

MACHINE LEARNING

357353 – Artificial Intelligence
Choopan Rattanapoka

การเรียนรู้ของเครื่องจักร(machine learning)

- เป็นการประยุกต์ของเทคโนโลยีปัญญาประดิษฐ์เพื่อใช้ในการควบคุมเครื่องจักรหรือเครื่องคอมพิวเตอร์
- เพื่อให้เครื่องจักรหรือเครื่องคอมพิวเตอร์สามารถตัดสินใจและแก้ไขปัญหาได้ใกล้เคียงกับมนุษย์
- การเรียนรู้ทั่วไปแบ่งได้ 2 ชนิด
 - * □ **Deductive** เป็นการเรียนรู้โดยอาศัยความรู้ที่เป็นจริงอยู่แล้ว มีความเป็นสากลซึ่งทุกคนยอมรับ
 - * □ **Inductive** เป็นการเรียนรู้จากเหตุการณ์หรือสิ่งที่สนใจ โดยทราบข้อมูลเพียงบางส่วน

การเรียนรู้แบบ Inductive

- เป็นการเรียนรู้จากข้อมูลเพียงบางส่วนก่อนจะดำเนินการหาความจริงที่เป็น สากล
- การเรียนรู้แบบ **Inductive** สามารถแบ่งออกเป็น 2 ประเภท

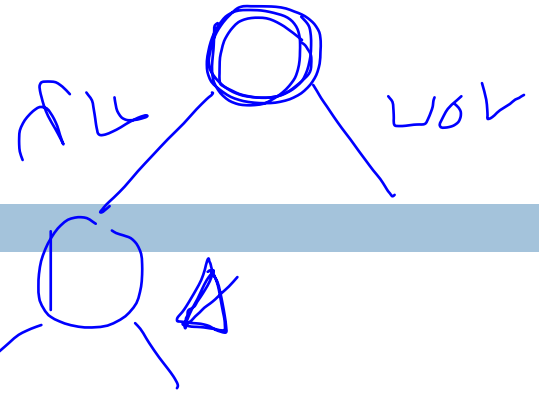
▣ Supervised Learning

- เป็นการเรียนรู้ที่สามารถนำเสนอและจำแนกข้อมูลภายในชุดข้อมูลว่ามีผลลัพธ์ที่ถูกหรือผิดได้
- ข้อมูลจะถูกนำไปใช้ประมาณค่าหรือพยากรณ์ค่าข้อมูล
- เทคนิคที่ใช้วิธีนี้ เช่น **Decision Tree, Neural network**

