

Classifier Evaluation Metrics

Content

- Confusion Matrix
 - True Positive (TP), True Negative (TN)
 - False Positive (FP), False Negative (FN)
- Precision and Recall
- F-Measure
- ROC Curve and AUC

Confusion Matrix

ID	Actual	Predicted
1	spam	spam
2	spam	spam
3	normal	normal
4	normal	spam
5	spam	spam
6	spam	spam
7	normal	spam
8	spam	normal
9	normal	normal
10	normal	normal
11	spam	spam
12	spam	spam
13	spam	normal
14	spam	normal
15	normal	normal

<div> <div>Actual</div> <div>Predicted</div> </div>	normal	spam
	True Positive	False Positive
normal	True Positive	False Positive
spam	False Negative	True Negative

True Positive

ID	Actual	Predicted
1	spam	spam
2	spam	spam
3	normal	normal
4	normal	spam
5	spam	spam
6	spam	spam
7	normal	spam
8	spam	normal
9	normal	normal
10	normal	normal
11	spam	spam
12	spam	spam
13	spam	normal
14	spam	normal
15	normal	normal

Actual Predicted	normal	spam
normal	4	False Positive
spam	False Negative	True Negative

True Negative

ID	Actual	Predicted
1	spam	spam
2	spam	spam
3	normal	normal
4	normal	spam
5	spam	spam
6	spam	spam
7	normal	spam
8	spam	normal
9	normal	normal
10	normal	normal
11	spam	spam
12	spam	spam
13	spam	normal
14	spam	normal
15	normal	normal

Actual Predicted	normal	spam
	4	False Positive
normal	4	False Positive
spam	False Negative	6

False Positive

ID	Actual	Predicted
1	spam	spam
2	spam	spam
3	normal	normal
4	normal	spam
5	spam	spam
6	spam	spam
7	normal	spam
8	spam	normal
9	normal	normal
10	normal	normal
11	spam	spam
12	spam	spam
13	spam	normal
14	spam	normal
15	normal	normal

Actual Predicted	normal	spam
normal	4	3
spam	False Negative	6

False Negative

ID	Actual	Predicted
1	spam	spam
2	spam	spam
3	normal	normal
4	normal	spam
5	spam	spam
6	spam	spam
7	normal	spam
8	spam	normal
9	normal	normal
10	normal	normal
11	spam	spam
12	spam	spam
13	spam	normal
14	spam	normal
15	normal	normal

Actual Predicted	normal	spam
normal	4	3
spam	2	6

Precision ของ normal

Predicted \ Actual	normal	spam
	normal	spam
normal	4 (True Positive)	3 (False Positive)
spam	2 (False Negative)	6 (True Negative)

$$\text{Precision ของ normal} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Precision ของ normal} = \frac{4}{4 + 3} = \frac{4}{7}$$

Precision ของ spam

Predicted \ Actual	normal	spam
	normal	spam
normal	4 (True Positive)	3 (False Positive)
spam	2 (False Negative)	6 (True Negative)

$$\text{Precision ของ spam} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}}$$

$$\text{Precision ของ spam} = \frac{6}{2 + 6} = \frac{6}{8}$$

Recall ของ normal

Predicted \ Actual	normal	spam
	normal	spam
normal	4 (True Positive)	3 (False Positive)
spam	2 (False Negative)	6 (True Negative)

$$\text{Recall ของ normal} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Recall ของ normal} = \frac{4}{4 + 2} = \frac{4}{6}$$

Recall ของ spam

Predicted \ Actual	normal	spam
normal	4 (True Positive)	3 (False Positive)
spam	2 (False Negative)	6 (True Negative)

$$\text{Recall ของ spam} = \frac{\text{True Negative}}{\text{False Negative} + \text{True Negative}}$$

$$\text{Recall ของ spam} = \frac{6}{3 + 6} = \frac{6}{9}$$

F-Measure ของ normal

- F-Measure คือ ค่าเฉลี่ยของ Precision และ Recall

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Precision \text{ ของ } normal = \frac{4}{7} = 0.57 \quad Recall \text{ ของ } normal = \frac{4}{6} = 0.67$$

$$F - Measure \text{ ของ } normal = \frac{2 \times 0.57 \times 0.67}{0.57 + 0.67} = 0.62$$

F-Measure ของ spam

- F-Measure คือ ค่าเฉลี่ยของ Precision และ Recall

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Precision \text{ ของ spam} = \frac{6}{8} = 0.75 \quad Recall \text{ ของ spam} = \frac{6}{9} = 0.67$$

$$F - Measure \text{ ของ normal} = \frac{2 \times 0.75 \times 0.67}{0.75 + 0.67} = 0.71$$

