Hw 2

features — class

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

$$Info_{student}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

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

$$= \frac{7}{14}(0.524 + 0.461) + \frac{7}{14}(0.191 + 0.401)$$

$$= 0.493 + 0.296$$

$$= 0.789$$

$$Info_{credit}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

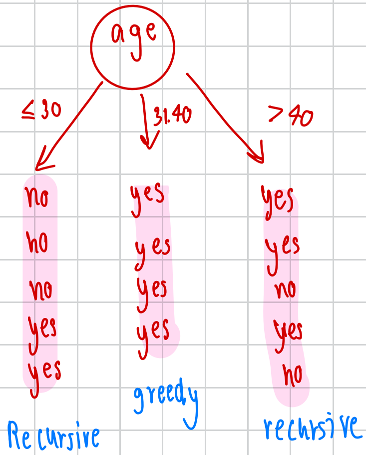
$$= \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$$

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

$$= \frac{8}{14}(0.311 + 0.5) + \frac{6}{14}(0.5 + 0.5)$$

$$= 0.464 + 0.429$$

$$= 0.893$$



643020525-2 หทัยชนก สร้างชัยภูมิ

• คำนวณ class

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

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

(yes, no)

$$= 0.41 + 0.53$$

$$= 0.94$$

• คำนวณ Feature

$$Info_{age}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{0}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right] + \frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$$= \frac{5}{14}(0.529 + 0.442) + \frac{5}{14}(0.442 + 0.529)$$

$$= 0.347 + 0.347$$

$$= 0.69352 \approx 0.694$$

$$Info_{income}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

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

$$= 0.286 + 0.394 + 0.232$$

$$= 0.912$$

** $Gain(A) = Info(D) - Info_A(D)$ มากสุดเป็น root node

$Gain(age) = 0.94 - 0.649 = 0.291$

$Gain(income) = 0.94 - 0.912 = 0.028$

$Gain(student) = 0.94 - 0.789 = 0.151$

$Gain(credit) = 0.94 - 0.893 = 0.047$

6430205225-2 นพชัย ชนก สีรางชัยภูมิ

(Fi) age ≤ 30

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

• คำนวณ class

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

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

$$= 0.5288 + 0.4422$$

$$= 0.971$$

$$Info_{income}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) + -\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] + \left[\frac{1}{5}\left[-1\log_2\left(\frac{1}{5}\right) - 2\log_2(0)\right]\right]$$

$$= \frac{2}{5}(0.5 + 0.5)$$

$$= 0.4$$

$$Info_{student}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

$$= \frac{3}{5}\left[-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\log_2\left(\frac{3}{3}\right)\right] + \frac{2}{5}\left[-\frac{2}{2}\log_2\left(\frac{1}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right]$$

$$= 0$$

$$Info_{credit}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

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

$$= 0.4 + 0.551$$

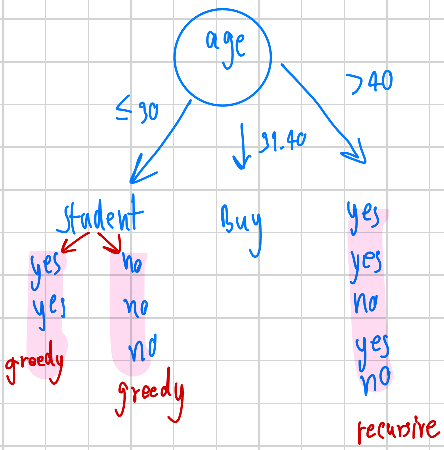
$$= 0.451$$

* ※ $Gain(A) = Info(D) - Info_A(D)$

Gain (income) = 0.971 − 0.4 = 0.571

Gain (student) = 0.971 − 0 = 0.971 → มากสุด

Gain (credit) = 0.971 − 0.951 = 0.020



643020525-2 หทัยชนก สร้างชั

F_2 age >40

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

• ก่อนแบ่ง class

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

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

$$= 0.4422 + 0.5288$$

$$= 0.971$$

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

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

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

$$= 0.551 + 0.4$$

$$= 0.951$$

$$Info_{credit}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

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

$$= \frac{3}{5}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] + \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right]$$

$$= 0$$

$$Info_{student}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D)$$

$$= \frac{3}{5} I(3,0) + \frac{2}{5} I(2,1)$$

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

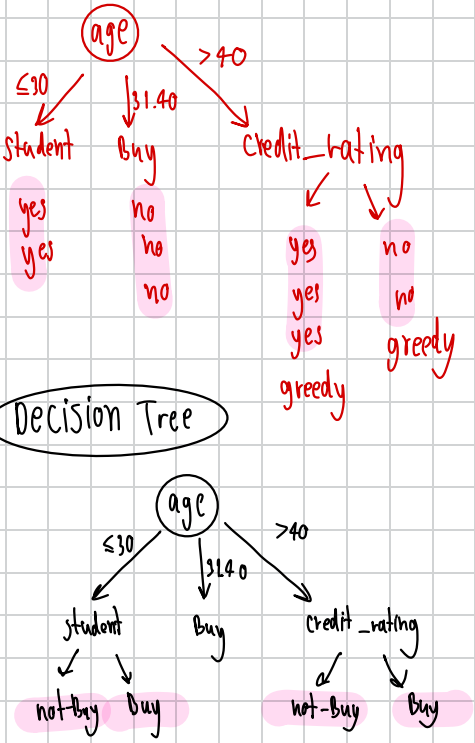
$$= 0.951$$

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

Gain(income) = 0.971 − 0.951 = 0.2

Gain(student) = 0.971 − 0.951 = 0.2

Gain(credit) = 0.971 − 0 = 0.9710 → มากสุด



Decision Tree

6430205525-2 หทัยชนก สีรางชัยภูมิ