

## Class (buys)

$$\text{Info } (0) = - \sum_{i=1}^n p_i \log_2 (cpi)$$

$$\begin{aligned}
 &= I(1, 5) \\
 &= -\left(\frac{1}{14} \log_2 \frac{1}{14}\right) + \left(-\frac{5}{14} \log_2 \frac{5}{14}\right) \\
 &= -\frac{1}{14} (-0.657) - \frac{5}{14} (-1.485) \\
 &= 0.940 \times
 \end{aligned}$$

## Feature

$$\text{Info age } (0) = \sum_{j=1}^k \left| \frac{D_j}{D} \right| \times \text{Info } (D_j) \quad \log_2 \rightarrow 0$$

$$\begin{aligned}
 &= \frac{5}{14} I(2, 3) + \frac{4}{14} I(4, 0) + \frac{5}{14} I(3, 2) \\
 &\leq 30 \quad \begin{matrix} Y \\ N \end{matrix} \quad \begin{matrix} \geq 40 \\ \geq 40 \end{matrix} \quad \begin{matrix} 0 \\ \text{unbalanced} \end{matrix} \\
 &= \frac{5}{14} \left[ \frac{-2}{5} \log_2 \left( \frac{2}{5} \right) - \frac{3}{5} \log_2 \left( \frac{3}{5} \right) \right] + \frac{4}{14} \left[ -\frac{4}{4} \log_2 \left( \frac{0}{4} \right) - \frac{0}{4} \log_2 \left( \frac{0}{4} \right) \right] + \frac{5}{14} \left[ -\frac{3}{5} \log_2 \left( \frac{3}{5} \right) - \frac{2}{5} \log_2 \left( \frac{2}{5} \right) \right] \\
 &> 30 \quad \begin{matrix} Y \\ N \end{matrix} \quad \begin{matrix} 0 \\ \text{unbalanced} \end{matrix} \\
 &= \frac{5}{14} (0.529 + 0.442) + \frac{4}{14} (0 + \text{unbalanced}) + \frac{5}{14} (0.442 + 0.529) \\
 &= \frac{5}{14} (0.971) + \frac{5}{14} (0.971) \\
 &= 0.349 + 0.349 \\
 &\approx 0.694 \times
 \end{aligned}$$

$$\text{Info income } (0) = \sum_{j=1}^k \left| \frac{D_j}{D} \right| \times \text{Info } (D_j)$$

$$\begin{aligned}
 &= \frac{4}{14} I(2, 2) + \frac{6}{14} I(4, 2) + \frac{4}{14} I(3, 1) \\
 &n \quad \begin{matrix} 2 \\ 2 \\ 14 \end{matrix} \quad \begin{matrix} M \\ M \end{matrix} \\
 &= \frac{4}{14} \left[ -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) - \frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right] + \frac{6}{14} \left[ -\frac{4}{6} \log_2 \left( \frac{4}{6} \right) - \frac{2}{6} \log_2 \left( \frac{2}{6} \right) \right] + \frac{4}{14} \left[ -\frac{3}{4} \log_2 \left( \frac{3}{4} \right) - \frac{1}{4} \log_2 \left( \frac{1}{4} \right) \right] \\
 &= \frac{4}{14} (0.5 + 0.5) + \frac{6}{14} (0.390 + 0.528) + \frac{4}{14} (0.31 + 0.5) \\
 &= \frac{4}{14} + \frac{6}{14} (0.918) + \frac{4}{14} (0.811) \\
 &= 0.286 + 0.591 + 0.286 \\
 &\approx 0.912 \times
 \end{aligned}$$

