

## 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) refund

$$\frac{6}{10} \left( 1 - \left( \frac{4}{6} \right)^2 - \left( \frac{2}{6} \right)^2 \right) + \frac{4}{10} \left( 1 - \left( \frac{0}{4} \right)^2 - \left( \frac{4}{4} \right)^2 \right) = 0.29$$

2) marital status

$$\frac{6}{10} \left( 1 - \left( \frac{2}{6} \right)^2 - \left( \frac{4}{6} \right)^2 \right) + \frac{4}{10} \left( 1 - \left( \frac{2}{4} \right)^2 - \left( \frac{2}{4} \right)^2 \right) = 0.47$$

3) income

$$\frac{7}{10} \left( 1 - \left( \frac{4}{7} \right)^2 - \left( \frac{3}{7} \right)^2 \right) + \frac{3}{10} \left( 1 - \left( \frac{0}{3} \right)^2 - \left( \frac{3}{3} \right)^2 \right) = 0.34$$

∴ 최초 split 가점을 refund이다.