Homework 3

Yes : 6	S : 6	Over : 7	Yes : 4
No : 4	m : 4	Under : 3	No : 6

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.

$$\begin{cases} Gini(A) = \sum_{j=1}^{2} \frac{|D_j|}{|D|} * Gini(D_n) \\ Gini(D_n) = 1 - \sum_{j=1}^{q} P_j^2 \end{cases}$$

$Gini(Refund) = \frac{6}{10}\left(1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2\right) + \frac{4}{10}\left(1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2\right) = 0.2667$

$Gini(Marital\ Status) = \frac{6}{10}\left(1 - \left(\frac{2}{6}\right)^2 - \left(\frac{4}{6}\right)^2\right) + \frac{4}{10}\left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right) = 0.4667$

$Gini(Taxable\ Income) = \frac{7}{10}\left(1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2\right) + \frac{3}{10}\left(1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2\right) = 0.3429$

∴ 지니계수가 가장 작은 Refund를 기준으로 split 한다.