Classification and Confusion Matrix

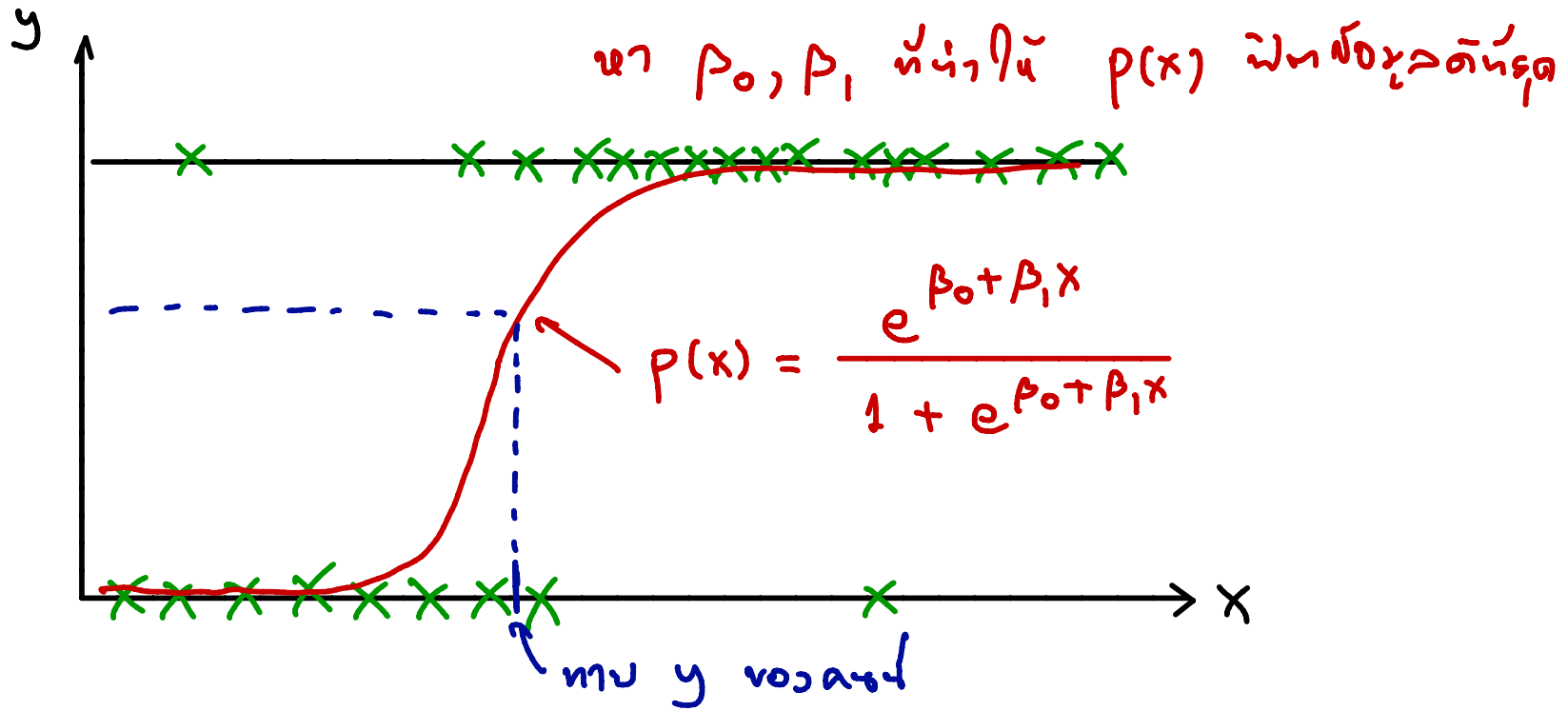
Petchara Pattarakijwanich

Introduction to Data Science, 16 September 2022

Goal of this week

- Classification Problem
- Confusion Matrix
  - Accuracy
  - Precision vs Recall
  - Sensitivity vs Specificity
  - F1 Score
- ROC curve

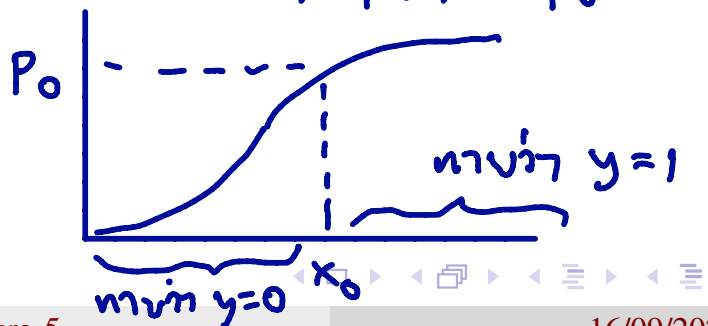
Classification Problems



x = เกรดปี 1

$y = \begin{cases} 1 & \text{, จบ} \\ 0 & \text{, ไม่จบ} \end{cases}$

$\hat{y} = \begin{cases} 1 & , p(x) \geq p_0 \\ 0 & , p(x) < p_0 \end{cases}$



How to Measure Model Accuracy?