$$\begin{aligned}
 \text{Info student (o)} &= \sum_{j=1}^2 \left| \frac{D_j}{D} \right| \times \text{Info}(D_j) \\
 &= \frac{9}{14} \underbrace{\text{I}(3,4)}_{\text{no}} + \frac{7}{14} \underbrace{\text{I}(b,1)}_{\text{yes}} \\
 &= \frac{9}{14} \left[ -\frac{3}{4} \log_2 \left( \frac{3}{4} \right) - \frac{4}{4} \log_2 \left( \frac{4}{4} \right) \right] + \frac{7}{14} \left[ -\frac{b}{7} \log_2 \left( \frac{b}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right) \right] \\
 &= \frac{9}{14} (0.524 + 0.461) + \frac{7}{14} (0.191 + 0.409) \\
 &= \frac{9}{14} (0.985) + \frac{7}{14} (0.592) \\
 &= 0.493 + 0.296 \\
 &= 0.749 \times
 \end{aligned}$$

$$\begin{aligned}
 \text{Info credit (D)} &= \sum_{j=1}^2 \left| \frac{D_j}{D} \right| \times \text{Info}(D_j) \\
 &= \frac{8}{14} \text{I}(b,2) + \frac{b}{14} \text{I}(3,3) \\
 &= \frac{8}{14} \left[ -\frac{1}{8} \log_2 \left( \frac{1}{8} \right) - \frac{2}{8} \log_2 \left( \frac{2}{8} \right) \right] + \frac{b}{14} \left[ -\frac{3}{b} \log_2 \left( \frac{3}{b} \right) - \frac{3}{b} \log_2 \left( \frac{3}{b} \right) \right] \\
 &= \frac{8}{14} (0.311 + 0.5) + \frac{b}{14} (0.5 + 0.5) \\
 &= \frac{8}{14} (0.811) + \frac{b}{14} \\
 &= 0.464 + 0.429 \\
 &= 0.893
 \end{aligned}$$

### Gain

$$\text{Gain (age)} = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.619 = 0.291$$

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.940 - 0.912 = 0.028$$

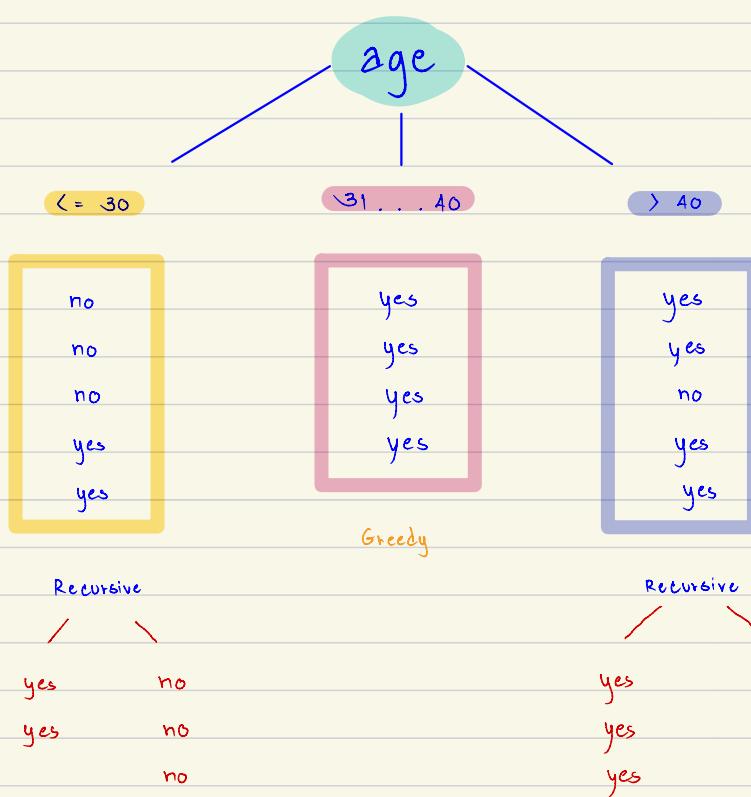
$$\text{Gain (Student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.749 = 0.191$$

$$\text{Gain (credit-rating)} = \text{Info}(D) - \text{Info}_{\text{credit}}(D) = 0.940 - 0.893 = 0.047$$