▣ Unsupervised Learning

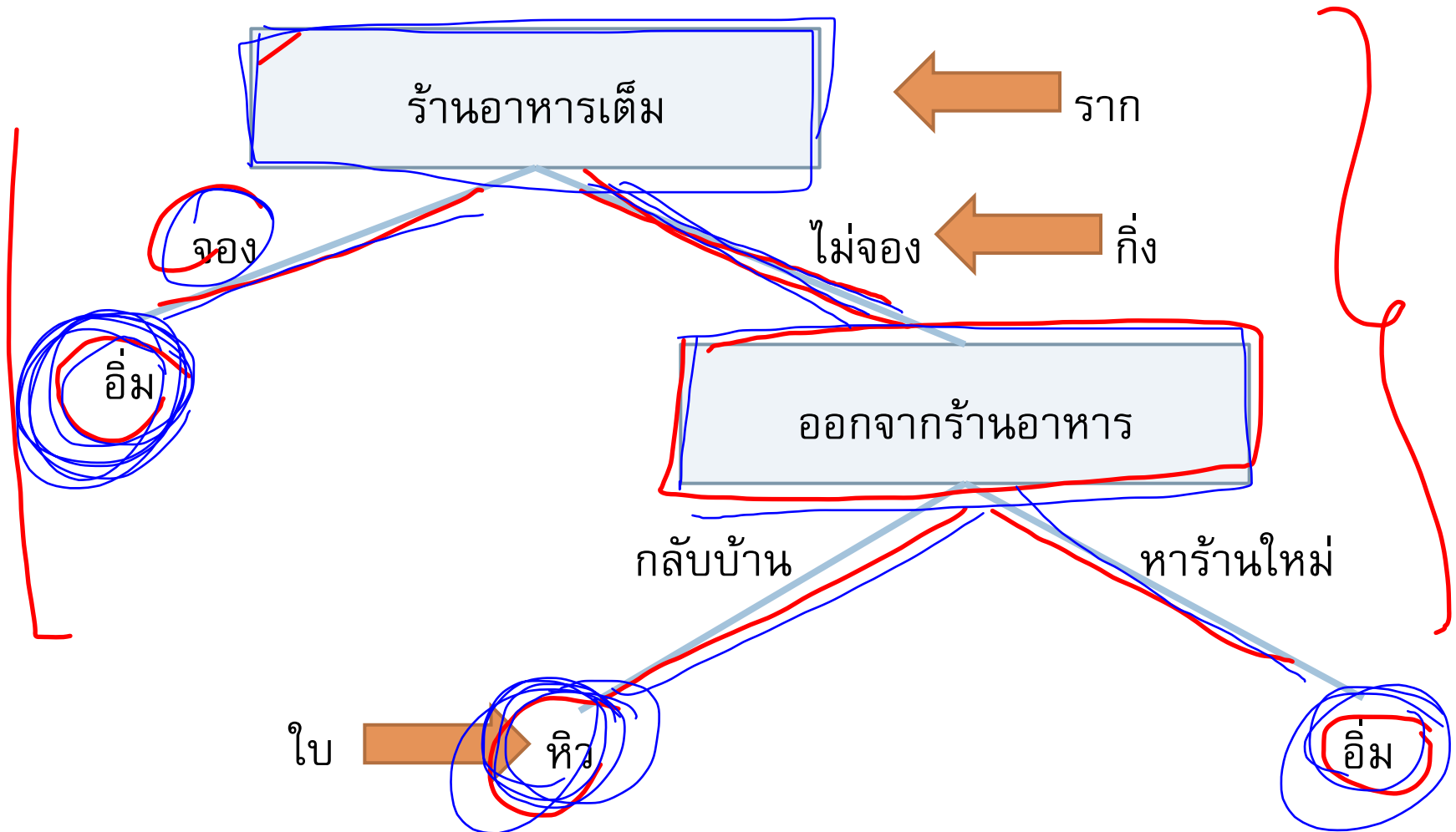
- เป็นการเรียนรู้ที่ไม่มีการกำหนดข้อมูลที่สนใจภายในชุดข้อมูล ไม่มีการจำแนกข้อมูลว่า ผลลัพธ์จะเป็นเช่นไร
- เทคนิคที่ใช้วิธีนี้ เช่น Nearest Neighbor Classification //