Accuracy

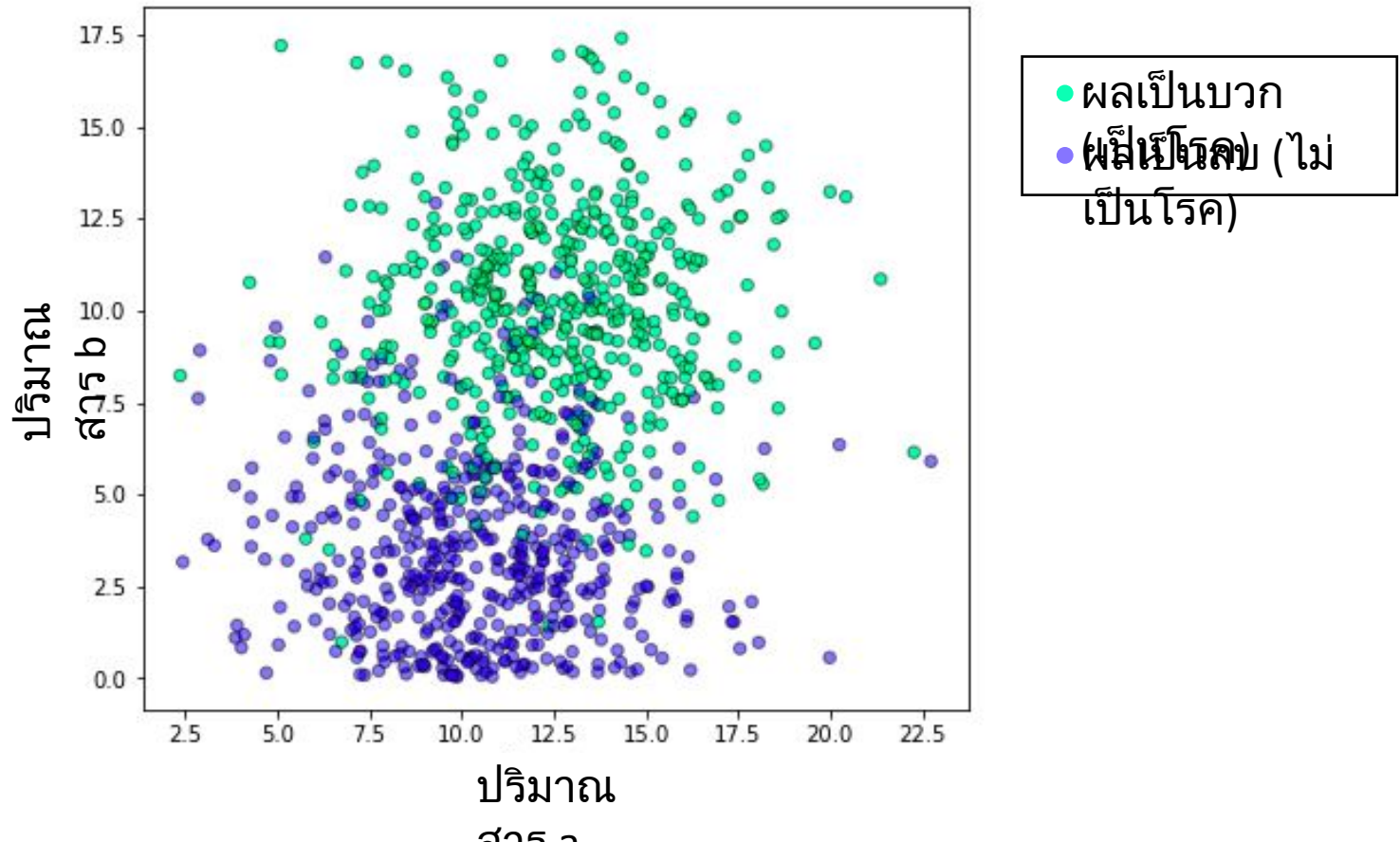
- ความแม่นยำของโมเดล

Predicted \ Actual	normal	spam
	normal	spam
normal	True Positive	False Positive
spam	False Negative	True Negative

