrition data set의 일부 변수만 사용한 분류 트리와 나이브 베이지안 모델 만들어서 정확도 비교

1. 모든 변수를 쓴 나이브 베이지안 모델과 일부 변수 쓴 나이브 베이지안 모델 만든 후 정확도 비교 및 고찰