

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.

$$Gini_{Refund=Yes} = \frac{6}{10} \left(1 - \left(\frac{4}{6} \right)^2 - \left(\frac{2}{6} \right)^2 \right) + \frac{4}{10} \left(1 - \left(\frac{4}{4} \right)^2 - \left(\frac{0}{4} \right)^2 \right) = \frac{4}{15}$$

$$Gini_{Refund=No} = \frac{4}{10} \left(1 - \left(\frac{4}{4} \right)^2 - \left(\frac{0}{4} \right)^2 \right) + \frac{6}{10} \left(1 - \left(\frac{4}{6} \right)^2 - \left(\frac{2}{6} \right)^2 \right) = \frac{4}{15}$$

$$\rightarrow \min(Gini_{Refund}) = \frac{4}{15}$$

$$Gini_{Marital\ Status=Single} = \frac{6}{10} \left(1 - \left(\frac{2}{6} \right)^2 - \left(\frac{4}{6} \right)^2 \right) + \frac{4}{10} \left(1 - \left(\frac{1}{4} \right)^2 - \left(\frac{3}{4} \right)^2 \right) = \frac{1}{15}$$

$$Gini_{Marital\ Status=Married} = \frac{1}{15}$$

$$\rightarrow \min(Gini_{Marital\ Status}) = \frac{1}{15}$$

$$Gini_{Taxable\ Income=Over\ 100k} = \frac{7}{10} \left(1 - \left(\frac{9}{7} \right)^2 - \left(\frac{3}{7} \right)^2 \right) + \frac{3}{10} \left(1 - \left(\frac{0}{3} \right)^2 - \left(\frac{3}{3} \right)^2 \right) = \frac{12}{35}$$

$$Gini_{Taxable\ Income=Under\ 100k} = \frac{12}{35}$$

$$\rightarrow \min(Gini_{Taxable\ Income}) = \frac{12}{35}$$

\therefore Marital Status