

Homework 3

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 최대 깊이 분류 트리(maximum depth classification tree)를 찾으세요.

$$\text{Refund} : Y=6, N=4 \quad \text{Taxable} : O=7, U=3$$

$$\text{Marital} : S=6, M=4 \quad \text{Cheat} : Y=4, N=6$$

$$G(\text{Refund} = \text{Yes}) = \frac{6}{10} \left(1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2\right) + \frac{4}{10} \left(1 - \left(\frac{4}{4}\right)^2\right) = 0.267$$

$$G(\text{Marital} = \text{Single}) = \frac{6}{10} \left(1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2\right) + \frac{4}{10} \left(1 - \left(\frac{4}{4}\right)^2\right) = 0.467$$

$$G(\text{Taxable} = \text{Over}) = \frac{7}{10} \left(1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2\right) + \frac{3}{10} \left(1 - \left(\frac{3}{3}\right)^2\right) = 0.347$$