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

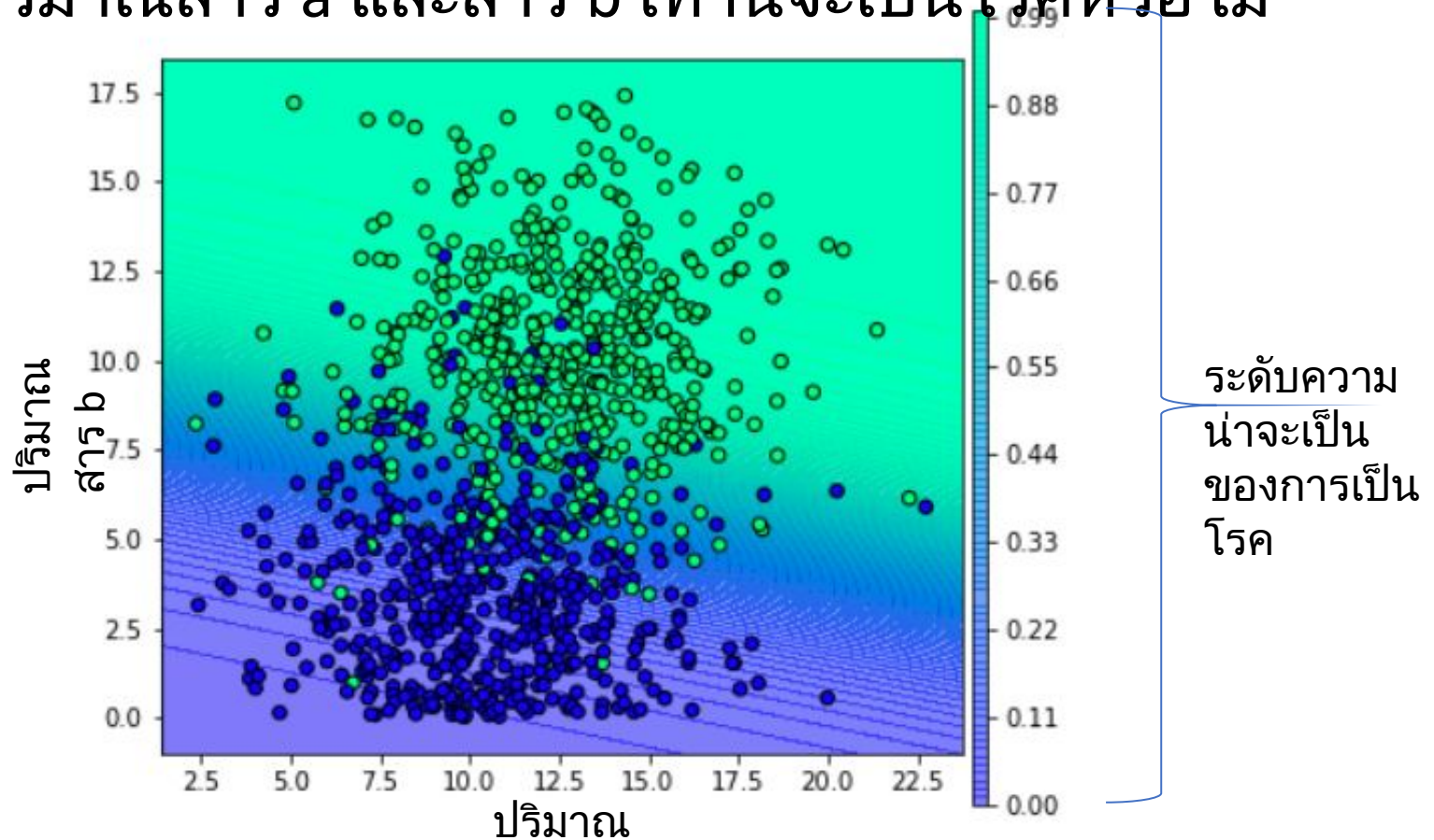
ROC Curve and AUC

- สมมติว่า มีข้อมูลการตรวจโรคจากสาร 2 ชนิดคือ สาร a และสาร b จำนวน 1000 คน แสดงผลได้ดังนี้