ពីនេះ Gain (age) មានរយៈព័ត៌មានលម្អិតបន្ថែមជាដុំដាក់

## Training data set: Who buys computer?

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



F1 age L=30

age	income	student	credit	buys
<=30	high	no	fair	no
L=30	high	no	excellent	no
<=30	medium	no	fair	no
L=30	low	yes	fair	yes
L=30	medium	yes	excellent	yes

$$\begin{aligned} \text{Info}(D) &= -\sum_{i=1}^m p_i \log_2(p_i) \\ &= I(2,3) \\ &= -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \\ &= 0.941 \end{aligned}$$

$$\begin{aligned} \text{Info income} &= \frac{2}{5} I(0,2) + \frac{2}{5} I(1,1) + \frac{1}{5} I(1,0) \\ &= \frac{2}{5} \left[ -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \right] + \frac{2}{5} \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] + \frac{1}{5} \left[ -1 \log_2\left(\frac{1}{5}\right) - 2 \log_2(0) \right] \\ &= 0.4 \times \end{aligned}$$

$$\begin{aligned} \text{Info student (D)} &= \frac{3}{5} I(0,3) + \frac{2}{5} I(2,0) \\ &= \frac{3}{5} \left[ -\frac{0}{3} \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \log_2\left(\frac{3}{3}\right) \right] + \frac{2}{5} \left[ -\frac{2}{2} \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2(0) \right] \\ &= 0 \times \end{aligned}$$

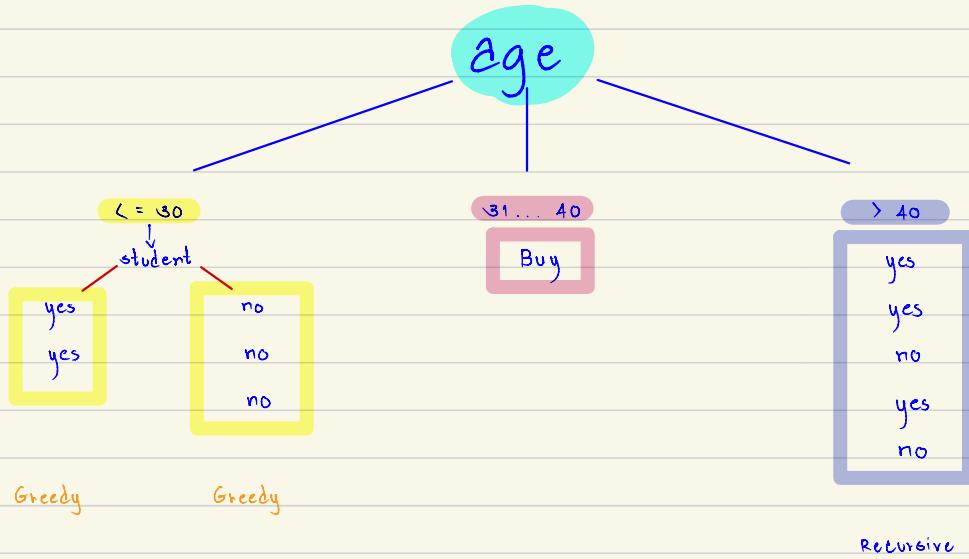
$$\begin{aligned} \text{Info credit (D)} &= \frac{3}{5} I(1,2) + \frac{2}{5} I(1,1) \\ &= \frac{3}{5} \left[ -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] + \frac{2}{5} \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\ &= 0.551 + 0.4 \\ &= 0.951 \times \end{aligned}$$

Gain

$$\text{Gain (income)} = \text{Info}(D) - \text{Info income}(D) = 0.941 - 0.4 = 0.571$$

$$\text{Gain (student)} = \text{Info}(D) - \text{Info student}(D) = 0.941 - 0 = 0.941$$

$$\text{Gain credit} = \text{Info}(D) - \text{Info credit (D)} = 0.941 - 0.951 = 0.020$$



