1、特征 $A_1 \cdots A_4$ 表示 年龄，工作，房子，信贷 情况

$$H(D) = -\frac{6}{15}\log_2\frac{6}{15} - \frac{9}{15}\log_2\frac{9}{15} = 0.971$$

$$g(D,A_1) = H(D) - H(D|A_1) = 0.971 + \left[\frac{5}{15}\times\left(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}\right) + \frac{5}{15}\times\left(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\times\log_2\frac{3}{5}\right) + \frac{5}{15}\left(\frac{4}{5}\times\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5}\right)\right]$$

$$= 0.083$$

$$g(D,A_2) = 0.971 - \left[\frac{5}{15}\times 0 + \frac{10}{15}\times\left(-\frac{6}{10}\log_2\frac{6}{10} - \frac{4}{10}\log_2\frac{4}{10}\right)\right] = 0.324$$

$$g(D,A_3) = 0.971 - \left[\frac{6}{15}\times 0 + \frac{9}{15}\times\left(-\frac{3}{9}\log_2\frac{3}{9} - \frac{6}{9}\log_2\frac{6}{9}\right)\right] = 0.420$$

$$g(D,A_4) = 0.971 - \left[\frac{5}{15}\times\left(-\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5}\right) + \frac{6}{15}\times\left(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}\right) + \frac{4}{15}\times 0\right] = 0.363$$

$$H_{A_1}(D) = -\frac{5}{15}\log_2\frac{5}{15}\times 3 = 1.585$$

$$H_{A_2}(D) = -\frac{5}{15}\log_2\frac{5}{15} - \frac{10}{15}\log_2\frac{10}{15} = 0.918$$

$$H_{A_3}(D) = -\frac{6}{15}\log_2\frac{6}{15} - \frac{9}{15}\log_2\frac{9}{15} = 0.971$$

$$H_{A_4}(D) = -\frac{5}{15}\log_2\frac{5}{15} - \frac{6}{15}\log_2\frac{6}{15} - \frac{4}{15}\log_2\frac{4}{15} = 1.566$$

$$\Rightarrow \quad g_R(D,A_1) = \frac{g(D,A_1)}{H_{A_1}(D)} = 0.052$$

$$g_R(D,A_2) = \cancel{0.354}\ 0.353$$

$$g_R(D,A_3) = 0.433$$

$$g_R(D,A_4) = 0.232$$

选择 A_3（是否有房子）作为最优特征对 D 进行划分，分为 D_1（有房子） D_2（没有房子）

D_1中类均为"是"，将其作为叶结点

对D_2: $H(D_2) = -(\frac{3}{9}\log_2\frac{3}{9} + \frac{6}{9}\log_2\frac{6}{9}) = 0.918$

有 ~~$g(D_2,A_1) = H(D_2) + [\frac{4}{9}\log_2\frac{4}{9} + \frac{2}{9}\log_2\frac{2}{9} + \frac{3}{9}\log_2\frac{3}{9}]$~~

$g(D_2,A_1) = 0.918 - [\cancel{5}\frac{4}{9}\times(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) + \frac{2}{9}\times 0 + \frac{3}{9}\times(-\log_2\frac{2}{3}\times\frac{2}{3} - \frac{1}{3}\log_2\times\frac{1}{3})] = 0.251$

$g(D_2,A_2) = 0.918 - [\frac{6}{9}\times 0 + \frac{3}{9}\times 0] = 0.918$

$g(D_2,A_4) = 0.918 - [\frac{4}{9}\times 0 + \frac{4}{9}\times(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}) + \frac{1}{9}\times 0] = 0.474$

$H_{A_1}(D_2) = -\frac{4}{9}\log_2\frac{4}{9} - \frac{2}{9}\log_2\frac{2}{9} - \frac{3}{9}\log_3\frac{3}{9} = 1.53$

$H_{A_2}(D_2) = -\frac{6}{9}\log_2\frac{6}{9} - \frac{3}{9}\log_2\frac{3}{9} = 0.918$