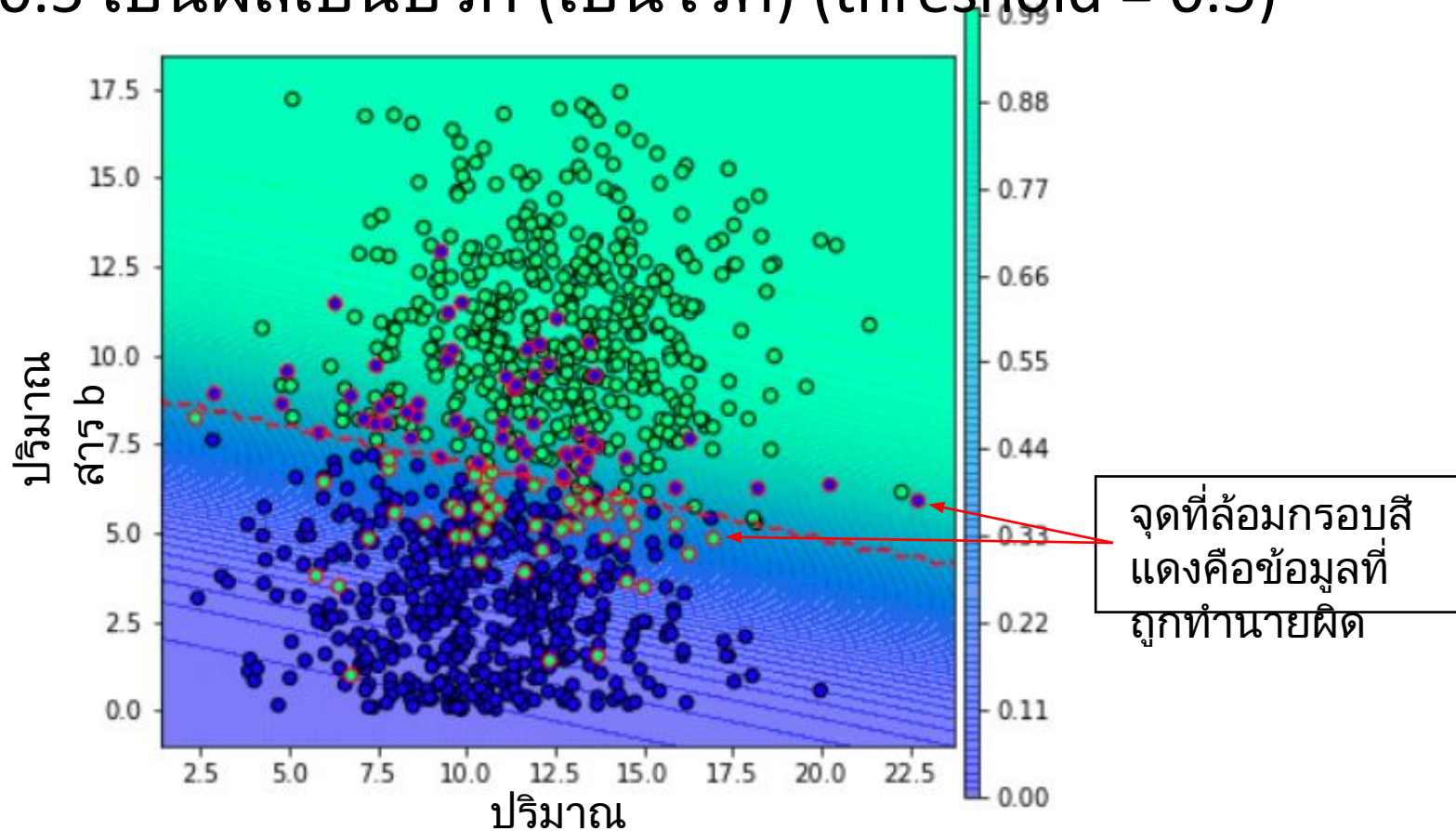
ROC Curve and AUC

- สร้างโมเดล ML ขึ้นมาตัวหนึ่งเพื่อทำนายว่าถ้าตรวจพบปริมาณสาร a และสาร b เถ้านี้จะเป็นโรคหรือไม่



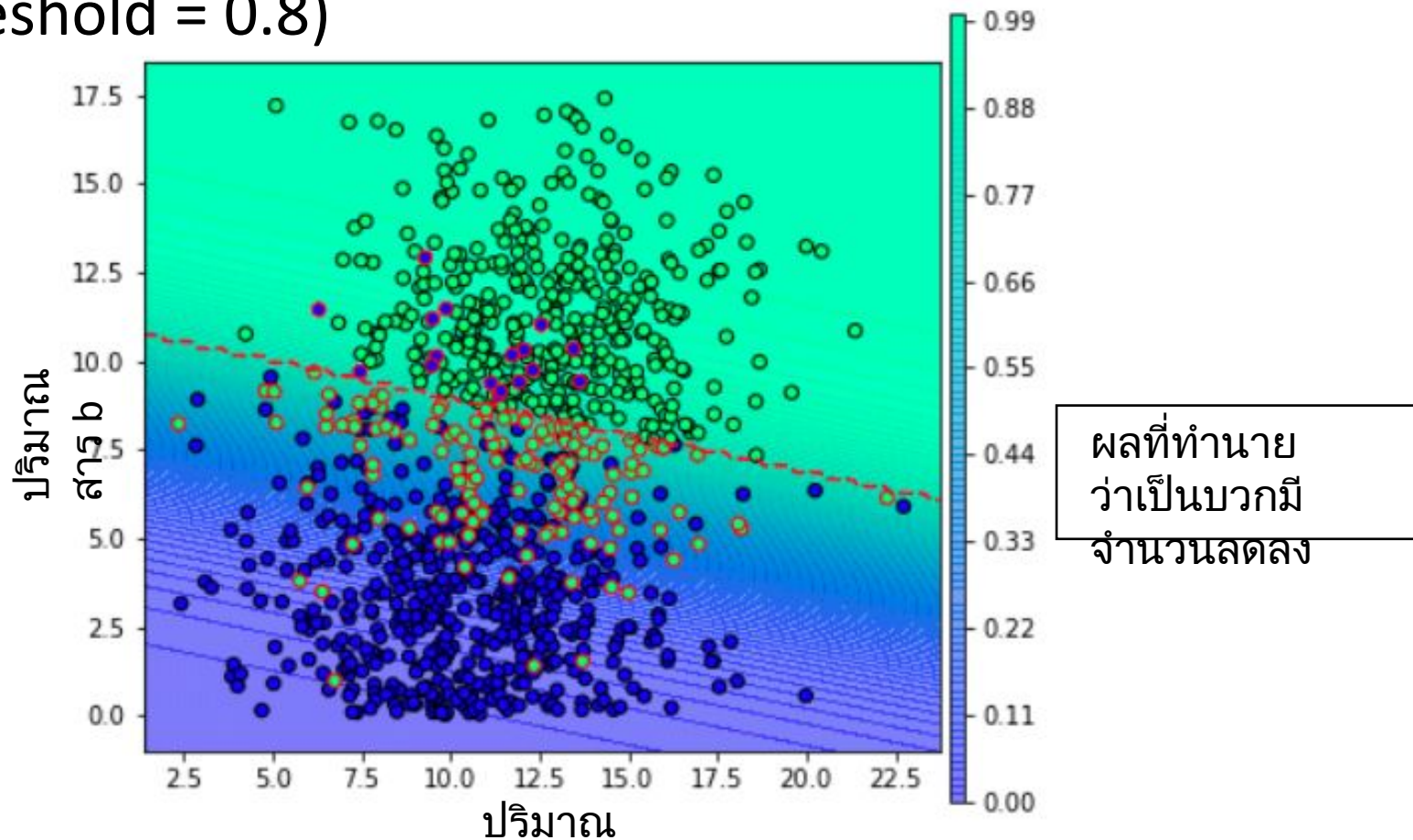
ROC Curve and AUC

- โดยทั่วไป มักตัดสินใจให้ผลที่มีค่าความน่าจะเป็นมากกว่า 0.5 เป็นผลเป็นบวก (เป็นโรค) (threshold = 0.5)



