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.

Target : "cheat"

$Gini_r(D)$: $Gini_{r=Yes} = \frac{6}{10}(1-(\frac{4}{6})^2-(\frac{2}{6})^2)+\frac{4}{10}(1-0^2-1^2) = 0.2667$

$Gini_{r=No} = \frac{4}{10}(1-0^2+1^2)+\frac{6}{10}(1-(\frac{4}{6})^2-(\frac{2}{6})^2) = 0.2667$

$Gini_m(D)$: $Gini_{m=Married} = \frac{4}{10}(1-(\frac{2}{4})^2-(\frac{2}{4})^2)+\frac{6}{10}(1-(\frac{2}{6})^2-(\frac{4}{6})^2) = 0.4667$

$Gini_{m=Single} = \frac{6}{10}(1-(\frac{2}{6})^2-(\frac{4}{6})^2)+\frac{4}{10}(1-(\frac{2}{4})^2-(\frac{2}{4})^2) = 0.4667$

$Gini_T(D)$: $Gini_{T=Over} = \frac{7}{10}(1-(\frac{4}{7})^2-(\frac{3}{7})^2)+\frac{3}{10}(1-0^2-1^2) = 0.3429$

$Gini_{T=Under} = \frac{3}{10}(1-0^2-1^2)+\frac{7}{10}(1-(\frac{4}{7})^2-(\frac{3}{7})^2) = 0.3429$

최초 split 기준 : "Refund"