Regression

test data — predicted

x	y	$\hat{y} = f(x)$
.	1.1	1.4
.	2.5	2.2
.	.	.
.	.	.
.	.	.
.		

เทียบ 2 คอลัมน์

$$RSS = \sum (y - \hat{y})^2$$

ต่างกัน → โมเดลแม่น

Classification

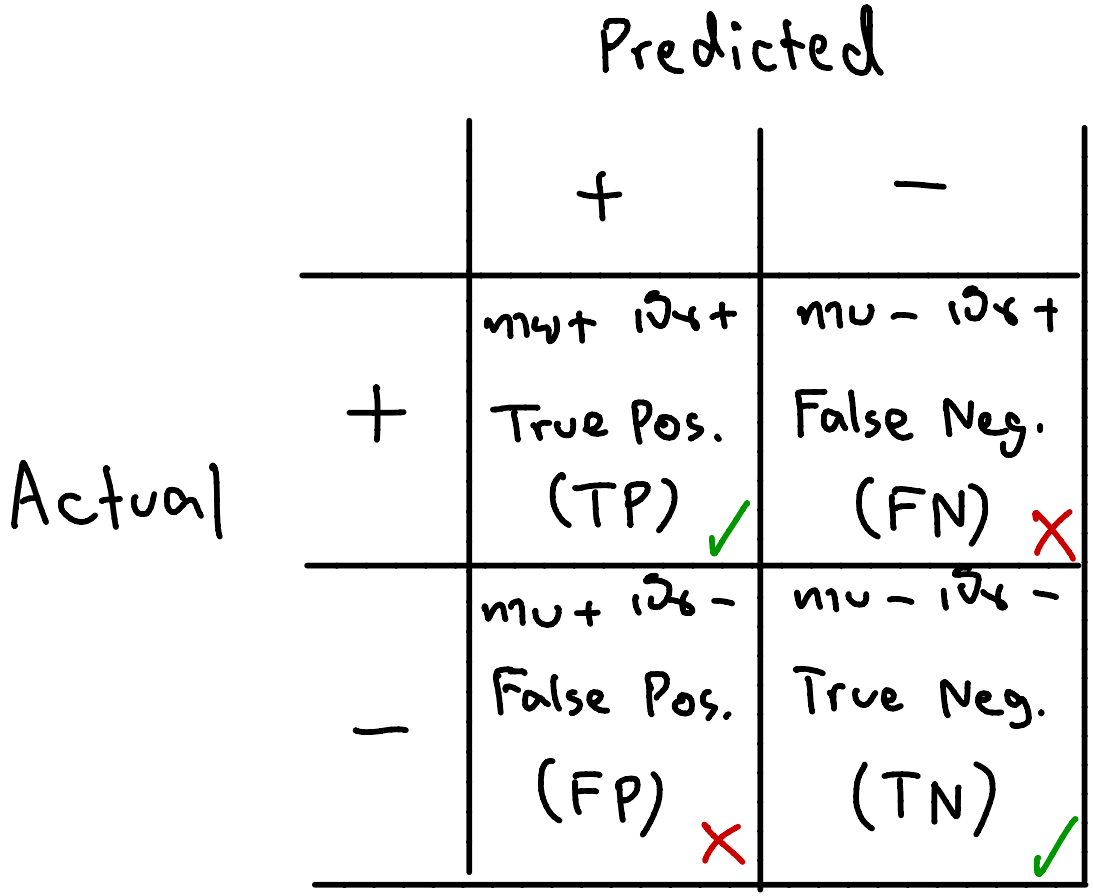
test data — predicted

x	y	$\hat{y} = f(x)$
.	1	1
.	1	0
.	0	1
.	0	0
.	.	.
	.	.