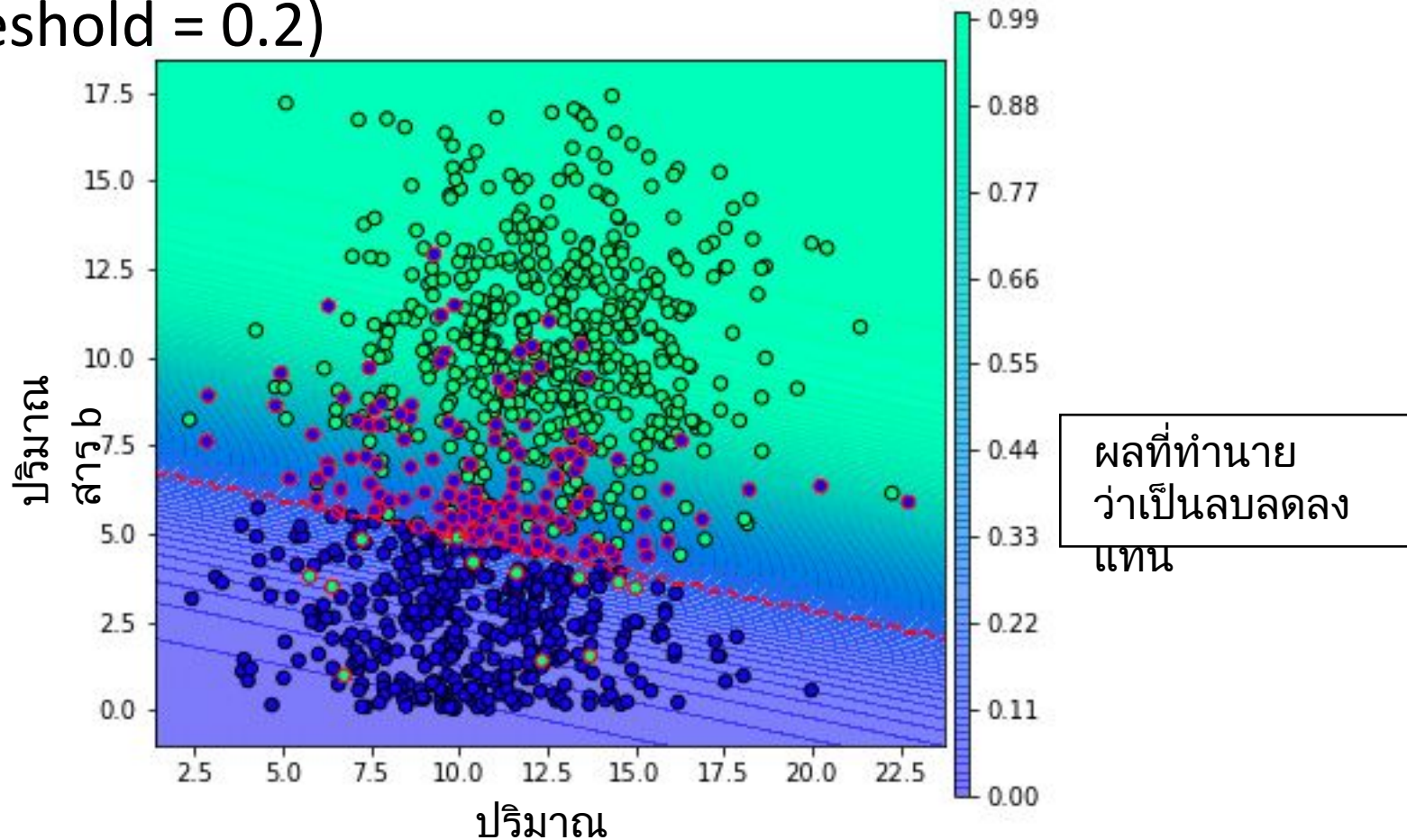
ROC Curve and AUC

- ถ้าปรับค่าความน่าจะเป็นที่ใช้ตัดสินเป็น 0.8 (threshold = 0.8)



