Homework 2

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

$$6 \quad 4 \qquad 6 \quad 4 \qquad 7 \quad 3$$

지니계수(Gini index)를 사용하여 최초 split이 되는 기준을 찾으세요.

1) Refund

$$Gini_{Refund} = \frac{6}{10} Gini_{Yes} + \frac{4}{10} Gini_{No} = \frac{6}{10} \cdot 0.5 + \frac{4}{10} \cdot 0.375 = 0.45$$

$$Gini_{Yes} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

$$Gini_{No} = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 0.375$$

2) Martial Status

$$Gini_{Martial} = \frac{6}{10} \cdot Gini_{single} + \frac{4}{10} Gini_{married} = \frac{6}{10} \cdot 0.5 + \frac{4}{10} \cdot 0.375 = 0.45$$

$$Gini_{single} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

$$Gini_{married} = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 0.375$$

3) Taxable Income

$$Gini_{Income} = \frac{7}{10} \cdot Gini_{over} + \frac{3}{10} Gini_{under} = \frac{7}{10} \cdot 0.49 + \frac{3}{10} \cdot 0 = 0.343$$

$$Gini_{over} = 1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2 = 0.49$$

$$Gini_{under} = 1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2 = 0$$

Taxable Income의 Gini Index가 가장 작으므로 최초의 분할기준으로 선택