$F_2: \text{age} > 40$

age	income	student	credit	buys
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
> 40	medium	yes	fair	yes
> 40	medium	no	excellent	yes

$$\text{Info}(D) = I(3,2)$$

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

$$= 0.941$$

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

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

$$= 0.651 + 0.1$$

$$= 0.951 \times$$

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

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

$$= 0.4 + 0.551$$

$$= 0.951 \times$$

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

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

$$= 0.4$$

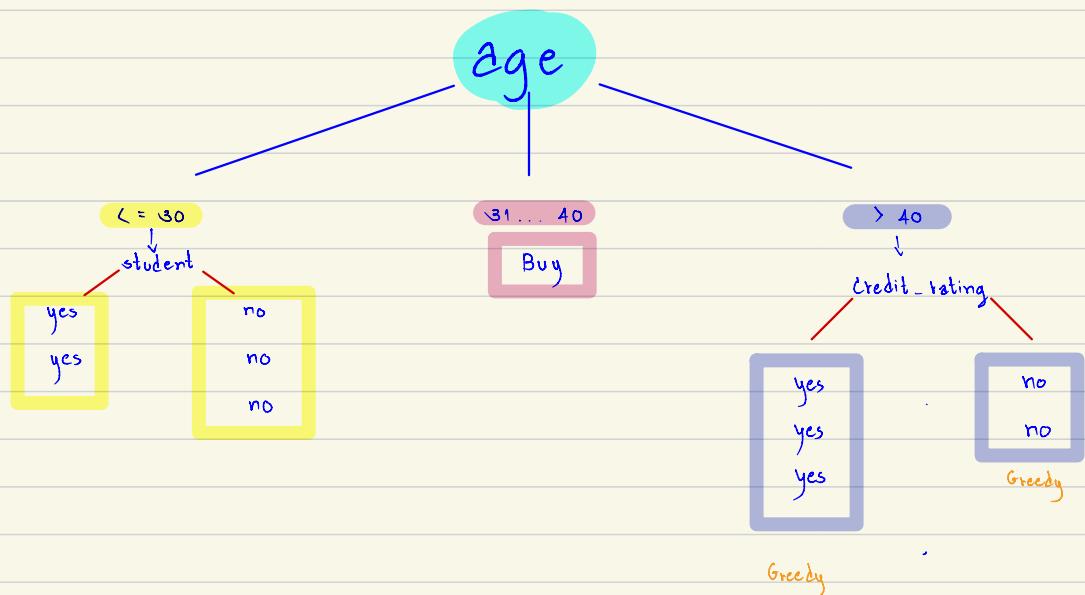
Gain

$$\text{Gain (income)} = \text{Info}(D) - \text{Info income}(D) = 0.971 - 0.951 = 0.2$$

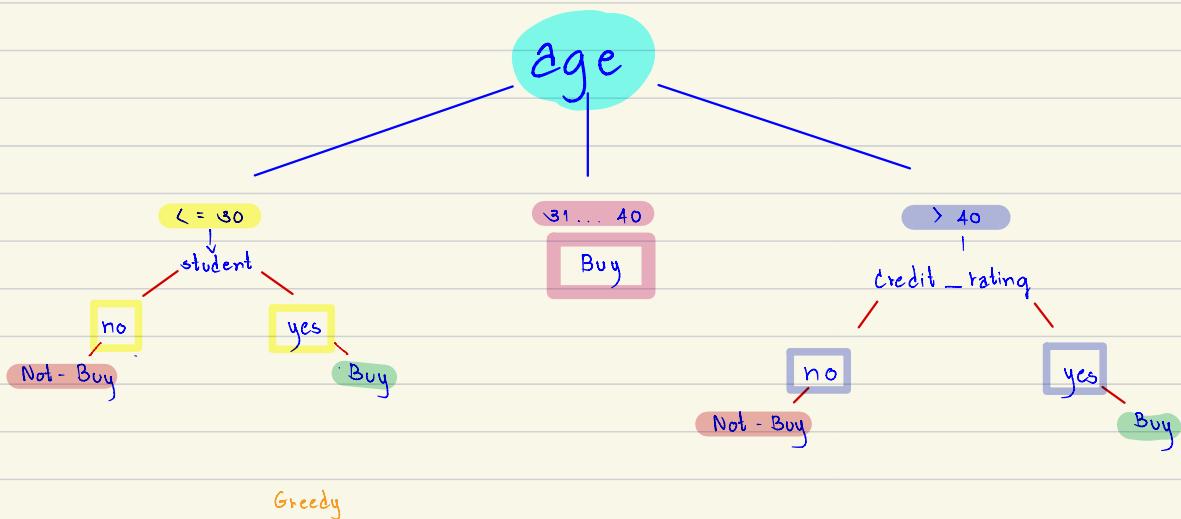
$$\text{Gain (student)} = \text{Info}(D) - \text{Info student}(D) = 0.971 - 0.961 = 0.1$$