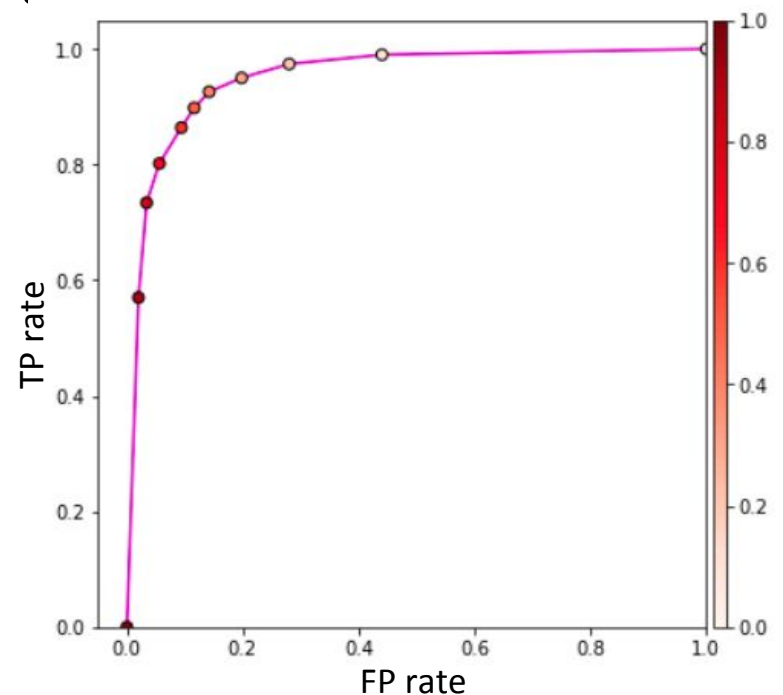
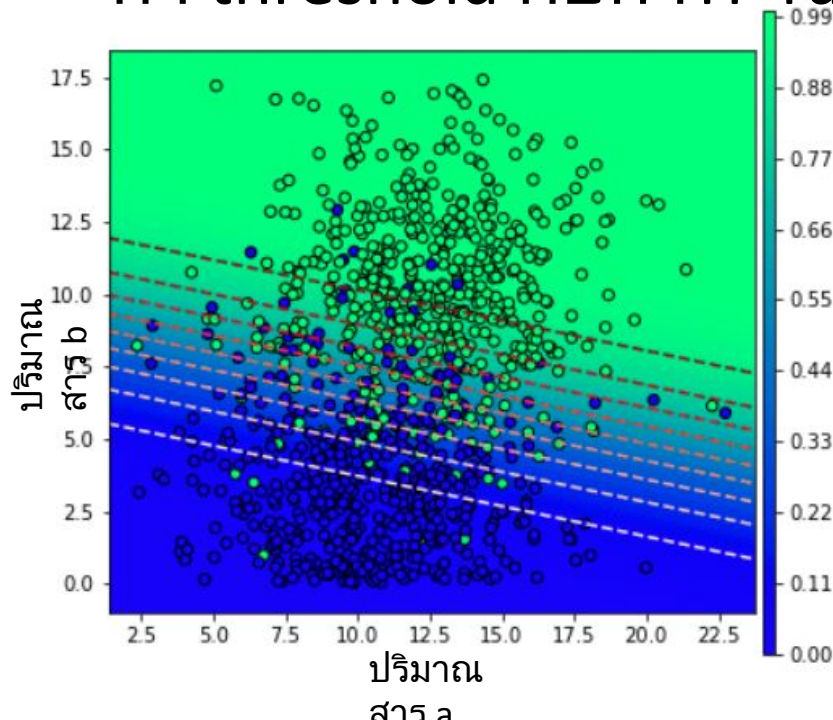
ROC Curve and AUC

- แต่ถ้าปรับค่าความน่าจะเป็นที่ใช้ตัดสินเป็น 0.2 (threshold = 0.2)



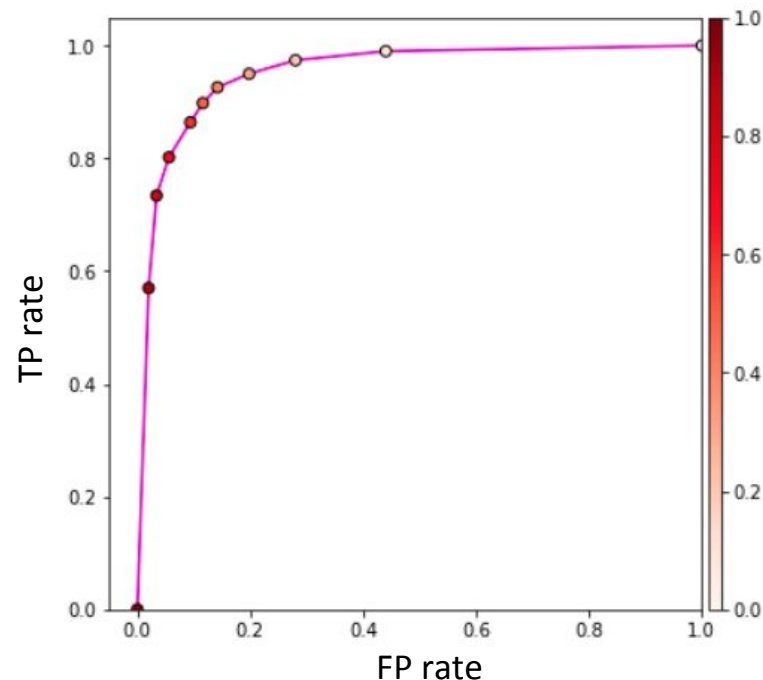
ROC Curve and AUC

- การปรับค่าความน่าจะเป็นที่ใช้ตัดสินหรือ threshold ทำให้ TP rate และ FP rate เปลี่ยนแปลง
- ถ้านำข้อมูลนี้มาวาดกราฟแสดงความสัมพันธ์ระหว่างค่า threshold กับค่า TP rate และ FP rate จะได้ดังนี้