Decision Tree



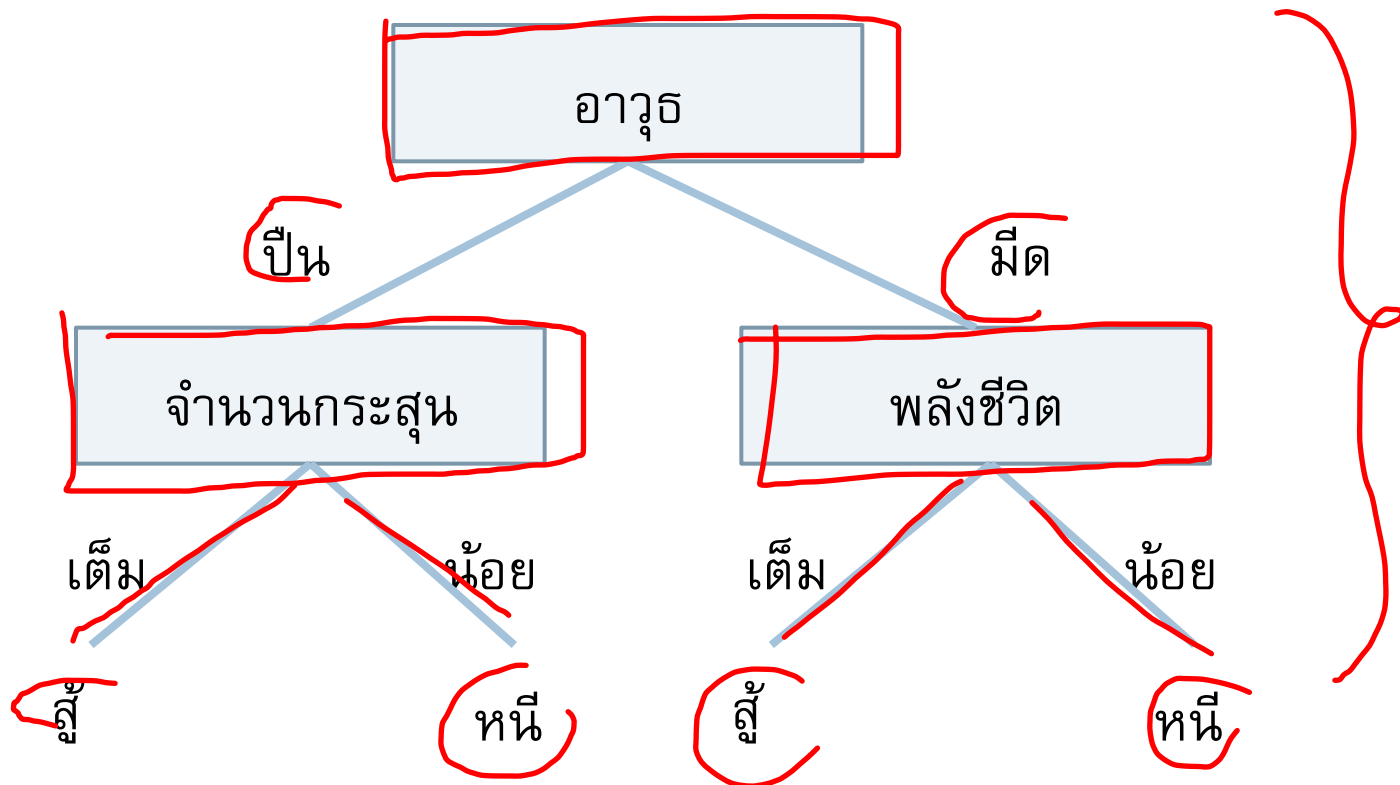
- เรียกอีกชื่อว่า **Classification tree**
- เป็นเทคนิคแบบ **Supervised Learning**
- เป็นวิธีเรียนรู้ที่ไม่ซับซ้อนมากนัก โดยมีโครงสร้างเป็นต้นไม้ แตกแขนงไปตามเงื่อนไขหรือข้อมูลที่ได้คาดคะเนไว้ว่าจะเกิดขึ้น
- โครงสร้างของ **Decision tree** จะประกอบด้วย
 - ใบ (Leaf) เป็นส่วนของข้อมูลที่เราสนใจ ซึ่งอาจเป็นข้อมูลที่เกิดขึ้นโดยสิ่งแวดล้อมตามสถานการณ์นั้น หรือเป็นสิ่งที่กำหนดตามการคาดคะเนว่ามีโอกาสที่จะเกิดขึ้น
 - กิ่ง (Branch) เป็นตัวเชื่อมระหว่างโหนด
 - ราก (Root) เป็นจุดเริ่มต้นของเหตุการณ์

ตัวอย่าง : Decision Tree



การเรียนรู้ด้วย Decision Tree

- การเรียนรู้ด้วย **Decision tree** เป็นการเรียนรู้จากการคาดคะเนเหตุการณ์ต่างๆ ที่อาจเกิดขึ้น
- อาศัยเงื่อนไขเป็นตัวช่วยในการตัดสินใจว่าเมื่อมีเหตุการณ์ใด ๆ ขึ้น จะแสดงผลออกมาอย่างไร
- การคาดคะเนจะถูกนำเสนอด้วยรูปแบบการตัดสินใจ (if/then/else)