$$\text{Gain (credit)} = \text{Info}(D) - \text{Info credit}(D) = 0.971 - 0.4 = 0.571$$

$\therefore$  Gain ~~is increasing~~  $\Rightarrow$  Gain (credit)



### Decision Tree Induction



## ขั้นตอนการแก้ปัญหาด้วยเทคนิค 2 ระยะ (Two Phase Method)

เนื่องจากการแก้ปัญหาโดยเทคนิค Big-M นั้น การคำนวณอาจผิดพลาดได้ง่าย เพราะติดค่า M และในบางครั้ง เมื่อวิเคราะห์แล้วปัญหานั้นมีคำตอบ ดังนั้นเพื่อให้คำนวณง่ายขึ้น และเป็นการตรวจสอบก่อนว่าปัญหานั้นมีคำตอบหรือไม่ จึงมีการใช้เทคนิค 2 ระยะช่วยในการแก้ปัญหาตัวแปรเทียม ดังรายละเอียดต่อไปนี้

**ระยะที่ 1 (phase I)** ทดสอบว่ามีผลลัพธ์ที่เป็นไปได้หรือไม่ (Feasible Solution) ซึ่งเป็นผลลัพธ์ที่ทำให้เงื่อนไขบังคับทุกข้อเป็นจริง โดยใช้สร้างสมการเป้าหมายใหม่เป็นการหาค่าต่ำสุดของผลรวมของตัวแปรเทียม โดยมีข้อจำกัดของปัญหาเดิม และสร้างตาราง Simplex วิเคราะห์ปัญหานี้มา ถ้าตารางสุดท้ายได้ค่าสมการเป้าหมายเป็น 0 แสดงว่าปัญหานั้นมีคำตอบ ให้ทำระยะ 2 ต่อไป ถ้าค่าสมการเป้าหมายในระยะที่ 1 ไม่เป็น 0 แสดงว่าปัญหานั้นไม่มีคำตอบ จึงเรียกผลเฉลยลักษณะนี้ว่า ไม่มีคำตอบที่เป็นไปได้ (infeasible Solution) เพราะไม่สามารถหาตัวแปรมูลฐานใหม่ที่จะเข้าแทนที่ตัวแปรเทียมได้จึงไม่ต้องทำระยะที่ 2 อีก

**ระยะที่ 2 (phase II)** ถ้าค่าสมการเป้าหมายในระยะที่ 1 เป็น 0 จากนั้นจะหาผลลัพธ์ที่ดีที่สุดของปัญหาจากระยะที่ 2 โดยใช้คำตอบมูลฐานที่เหมาะสมที่ได้จากระยะที่ 1 ไปเป็นข้อจำกัดใหม่ในตาราง Simplex เริ่มต้นของระยะที่ 2 โดยตัดคอลัมน์ของ  $A_i$  ทั้งจากตาราง Simplex สุดท้ายที่ได้จากระยะที่ 1 (เพราะข้อจำกัดใหม่ที่ได้นี้ ทำให้ตัวแปรเทียมทุกตัวเป็นศูนย์แล้ว) และนำตารางที่ได้ไปเป็นตาราง Simplex เริ่มต้นของระยะที่ 2 โดยใช้สมการเป้าเดิมของโจทย์ (ค่าสูงสุดหรือค่าต่ำสุด) จากนั้นแก้ปัญหาโดยวิธี Simplex ต่อไปจนกระทั่งได้คำตอบที่เหมาะสมของปัญหานี้

