

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 기준: Cheat (Yes: 4, No: 2)

Refund Yes: Cheat (Y: 4, N: 2) $1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = \frac{4}{9}$

No: Cheat (0, 4) ~~$1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 = 0$~~

~~$\frac{2}{6} \times \frac{4}{6} + \frac{4}{6} \times \frac{2}{6} = \frac{8}{18} = \frac{4}{9}$~~

$\frac{4}{6}$

$2 \cdot 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = \frac{4}{9}$

Marital Status 기준: Single (4, 4)

M

$\frac{2}{2} \cdot 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = \frac{2}{3}$

$\frac{4}{9} \times \frac{6}{18} + \frac{4}{18} \times \frac{1}{2} = \frac{1}{3}$

$\frac{4}{15} + \frac{1}{5} = \frac{1}{3}$

Taxable Income 12

16.

Overlook: $C_{N 3}^{4 4} 1 - \frac{4}{5} - \frac{3}{5} = \frac{2}{5}$

Underlook: $C_3^0 \quad \frac{8}{9}$

$$\frac{24}{49} \times \frac{7}{10} + \frac{8}{9} \times \frac{3}{10} = \frac{24}{50} + \frac{8}{30}$$

$\frac{1}{2} \times \frac{1}{2} = 0$ Refund) $C_{\frac{2}{5} \frac{4}{5} \frac{0}{5} \frac{0}{5}}$

Refund