$H_{A_4}(D_2) = -\frac{4}{9}\log_2\frac{4}{9} - \frac{4}{9}\log_2\frac{4}{9} - \frac{1}{9}\log_2\frac{1}{9} = 1.392$

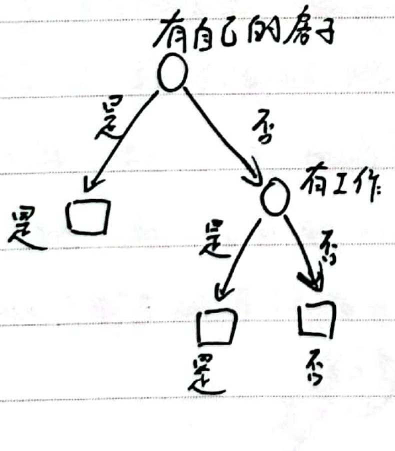
$g_R(D_2,A_1) = \frac{g(D_2,A_1)}{H_{A_1}(D_2)} = 0.164$

$g_R(D_2,A_2) = 1$

$g_R(D_2,A_4) = 0.34$

选择 A_2（有工作）作为最优特征，分为D_1'(有工作), D_2'(没有工作)

D_1'中类均为是，D_2'中类均为否。将D_1', D_2'作为叶结点，生成决策树



2. 选择划分点 s, $R_1(s)=\{x|x\le s\}$ $R_2(s)=\{x|x>s\}$

使 $\min_s\left[\min_{c_1}\sum_{x_i\in R_1(s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(s)}(y_i-c_2)^2\right]$

取 $s=1$, $c_1=y_1=4.5$, $c_2=\frac{1}{9}\sum_{i=2}^{10}y_i=6.85$

平方损失 $m(s)=0+\sum_{i=2}^{10}(y_i-c_2)^2=22.65$

取 $s=2$, $c_1=\frac{1}{2}(y_1+y_2)=4.625$, $c_2=\frac{1}{8}\sum_{i=3}^{10}y_i=7.12$

$m(s)=\sum_{i=1}^{2}(y_i-c_1)^2+\sum_{i=3}^{10}(y_i-c_2)^2=17.7$

分别取 $s=1,2\cdots10$，按相同方法计算 $m(s)$

得: $m(3)=12.19$, $m(4)=7.38$, $m(5)=3.36$, $m(6)=5.07$,

$m(7)=10.05$ $m(8)=15.18$ $m(9)=21.33$ $m(10)=27.63$

$m(5)$ 最小，取第一次分割点为 $s=5$

对 $x\le5$ 的部分，分别取 $s=1\cdots5$

$s=1$, $c_1=y_1=4.5$, $c_2=\frac{1}{4}\sum_{i=2}^{5}y_i=5.2$, $m(1)=0+\sum_{i=2}^{5}(y_i-5.2)^2=0.67$

按相同的方法计算 $m(s)$，得到: $m(2)=0.43$, $m(3)=0.19$, $m(4)=0.37$, $m(5)=1.06$

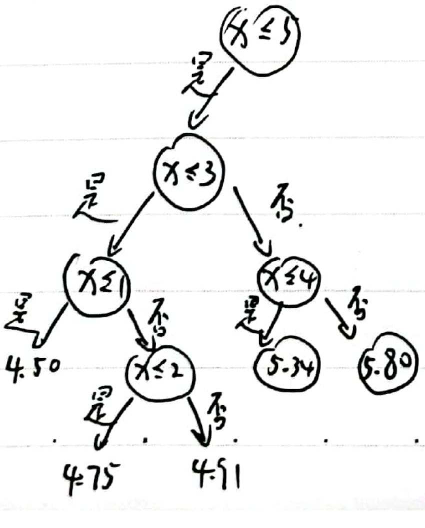
选择 $s=3$ 作为划分点，按相同的方法进行划分即可

$1\le x\le3$ 时，取 $s=1$ 时 $m(s)$ 最小，$m(1)=0+\left(y_2-\frac{y_2+y_3}{2}\right)^2+\left(y_3-\frac{y_2+y_3}{2}\right)^2=0.013$

