date. page.

1° 设10个 数据 编号分别 为1,2,3,4,5,...,10,

则 root node 的 熵为:

$H_0 = -0.6 \times \log\frac{6}{10} + -0.4 \times \log\frac{4}{10} = 0.97095$

Age 的中间值为: 37.5

Age > 37.5 的为: 2, 6, 8, 9, 10, Age < 37.5 的为: 1, 3, 4, 5, 7

二、以 Age 划分之后的 总分支 熵为:

$H_1 = \frac{5}{10} \times \left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) + \frac{1}{2}\left(\frac{4}{5}\log\frac{4}{5} + \frac{1}{5}\log\frac{1}{5}\right) = 0.84644$

Salary 的中间值为: 66,000,

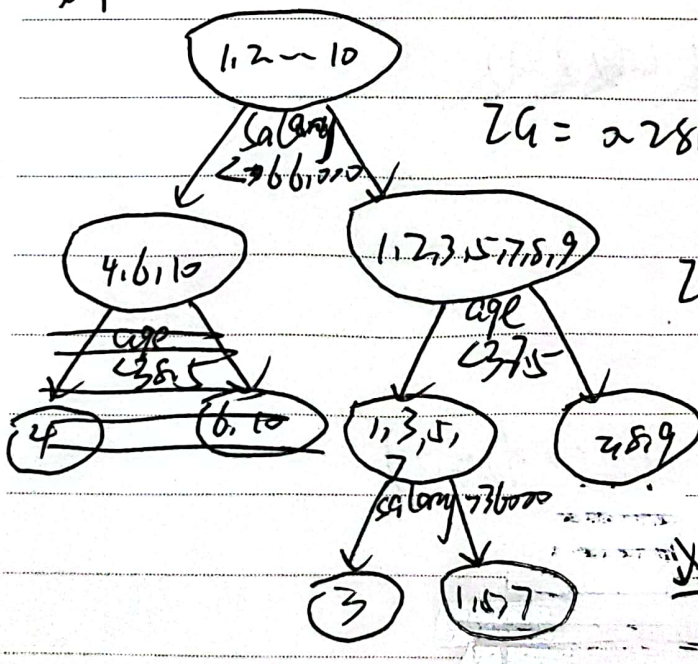
Salary > 66,000 的有: 4, 6, 10, Salary < 66,000, 有 1, 2, 3, 5, 7, 8, 9

二、以 Salary 划分之后总分支 熵为:

$H_2 = \frac{3}{10} \times 0 + \frac{7}{10} \times \left(\frac{4}{7}\log\frac{4}{7} + \frac{3}{7}\log\frac{3}{7}\right) = 0.68966$

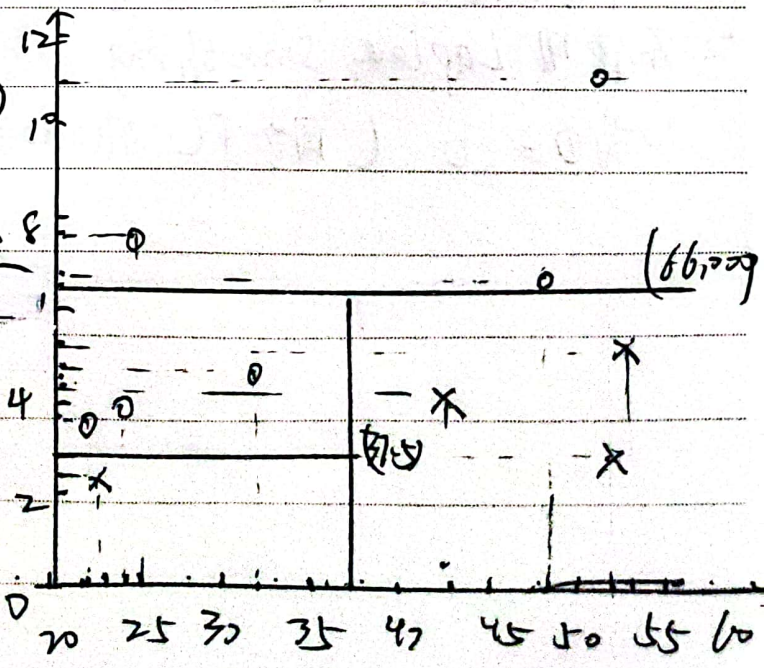
二、第一次划分 以 Salary 进行划分, 之后划分同理 decision tree