จากตัวอย่างของวิธี Big-M จะใช้วิธี Two phase Method เพื่อหาผลเฉลยที่เหมาะสมที่สุด

### Objective Function

$$\text{Maximize } Z = 4x_1 + x_2 + 2x_3$$

Subject to

$$x_1 + x_2 + x_3 \leq 1$$

$$x_1 + x_2 - x_3 = 0$$

$$x_1, x_2, x_3 \geq 0$$

## Solution

### Phase I

Minimize  $a = A$

Subject to

$$x_1 + x_2 + x_3 \leq 1$$

$$x_1 + x_2 - x_3 = 0$$

$$x_1, x_2, x_3 \geq 0$$

กำหนดให้

$$a - A = 0$$

เป็นสมการที่ 0

$$x_1 + x_2 + x_3 + s_1 = 1$$

เป็นสมการที่ 1

$$x_1 + x_2 - x_3 + A = 0$$

เป็นสมการที่ 2

ตัวแปรมูล ฐาน	สมการ ที่	สมมูล						ค่าคงที่ ขวามือ	อัตราส่วน
		a	X1	X2	X3	S1	A ↓		
a	0	1	0	0	0	0	-1	0	
S1	1	0	1	1	1	1	0	1	null
← A	2	0	1	1	-1	0	1	0	0/1 = 0

ตัวแปรมูล ฐาน	สมการ ที่	สมมูล						ค่าคงที่ ขวามือ	อัตราส่วน
		a	X1 ↓	X2	X3	S1	A		
a	0	1	1	1	-1	0	0	0	
S1	1	0	1	1	1	1	0	1	1/1 = 1

A	2	0	1	-1	0	1	0	$0/1 = 0$
---	---	---	---	----	---	---	---	-----------

ตัวแปรมูลฐาน	สมการที่	สัมประสิทธิ์						ค่าคงที่ความเมื่อ	อัตราส่วน
		a	X1	X2	X3	S1	A		
a	0	1	0	0	0	0	-1	0	
S1	1	0	0	0	2	1	-1	1	
X1	2	0	1	1	-1	0	1	0	

จากตาราง Simplex เริ่มต้นระยะที่ 1 กำหนดให้  $s_1$  และ  $A$  เป็นค่าตอบมูลฐานเริ่มต้น ดังนั้นต้องทำคอลัมน์ของ  $A$  ให้เป็นตัวแปรเข้าแล้ว จากนั้นแก้ปัญหาด้วยวิธี Simplex โดยตารางสุดท้ายของ Simplex ระยะที่ 1 พบว่าเป็นค่าตอบที่ดีที่สุดแล้วของสมการเป้าหมายต่ำสุด การดูที่สัมประสิทธิ์ในแถวที่ศูนย์หน้าตัวแปรไม่มูลฐาน ซึ่งในที่นี้สัมประสิทธิ์ในแถวที่ศูนย์หน้าตัวแปรไม่มูลฐาน มีค่าเป็นลบหมด แสดงว่าไม่สามารถลดค่า  $Z$  ได้อีก ดังนั้นจากผลเฉลยของระยะที่ 1 จะพบว่า  $A=0, a=0$  แสดงว่ามีผลลัพธ์ที่ทำให้ทุกสมการเงื่อนไขเป็นจริง จึงเรียกผลเฉลยลักษณะนี้ว่า มีค่าตอบที่เป็นไปได้ (Feasible Solution) จากนั้นทำระยะที่ 2 ต่อ แต่ถ้ากรณีที่อ่านผลเฉลยแล้วได้  $A > 0$  แสดงว่าไม่มีผลเฉลยที่ทำให้ทุกสมการเงื่อนไขเป็นจริง จึงเรียกผลเฉลยลักษณะนี้ว่า ไม่มีค่าตอบที่เป็นไปได้ (Infeasible Solution)

