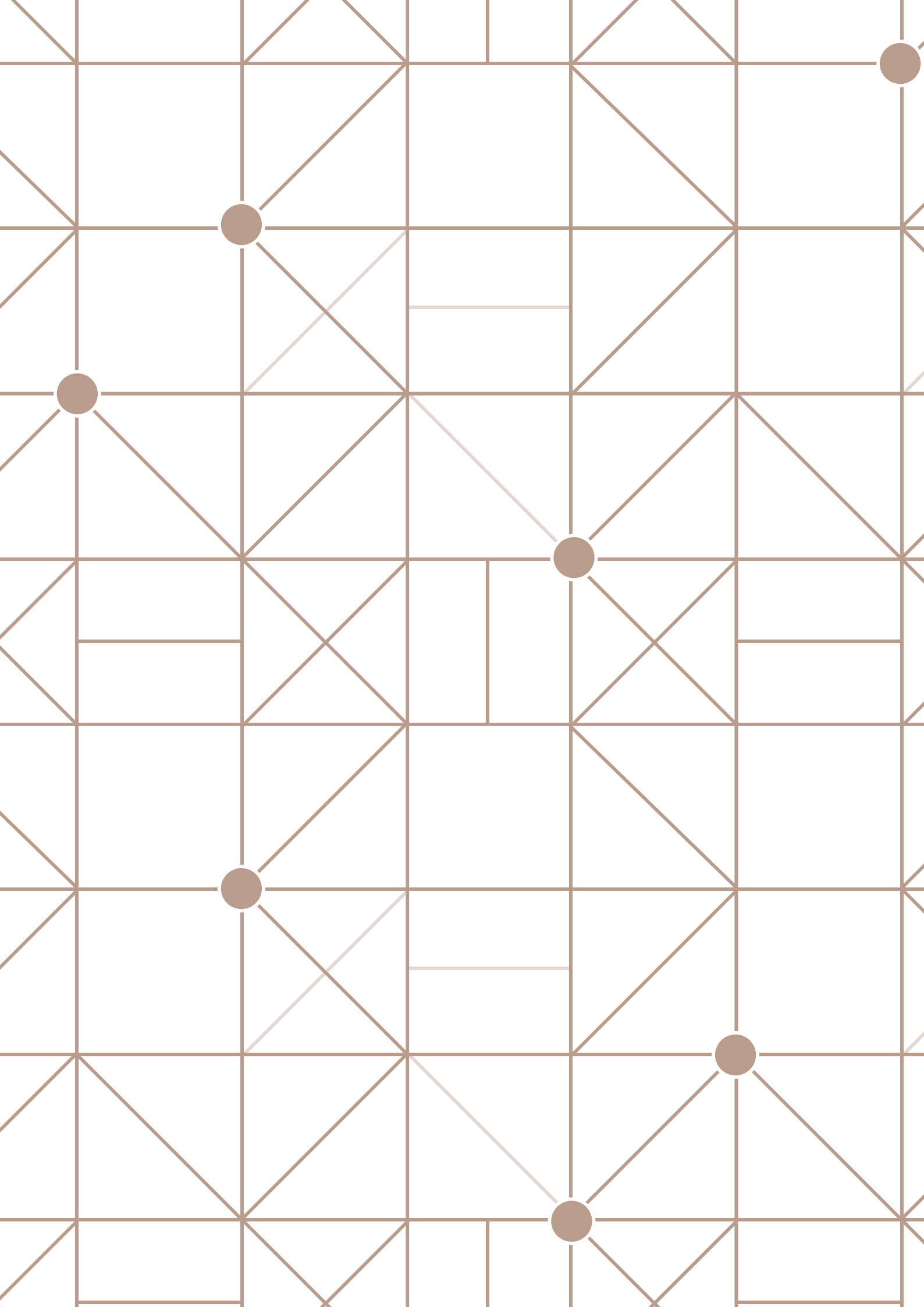

HW 5

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- Class P: buys_computer = "yes"
- Class N: buys_computer = "no"

Yes : 9

No : 5

age	p_i (yes)	n_i (no)	$I(p_i, n_i)$
<=30	2	3	0.971
31…40	4	0	0
>40	3	2	0.971

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$Info(D) = I(9,5) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

$Info_{age}(D) = \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2)$

$= \frac{5}{14}\left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) + \frac{4}{14}\left(-\frac{4}{4}\log_2\left(\frac{4}{4}\right)\right) + \frac{5}{14}\left(-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$

