Homework 3

Id	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	Over 100K	Yes
2	No	Married	Under 100K	No
3	Yes	Single	Over 100K	Yes
4	Yes	Married	Over 100K	Yes
5	No	Single	Over 100K	No
6	Yes	Single	Under 100K	No
7	No	Married	Over 100K	No
8	No	Single	Over 100K	No
9	Yes	Single	Under 100K	No
10	Yes	Married	Over 100K	Yes

지니계수(Gini index)를 사용하여 최대 깊이 분류 트리(maximum depth classification tree)를 찾으세요.

$$Gini = 1 - \left(\frac{\text{num of Yes}}{\text{total}}\right)^2 - \left(\frac{\text{num of No}}{\text{total}}\right)^2$$

$$= 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{6}{10}\right)^2$$

$$= 1 - 0.16 - 0.36 = 1 - 0.52$$

$$= 0.48$$

1. Refund 속성
   - Refund = Yes
     - Cheat = Yes : 4개
     - Cheat = No : 2개
   - Refund = No
     - Cheat = Yes : 0개
     - Cheat = No : 4개

   → 지니 계수 : 0.264

2. Material Status 속성
   - Material Status = Single
     - Cheat = Yes : 2개
     - Cheat = No : 4개
   - Material Status = married
     - Cheat = Yes : 1개
     - Cheat = No : 2개

   → 0.496

3. Taxable income 속성
   - over look
     - Cheat = Yes = 4
     - Cheat = No : 4
   - Under look
     - Cheat = Yes : 0
     - Cheat = No : 2

   → 0.4

Refund Gini 계수↓ : 최빈값 분할

→ 각 분할에 대해 반복