## Phase II

Objective Function

$$\text{Maximize } Z = 4x_1 + x_2 + 2x_3$$

Subject to

$$x_1 + x_2 + x_3 \leq 1$$

$$x_1 + x_2 - x_3 = 0$$

$$x_1, x_2, x_3 \geq 0$$

ตัวแปรมูลฐาน	สมการที่	สัมประสิทธิ์						ค่าคงที่ขวามือ	อัตราส่วน
		a	X1	X2	X3	S1	A		
a	0	1	0	0	0	0	-1	0	
S1	1	0	0	0	2	1	-1	1	
X1	2	0	1	1	-1	0	1	0	

ตัดคอลัมน์ตัวแปรเทียมทิ้ง

เนื่องจากตารางสุดท้ายของระยะที่ 1 ตัวแปรเทียมไม่ได้เป็นตัวแปรมูลฐานแล้ว จึงไม่จำเป็นต้องอาศัยตัวแปรเทียมในการหาคำตอบของระบบ方程ที่เป็นไปได้อีกต่อไป

จากตารางสุดท้ายของ Simplex ระยะที่ 1 พบร้า  $A = 0$  และ  $a = 0$  แสดงว่าสามารถทำระยะที่ 2 ต่อได้โดยการตัดคอลัมน์ของ  $A$  ทิ้ง 既然นั้นนำตารางสุดท้ายจากระยะที่ 1 ไปเป็นสมการเงื่อนไขใหม่ที่จะถูกใช้เป็นตารางเริ่มต้นของระยะที่ 2 และใส่พังก์ชันเป้าหมายของโจทย์เดิมลงใน列ที่ 0 既然นั้นจะได้ตารางเริ่มต้นของ Simplex ดังแสดงในตารางด้านล่าง หลังจากนั้นจึงแก้ปัญหาของตารางที่นี่ด้วยวิธี Simplex

ตัวแปรมูลฐาน	สมการที่	สัมประสิทธิ์						ค่าคงที่ขวามือ	อัตราส่วน
		Z	X1	X2	X3	S1			
Z	0	1	-4	-1	-2	0	0		
S1	1	0	0	0	2	1	1	Null	
X1	2	0	1	1	-1	0	0	0/1=0	

ตัวแปรมูล ฐาน	สมการ ที่	สมมูลต้องการ					ค่าคงที่ ข้ามมือ	อัตราส่วน
		Z	X1	X2	X3 ↓	S1		
Z	0	1	0	3	-6	0	0	
S1	1	0	0	0	2	1	1	$\frac{1}{2}=1/2$
X1	2	0	1	1	-1	0	0	null

ตัวแปรมูล ฐาน	สมการ ที่	สมมูลต้องการ					ค่าคงที่ ข้ามมือ	อัตราส่วน
		Z	X1	X2	X3	S1		
Z	0	1	0	3	0	3	3	
X3	1	0	0	0	1	1/2	1/2	
X1	2	0	1	1	-1	1/2	1/2	

จากตารางสุดท้ายของชิมเพล็กพบว่า เป็นค่าตอบที่ดีที่สุดแล้ว โดยดูที่สมมูลต้องการที่ศูนย์ของตัวแปรไม่มูลฐาน ซึ่งเจตนาขึ้นนี้สมการเป้าหมายสูงสุด และสมมูลต้องการที่ศูนย์ของตัวแปรไม่มูลฐาน มีค่าเป็นบวกหมด แสดงว่าไม่สามารถเพิ่มค่า Z ได้อีก จึงหยุดทำ แสดงว่าได้ค่าตอบที่ดีที่สุดแล้ว

ดังนั้นผลเฉลยที่ดีที่สุดคือ  $x_1 = 1/2, x_2 = 0, x_3 = 1/2$  และ  $Z = 3$