$= 0.694$

Age	Buy_Computer	
	yes	no
<=30	2	3
31..40	4	0
>40	3	2

$Info_{income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1)$

$= \frac{4}{14}\left(-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right) + \frac{6}{14}\left(-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right) + \frac{4}{14}\left(-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right)$

$= 0.911$

income	Buy_Computer	
	yes	no
high	2	2
medium	4	2
low	3	1

$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$

$= \frac{7}{14}\left(-\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right) + \frac{7}{14}\left(-\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right)$

$= 0.789$

student	Buy_Computer	
	yes	no
yes	6	1
no	3	4

$Info_{credit}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$

$= \frac{8}{14}\left(-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right) + \frac{6}{14}\left(-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right)$

$= 0.892$

Credit rating	Buy_Computer	
	yes	no
fair	6	2
excellent	3	3

$$Gain(A) = Info(D) - Info_A(D)$$

$Gain(age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246$

$Gain(income) = Info(D) - Info_{income}(D) = 0.940 - 0.911 = 0.029$

$Gain(student) = Info(D) - Info_{student}(D) = 0.940 - 0.789 = 0.151$

$Gain(credit\_rating) = Info(D) - Info_{credit}(D) = 0.940 - 0.892 = 0.048$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

yes : 2
No : 3

$$Info(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

$$Info_{income}(D) = \frac{2}{5}I(0,2) + \frac{2}{5}I(1,1) + \frac{1}{5}I(1,0)$$

$$= \frac{2}{5}\left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) + \frac{2}{5}\left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) + \frac{1}{5}\left(-\frac{1}{1}\log_2\left(\frac{1}{1}\right)\right)$$

$$= 0.4$$

Income
High Y:0 / N:2
Medium Y:1 / N:1
Low Y:1 / N:0

$$Info_{student}(D) = \frac{2}{5}I(2,0) + \frac{3}{5}I(0,3)$$

$$= \frac{2}{5}\left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) + \frac{3}{5}\left(-\frac{3}{3}\log_2\left(\frac{3}{3}\right)\right)$$

$$= 0$$

Student
Yes Y:2 / N:0
No Y:0 / N:3

$$Info_{credit}(D) = \frac{3}{5}I(1,2) + \frac{2}{5}I(1,1)$$

$$= \frac{3}{5}\left(-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) + \frac{2}{5}\left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$$

$$= 0.951$$

Credit
Fair Y:1 / N:2
Excellent Y:1 / N:1

$Gain(income) = Info(D) - Info_{income}(D) = 0.940 - 0.4 = 0.571$

$Gain(student) = Info(D) - Info_{student}(D) = 0.940 - 0 = 0.971$

$Gain(credit\_rating) = Info(D) - Info_{credit}(D) = 0.940 - 0.951 = 0.02$

age	income	student	credit_rating	buys_computer
31-40	high	no	fair	yes
31-40	low	yes	excellent	yes
31-40	medium	no	excellent	yes
31-40	high	yes	fair	yes

yes : 4
No : 0

Income

High	Y:0 / N:2
Medium	Y:1 / N:1
Low	Y:1 / N:0

Student

Yes	Y:2 / N:0
No	Y:0 / N:3

Credit

Fair	Y:1 / N:2
Excellent	Y:1 / N:1

age	income	student	credit_rating	buys_computer
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

yes : 3
No : 2

$$Info(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

$$Info_{income}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1)$$

$$= \frac{3}{5}\left(-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) + \frac{2}{5}\left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$$

$$= 0.951$$

Income

High	Y:0 / N:0
Medium	Y:2 / N:1
Low	Y:1 / N:1

$$Info_{student}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1)$$

$$= \frac{3}{5}\left(-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) + \frac{2}{5}\left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$$

$$= 0.951$$

Student

Yes	Y:2 / N:1
No	Y:1 / N:1

$$Info_{credit}(D) = \frac{3}{5}I(3,0) + \frac{2}{5}I(0,2)$$

$$= \frac{3}{5}\left(-\frac{3}{3}\log_2\left(\frac{3}{3}\right)\right) + \frac{2}{5}\left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right)$$

$$= 0$$

Credit

Fair	Y:3 / N:0
Excellent	Y:0 / N:2

$Gain(income) = Info(D) - Info_{income}(D) = 0.971 - 0.951 = 0.02$

$Gain(student) = Info(D) - Info_{student}(D) = 0.971 - 0.951 = 0.02$

$Gain(credit\_rating) = Info(D) - Info_{credit}(D) = 0.971 - 0 = 0.971$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no