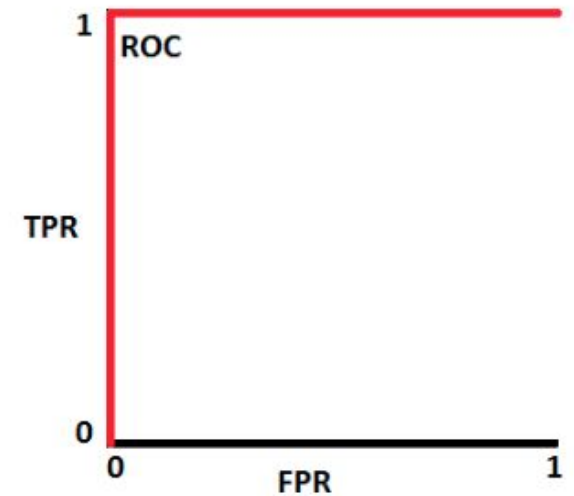
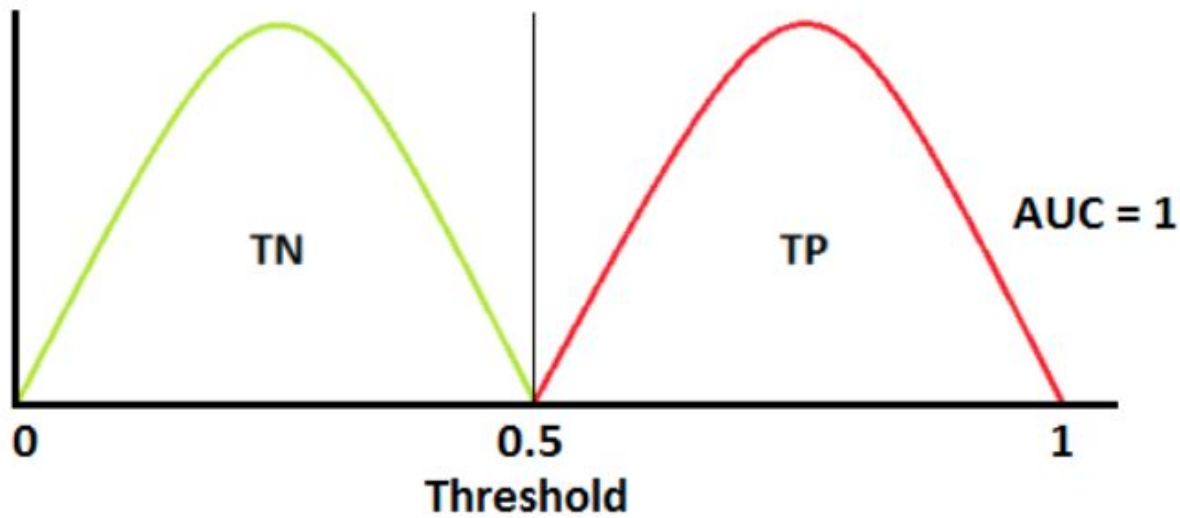


ROC Curve and AUC

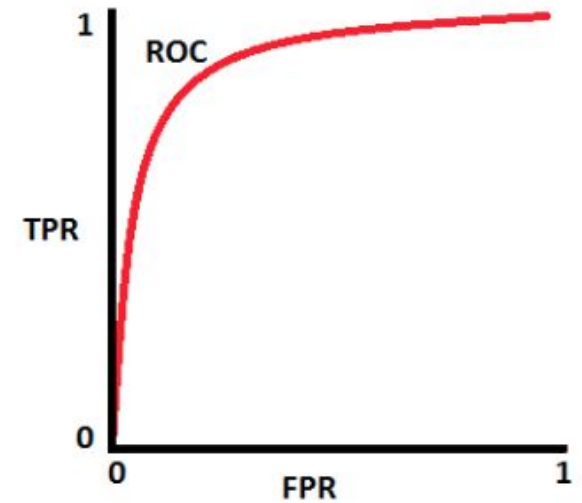
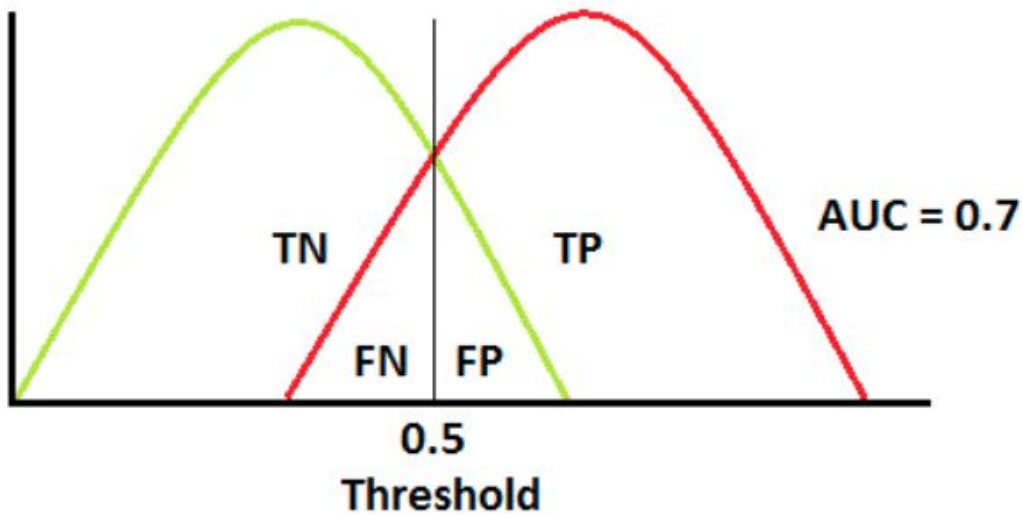
- กราฟนี้จะเรียกว่า ROC Curve หรือ Receiver Operating Characteristic curve
- ค่าพื้นที่ใต้กราฟจะเรียกว่า AUC (area under curve)



