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

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

1. $Gini_{refund}(D)$

$$Gini_{refund=yes} = \frac{6}{10}\left(1-\left(\frac{4}{6}\right)^2-\left(\frac{2}{6}\right)^2\right) + \frac{4}{10}\left(1-\left(\frac{0}{4}\right)^2-\left(\frac{4}{4}\right)^2\right)$$

$$= 0.27$$

2. $Gini_{marital\ status}(D)$

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

$$= 0.27 + 0.2 = 0.47$$

3. $Gini_{taxable\ income}(D)$

$$Gini_{TI=over\ 100k} = \frac{7}{10}\left(1-\left(\frac{4}{7}\right)^2-\left(\frac{3}{7}\right)^2\right) + \frac{3}{10}\left(1-\left(\frac{0}{3}\right)^2-\left(\frac{3}{3}\right)^2\right)$$

$$= 0.34$$

$$Min\ [\ 0.27,\ 0.47,\ 0.34\ ] = 0.27$$

∴ Refund를 첫번째 기준으로 선택한다.