如下:



$Z_G = 0.28129$

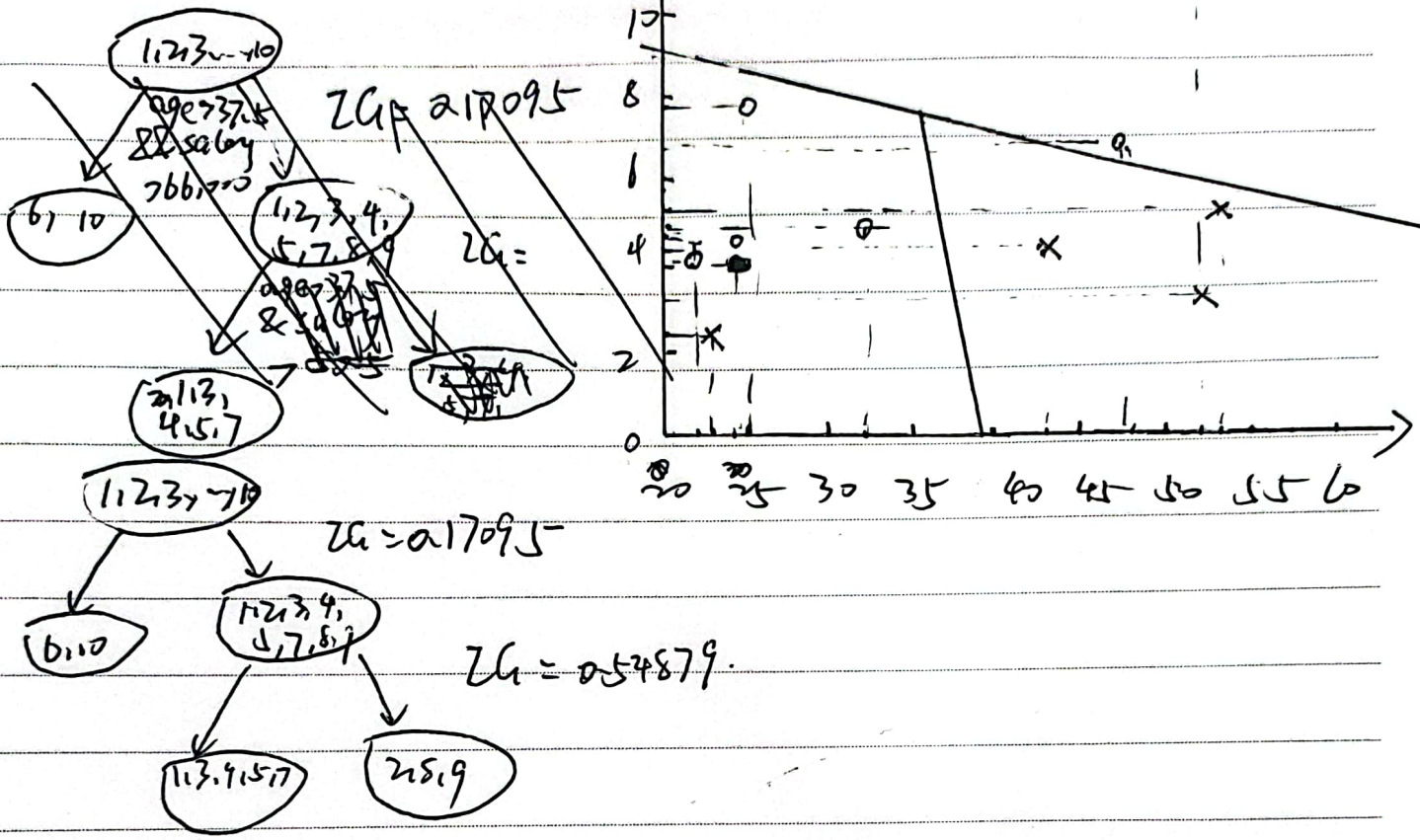
$Z_G = 0.2764$



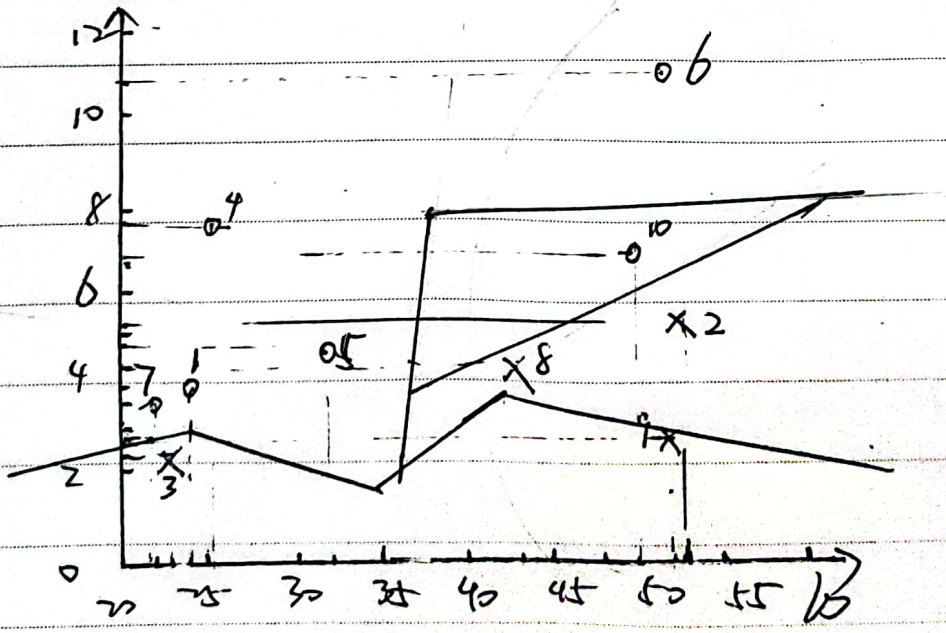
对于第一次划分，

b. 取 $\alpha>1$, $\beta>1$ 即只有：age > 37.5 && salary > 66,000 为平价未治依据

则 decision tree 如下：



c.



date. page.

a:

a. $P(\text{Cold}, \text{Headache}, \text{Cough}, \text{SoreThroat})$

$= P(\text{Headache}|\text{Cold})\,P(\text{Cough}|\text{Cold})\,P(\text{SoreThroat}|\text{Cold})\,P(\text{Cold})$

$+P(\text{Headache}|\neg\text{Cold})\,P(\text{Cough}|\neg\text{Cold})\,P(\text{SoreThroat}|\neg\text{Cold})\,P(\neg\text{Cold})$

b. $= \frac{1}{2} \times 1 \times \frac{2}{3} \times \frac{2}{3} = \frac{2}{9}$

$P(\text{Cold}|\neg\text{Headache}, \text{Cough}, \text{SoreThroat})$