การสร้าง Decision Tree

- ขั้นตอนที่สำคัญในเทคนิคการเรียนรู้แบบ **Decision Tree** คือการสร้างแบบจำลอง **Decision Tree**
- การสร้าง **Decision Tree** จะต้องเริ่มพิจารณาที่โหนด "ราก" เป็นอันดับแรก ก่อนจะดำเนินการพิจารณาใบ และกิ่งก้าน
- การหาโหนดรากจะใช้วิธีการคำนวณหาค่า "**Information gain**" ที่มากที่สุด
- **Information gain** จะประกอบไปด้วยข้อมูลอีกชนิดหนึ่งเรียกว่า **Entropy**

Entropy และ Information Gain

□ สมการหาค่า Entropy

$$E(S) = \sum_{i=1}^n -P(V_i) \log_2 P(V_i)$$

1

□ $E(S)$ = ค่า Entropy ของเซต(S) ข้อมูลทั้งหมด

* □ $S = P(V_1), P(V_2), \dots, P(V_n)$

□ $P(V_i)$ = ค่าความน่าจะเป็นของข้อมูลที่น่าสนใจ

□ สมการหาค่า Information Gain

$$\underline{Gain(S,A)} = \underline{E(S)} - \sum_{v \in Value(A)} \frac{|S_v|}{S} E(S_v)$$

2

□ $Gain(S,A)$ = ค่า Gain ของเหตุการณ์ที่น่าสนใจ

□ $E(S_v)$ = ค่า Entropy ของเซตข้อมูลที่น่าสนใจ

□ S_v = เซตข้อมูลที่น่าสนใจ

Information Gain

- จากสมการของ **Information Gain** ข้อมูลกลุ่มใดมีค่า **Information Gain** มากที่สุดจะนำมาใช้เป็นโหนดราก
- เมื่อได้โหนดรากแล้วก็นำข้อมูลที่เหลือมาคำนวณ **Information Gain** อีกครั้งเพื่อหาโหนดถัดไปที่เหมาะสม

$$\text{gain}(\text{order}) = E(S) - E \frac{|S_v|}{S} E(S_v)$$

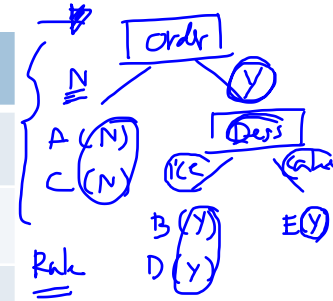
ตัวอย่าง 1

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

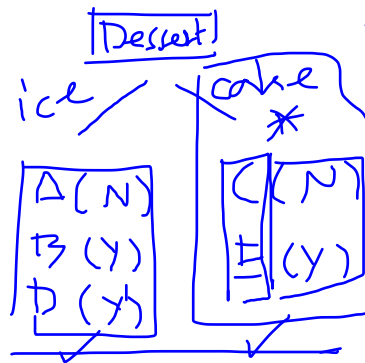
0.0199

- มีข้อมูลดิบที่ได้จากการสำรวจในร้านอาหารแห่งหนึ่งว่าลูกค้า 5 คนมีการสั่งของหวานหลังอาหารมือเย็นแตกต่างกันไปตั้งข้อมูลในตาราง

ลูกค้า	การสั่งของหวาน	ชนิดของหวาน	ความชอบ
A	ไม่สั่ง	ไอศกรีม	ไม่ชอบ
B	สั่ง	ไอศกรีม	ชอบ
C	ไม่สั่ง	ขนมเค้ก	ไม่ชอบ
D	สั่ง	ไอศกรีม	ชอบ
E	สั่ง	ขนมเค้ก	ชอบ