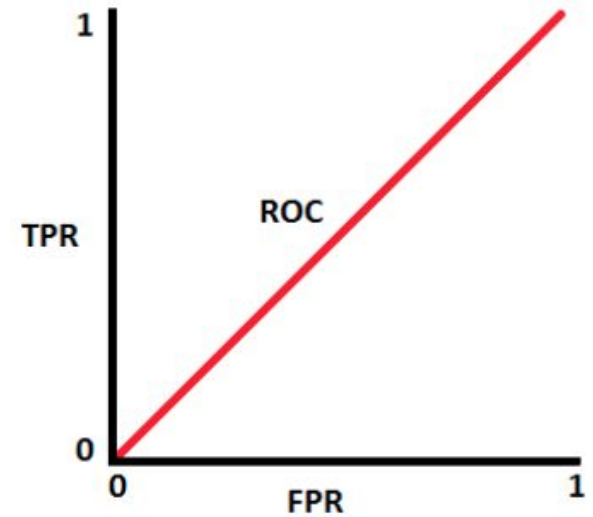
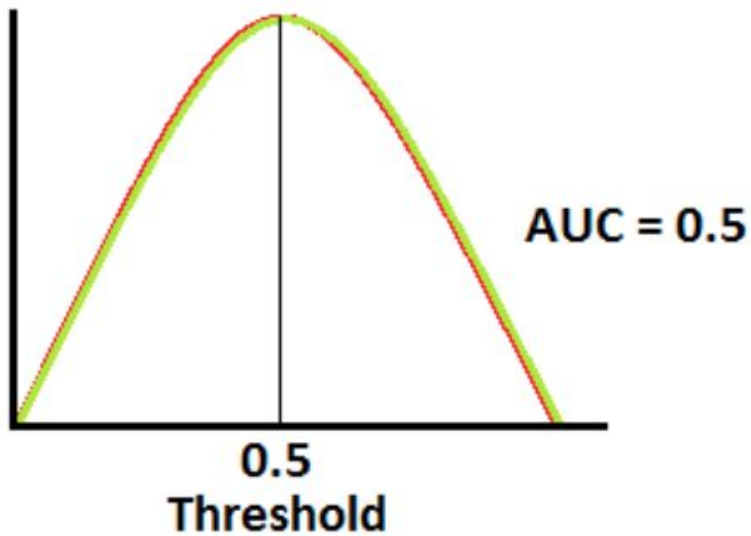
ROC Curve and AUC



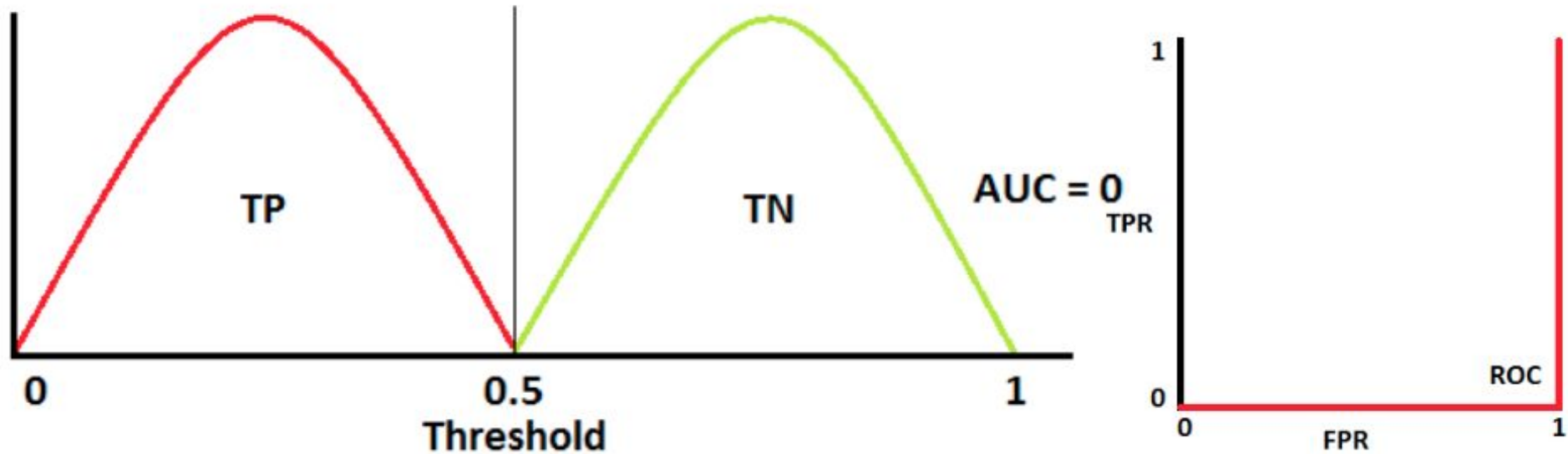
ROC Curve and AUC



ROC Curve and AUC



ROC Curve and AUC



ROC Curve and AUC

- $AUC = 0.50$: ไม่ต่างกับการเดาสุ่ม
- $AUC > 0.70$: เกณฑ์มาตรฐานสำหรับโมเดลส่วนใหญ่
- $AUC > 0.80$: โมเดลทำงานได้ดี
- $AUC > 0.90$: โมเดลทำงานได้ดีมาก