

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.

$$1. \text{Gini}_{\text{refund}} = \frac{6}{10} \left(1 - \left(\frac{2}{3} \right)^2 - \left(\frac{1}{3} \right)^2 \right) + \frac{4}{10} \left(1 - \left(\frac{4}{4} \right)^2 - \left(\frac{0}{4} \right)^2 \right) = \frac{2}{5} \cdot \frac{4}{9} + 0$$

$$= \left(\frac{4}{15} \right) \approx 0.266$$

$$2. \text{Gini}_{\text{M-S}} = \text{Single} = \frac{6}{10} \left(1 - \left(\frac{2}{6} \right)^2 - \left(\frac{4}{6} \right)^2 \right) + \frac{4}{10} \left(1 - \left(\frac{1}{2} \right)^2 - \left(\frac{1}{2} \right)^2 \right)$$

$$= \frac{3}{5} \left(1 - \frac{1}{9} - \frac{4}{9} \right) + \frac{2}{5} \cdot \frac{1}{2} = \frac{4}{15} + \frac{1}{5} = \left(\frac{7}{15} \right) \approx$$

$$\text{Gini}_{\text{M-S}} = \text{Married} = \frac{4}{10} \left(1 - \left(\frac{1}{2} \right)^2 - \left(\frac{1}{2} \right)^2 \right) + \frac{6}{10} \left(1 - \left(\frac{2}{6} \right)^2 - \left(\frac{4}{6} \right)^2 \right)$$

$$= \frac{4}{10} \cdot \frac{2}{4} + \frac{2}{5} \cdot \frac{4}{9}$$

$$= \frac{1}{5} + \frac{4}{15} = \left(\frac{7}{15} \right)$$

$$\left(\text{Gini}_{\text{M-S}} = \frac{7}{15} \right) \approx 0.466$$

$$3. \text{Gini}_{\text{T-I}} = \frac{7}{10} \left(1 - \left(\frac{4}{7} \right)^2 - \left(\frac{3}{7} \right)^2 \right) + \frac{3}{10} \left(1 - \left(\frac{3}{3} \right)^2 - \left(\frac{0}{3} \right)^2 \right)$$

$$= \frac{7}{10} \left(1 - \frac{25}{49} \right) + 0 = \frac{1}{10} \cdot \frac{24}{49} = \left(\frac{12}{35} \right) \approx 0.3428$$

따: 최초 split 되는 기준 Refund