เทียบ 2 คอลัมน์

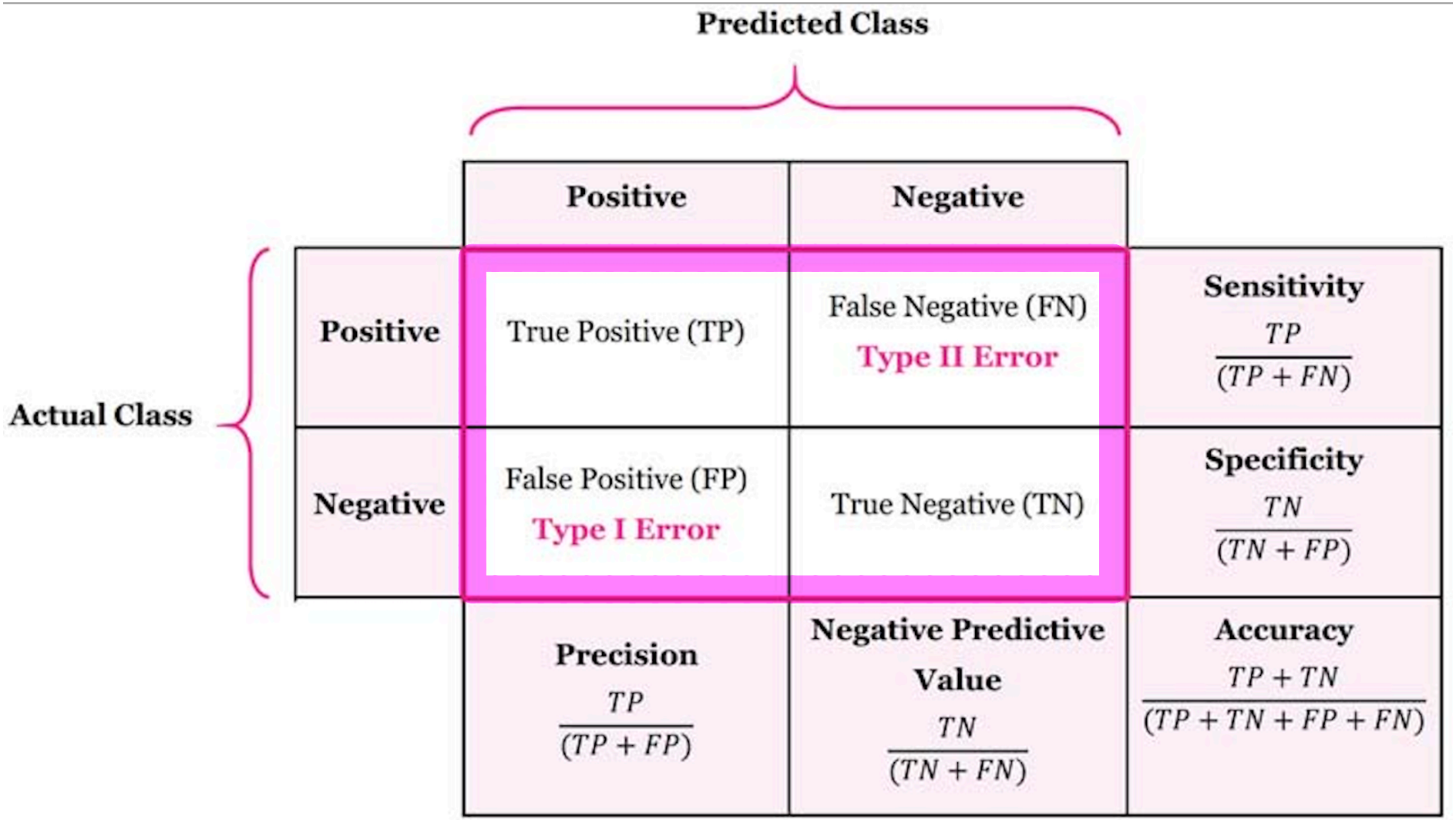
Accuracy = % ทายถูก

Confusion Matrix



False Negative

Confusion Matrix

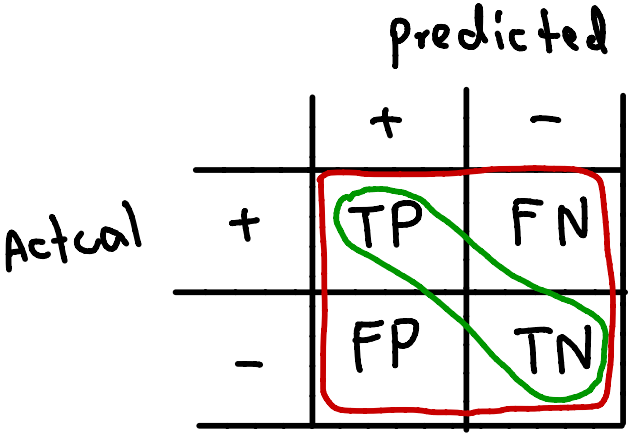


Confusion Matrix

Sources: [16][17][18][19][20][21][22][23][24] view · talk · edit

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP),** **hit**	**False negative (FN),** **type II error, miss,** **underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP),** **type I error, false alarm,** **overestimation**	**True negative (TN),** **correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN}$ $= 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

