

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.

- Refund

$$\begin{aligned}
 Gini_{\text{Refund}=\text{Yes}} &= \frac{6}{10} \left(1 - \frac{4^2}{6} - \frac{2^2}{6} \right) + \frac{4}{10} \left(1 - \frac{4^2}{4} \right) \\
 &= \frac{6}{10} \left(1 - \frac{4}{3} - \frac{1}{3} \right) = \frac{6}{10} \times \frac{4}{3} = \underline{0.267}
 \end{aligned}$$

$$\begin{aligned}
 Gini_{\text{Refund}=\text{No}} &= \frac{4}{10} \left(1 - \frac{4^2}{4} \right) + \frac{6}{10} \left(1 - \frac{4^2}{6} - \frac{2^2}{6} \right) \\
 &= \underline{0.267}
 \end{aligned}$$

- Marital Status

$$\begin{aligned}
 Gini_{\text{Marital Status}} &= \frac{6}{10} \left(1 - \frac{2^2}{6} - \frac{4^2}{6} \right) + \frac{4}{10} \left(1 - \frac{2^2}{4} - \frac{2^2}{4} \right) \\
 &= \frac{6}{10} \times \frac{4}{3} + \frac{4}{10} \times \frac{1}{2} = \underline{0.467}
 \end{aligned}$$

- Taxable Income

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

$$= \frac{7}{10} \times \frac{24}{49} = \underline{0.343}$$

∴ 2/2 Split 7/2 : Refund