

Homework 3 - 20기 모수아

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.

Try to guess value of cheat based on the other values

$$Gini(Refund=Yes) = \frac{6}{10} \left(1 - \frac{4}{9} - \frac{1}{9} \right) + \frac{4}{10} (1 - 0 - 1) = \frac{4}{15}$$

$$Gini(Refund=No) = \frac{4}{10} (1 - 0 - 1) + \frac{6}{10} \left(1 - \frac{4}{9} - \frac{1}{9} \right) = \frac{4}{15}$$

$$Gini(Marital\ Status=Single) = \frac{6}{10} \left(1 - \frac{1}{9} - \frac{4}{9} \right) + \frac{4}{10} \left(1 - \frac{1}{4} - \frac{1}{4} \right) = \frac{7}{15}$$

$$Gini(Marital\ Status=Married) = \frac{4}{10} \left(1 - \frac{1}{4} - \frac{1}{4} \right) + \frac{6}{10} \left(1 - \frac{1}{9} - \frac{4}{9} \right) = \frac{7}{15}$$

$$Gini(Taxable\ Income=Over\ 100K) = \frac{7}{10} \left(1 - \frac{16}{49} - \frac{9}{49} \right) + \frac{3}{10} (1 - 0 - 1) = \frac{12}{35}$$

$$Gini(Taxable\ Income=Under\ 100K) = \frac{3}{10} (1 - 0 - 1) + \frac{7}{10} \left(1 - \frac{16}{49} - \frac{9}{49} \right) = \frac{12}{35}$$