- การสั่งของหวาน (Order)
- ชนิดของหวาน (Dessert)



$$E(S) = \sum_{i=1}^N -p(x_i) \log_2 p(x_i)$$

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

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

ตัวอย่าง 2

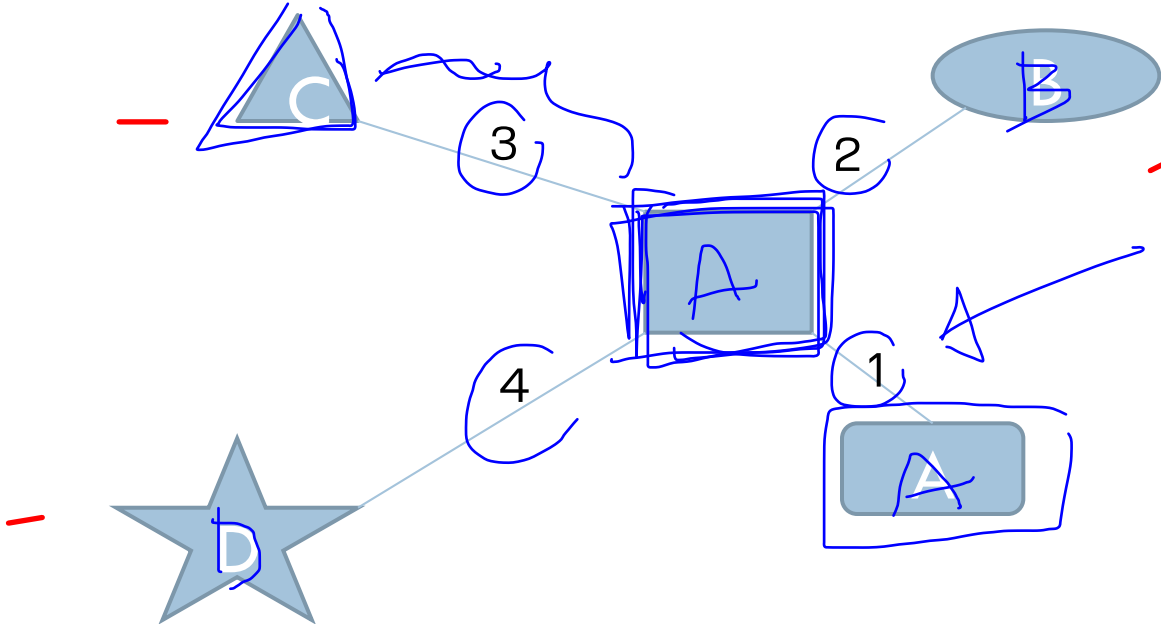
decision

□ ข้อมูลดิบที่ได้จากการสำรวจความมอดทนของลูกค้าที่รอใช้บริการ

ลูกค้า	สถานะของโต๊ะ	ราคา	ร้านอาหาร	ระยะเวลา รอ (นาที)	ลูกค้ารอ
A	เต็ม	แพง	ญี่ปุ่น	10-30	รอ
B	เต็ม	ถูก	ไทย	10-30	รอ
C	เต็ม	แพง	ไทย	10-30	ไม่รอ
D	เต็ม	แพง	ญี่ปุ่น	มากกว่า 30	ไม่รอ
E	เต็ม	แพง	อิตาลี	0-10	รอ
F	ว่าง	ถูก	ไทย	10-30	รอ
G	เต็ม	แพง	อิตาลี	10-30	ไม่รอ
H	ว่าง	แพง	ไทย	0-10	รอ

Nearest Neighbor Classification

- เป็นการเรียนรู้แบบ **Unsupervised Learning** ซึ่งเป็นการจำแนกหรือจัดกลุ่มที่มีวิธีการไม่ซับซ้อนมากนัก
- โดยจะพิจารณาจากชุดข้อมูลใกล้เคียงกับค่าของข้อมูลที่พิจารณามากที่สุด



การหาระยะทาง

- ในการหาระยะทางจะใช้ทฤษฎีการวัดค่าของ **Euclidean**

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

□ d = ระยะทาง

□ p_i = ค่าที่อยู่ในข้อมูลที่ต้องการจำแนก

□ q_i = ค่าของข้อมูลข้างเคียง

ตัวอย่าง 3

$$KV = \sqrt{(0-0)^2 + (1-1)^2 + (0)^2 + 0 + 0 + 0 + (0-1)^2 + (1-1)^2} = \sqrt{2}$$

กำหนดให้มีการจำแนกสัตว์ 5 ชนิด จำแนกจากคุณสมบัติ (1NN)