$= \dfrac{P(\text{Cold})\,P(\neg\text{Headache}|\text{Cold})\,P(\text{Cough}|\text{Cold})\,P(\text{SoreThroat}|\text{Cold})}{P(\neg\text{Headache}, \text{Cough}, \text{SoreThroat})}$ ①

$P(\text{Cold}) = \frac{1}{2}$, $P(\neg\text{Cold}) = \frac{1}{2}$,

$P(\neg\text{Headache}|\text{Cold}) = 0$, $P(\text{Headache}|\text{Cold}) = 1$

$P(\text{Cough}|\text{Cold}) = \frac{2}{3}$, $P(\neg\text{Cough}|\text{Cold}) = \frac{1}{3}$

$P(\text{SoreThroat}|\text{Cold}) = \frac{2}{3}$

$P(\neg\text{Headache}, \text{Cough}, \text{SoreThroa}$

⇒ 用 Laplace Smooth 后 ~~$P(\neg\text{Head}$~~ (取 $\alpha=1$)

$\hat{P}(\neg\text{Headache}|\text{Cold}) = \frac{1}{6}$,

$\hat{P}(\text{Cough}|\text{Cold}) = \frac{1}{2}$,

$\hat{P}(\text{SoreThroat}|\text{Cold}) = \frac{1}{2}$,

∴ 不使用 Laplace Smoothing,

∴ 式① = 0 (因为 $P(\neg\text{Headache}|\text{Cold}) = 0$)