$1<x\le3$ 时，取 $s=2$ $m(s)$ 最小，$m(2)=0$

$3<x\le5$ 时，取 $s=4$ 时 $m(s)$ 最小，$m(4)=0$

选择 $x\le5$ 部分的回归树为



对 $x>5$ 的部分同理可得

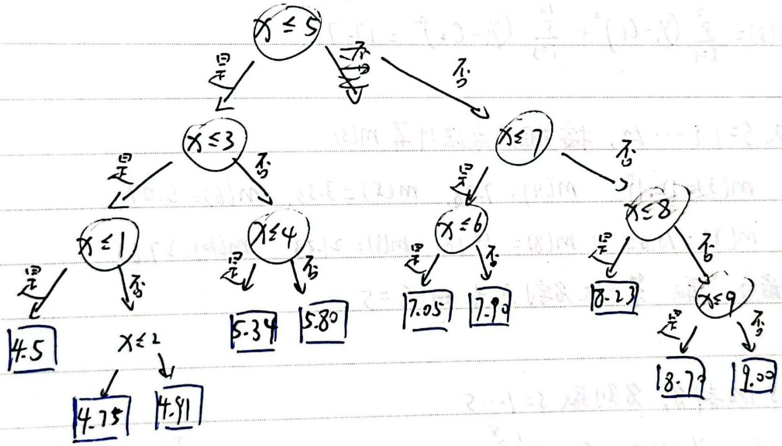
$s=7$ 时，$m(s)$ 最小，$x>5$ 可分为 $5<x\le 7$ 和 ~~$7<x\le 10$~~ $x>7$

对 $5<x\le 7$，取 $s=6$ 时 $m(s)$ 最小，分为 $5<x\le 6$ 和 $6<x\le 7$，为叶结点

对 $7<x\le 10$，取 $s=8$ 时 $m(s)$ 最小，分为 $7<x\le 8$ 和 ~~$8<x\le 10$~~ $x>8$

对 $8<x\le 10$，取 $s=9$ 时 $m(s)$ 最小，分为 $8<x\le 9$ 和 ~~$9<x\le 10$~~ $x>9$

综上，回归树为：



3. 当 ~~a为~~ ∀结点 b 不在 a 到根结点的路径上时（即 b 不是 a 的祖先）

剪枝前后 $g_k(b) = g_{k+1}(b)$ 不变，均大于 α_k，即：$g^{k+1}(b) > \alpha_k$

当 b 为 a 的祖先时：~~落~~

$$|T_b^{k+1}| = |T_b^k| - |T_a^k| + 1$$

$$C(T_b^{k+1}) = C(T_b^k) - C(T_a^k) + C(a)$$

剪枝前后 C(b) 不变

$$g^{k+1}(b) = \frac{C(b) - C(T_b^{k+1})}{|T_b^{k+1}| - 1} = \frac{C(b) - C(T_b^k) - [C(a) - C(T_a^k)]}{[|T_b^k| - 1] - [|T_a^k| - 1]}$$

$$g^k(b) = \frac{C(b) - C(T_b^k)}{|T_b^k| - 1} > \alpha_k \qquad g^k(a) = \frac{C(a) - C(T_a^k)}{|T_a^k| - 1} = \alpha_k$$

故有 $g^{k+1}(b) > \alpha_k$

因为有：$\frac{m}{n} > \frac{x}{y}$，$m, n, x, y > 0$

$$\Rightarrow \frac{m-x}{n-y} = \frac{n \times \frac{m}{n} - y \times \frac{x}{y}}{n-y} > \frac{n \times \frac{m}{n} - y \times \frac{m}{n}}{n-y} = \frac{m}{n} > \frac{x}{y}$$

$$\therefore \alpha_{k+1} = \min_b g^{k+1}(b) > \alpha_k$$