$\sqrt{1} = 1$

	ออกลูกเป็นตัว	ออกลูกเป็นไข่	เลี้ยงลูกด้วยนม	มีขนตามร่างกาย	ผิวหนังหรือมีเกล็ด	สัตว์เลือดอุ่น	สัตว์เลือดเย็น	อยู่ทั้งบกและน้ำ	หายใจด้วยเหงือก	
กบ	0	1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
เปิด		1		1		1		1		สัตว์จำพวกนก
ค่างคาว	1		1	1						สัตว์เลี้ยงลูกด้วยนม
งู		1			1		1		1	สัตว์เลื้อยคลาน
ปลาแซลมอล		1			1		1		1	สัตว์จำพวกปลา
สัตว์ปริศนา	0	1						1		???

$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$

$\sqrt{0} = 0$

ตัวอย่าง 4

กำหนดให้มีการจำแนกสัตว์ 5 ชนิด จำแนกจากคุณสมบัติ (3NN)

	ออกลูกเป็นตัว	ออกลูกเป็นไข่	เลี้ยงลูกด้วยนม	มีขนตามร่างกาย	ผิวหนังบางหรือมีเกล็ด	สัตว์เลือดอุ่น	สัตว์เลือดเย็น	อยู่ทั้งบกและน้ำ	หายใจด้วยเหงือก	
กบ		1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
เปิด		1		1		1		1		สัตว์จำพวกนก
ค่างคาว	1		1	1						สัตว์เลี้ยงลูกด้วยนม
งู		1			1		1		1	สัตว์เลื้อยคลาน
ปลาแซลมอล		1			1		1		1	สัตว์จำพวกปลา
จระเข้		1			1		1			สัตว์เลื้อยคลาน
แมว	1		1	1		1				สัตว์เลี้ยงลูกด้วยนม
นกกระจอกเทศ		1		1		1				สัตว์จำพวกนก
สัตว์ปริศนา			1	1		1				???

แบบฝึกหัด 1

- ซูเปอร์มาร์เก็ตแห่งหนึ่งต้องการสำรวจความต้องการซื้อสินค้าชนิดหนึ่ง ซึ่งมีทั้งหมด 3 ยี่ห้อคือ **A, B** และ **C** โดยพิจารณาข้อมูลที่สำคัญ 3 อย่างคือ ยี่ห้อ, ของแถม, และราคา จงเขียน **Decision tree** โดยให้ความสนใจกับการตัดสินใจที่จะซื้อสินค้า

ลูกค้า	ยี่ห้อ	ของแถม	ราคา	ความต้องการที่จะซื้อสินค้า
สมชาย	A	มี	ถูก	ซื้อ
สมศรี	B	มี	แพง	ไม่ซื้อ
สมปอง	A	ไม่มี	ถูก	ซื้อ
สมศักดิ์	C	มี	แพง	ซื้อ
สมหมาย	C	ไม่มี	ถูก	ไม่ซื้อ
สมหญิง	B	มี	ถูก	ซื้อ
สมใจ	C	ไม่มี	แพง	ไม่ซื้อ
สมจิตร	A	ไม่มี	แพง	ซื้อ

แบบฝึกหัด 2

	ออกธูปเป็นตัว	ออกธูปเป็นไข่	เลี้ยงลูกด้วยนม	สืบตามร่างกาย	ผิวหนังหรือมีเกล็ด	สัตว์เลือดอุ่น	สัตว์เลือดเย็น	อยู่ทั้งบกและน้ำ	หายใจด้วยเหงือก	
กบ		1					1	1		สัตว์ครึ่งบกครึ่งน้ำ
ปลาหมอ		1			1		1		1	สัตว์จำพวกปลา
จระเข้		1			1		1			สัตว์เลื้อยคลาน
ปลาทอง		1			1		1		1	สัตว์จำพวกปลา
เต่า		1					1			สัตว์เลื้อยคลาน
สุนัข	1		1	1		1				สัตว์เลี้ยงลูกด้วยนม
นกกระจอกเทศ		1		1		1				สัตว์จำพวกนก
สัตว์ปริศนา (1NN)	1			1						???
สัตว์ปริศนา (3NN)		1			1		1			???