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.

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

$$= \frac{6}{10} \cdot \frac{4}{9} + \frac{4}{10} \cdot 0 = \boxed{0.267}$$

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

$$= \frac{6}{10} \cdot \frac{5}{9} + \frac{4}{10} \cdot \frac{1}{2} = 0.467$$

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

$$= \frac{7}{10} \cdot \frac{24}{49} + \frac{3}{10} \cdot 0 = 0.343$$

지니 계수가 0에 가까울수록 클래스에 속한 불순도가 낮으므로

지니계수가 가장 0에 가까운 Refund를 기준으로 split한다.