Accuracy



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

$$= \text{\% ทายถูก}$$

Accuracy สูง $\Rightarrow$ FP + FN ต่ำ

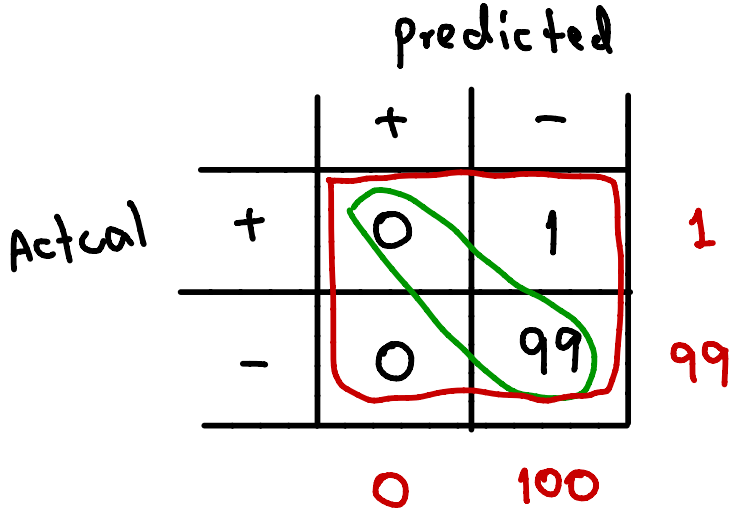
$\Rightarrow$ เส้นรอบแบ่งได้ดีเท่านั้น

Accuracy

Situation where high accuracy may not necessarily be good?

โรคหายาก มีคน 1% ของประชากร

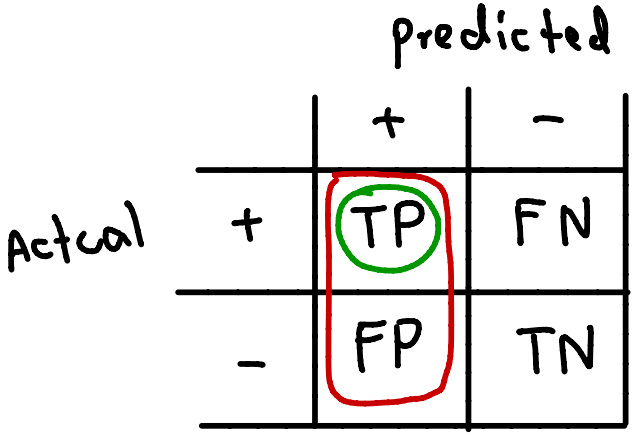
ถ้าทายผิด ⇒ ของว่า negative เสมอ (naive classifier)



$$\text{Accuracy} = \frac{99}{100} = 99\%$$

← (Imbalance Sample)

Precision



$$\text{Precision} = \frac{TP}{TP + FP}$$

= เมื่อทายว่าบวก ทายถูกกี่ %

Precision สูง ⇒ FP ต่ำ

Precision

Situation where one needs high precision?

โมเดลทำนายตลาดหุ้น → บวก = หุ้นขึ้น ⇒ ซื้อหุ้น

→ ลบ = หุ้นลง ⇒ อย่าเลย

	+	−
+	กำไร	ตกรถ (ไม่เจ๊ง)
−	เจ๊ง	รอด (ไม่เจ๊ง)

$$\text{Precision} = \frac{\#\text{กำไร}}{\#(\text{ทายว่าบวก})}$$

	+	−
+	9	80
−	1	10

Accuracy ต่ำ

Precision สูง

Sensitivity (อีกชื่อ Recall)

		predicted +	predicted −
Actual	+	TP	FN
	−	FP	TN

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN}$$

= ในบรรดาที่เป็นบวกจริง เก็บมาได้กี่ %

Sensitivity สูง ⇒ FN ต่ำ

⇒ ไม่มีที่เป็นบวกหลุดรอด

Sensitivity

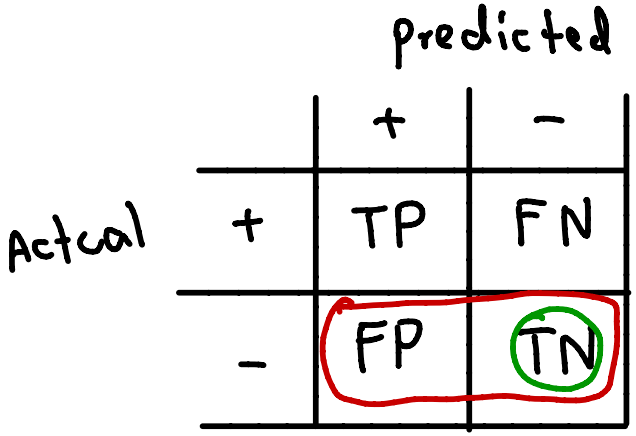
Situation where one needs high sensitivity?

หมอมะเร็ง ↗ บวก = เจอมะเร็ง ⇒ รักษา
↘ ลบ = ไม่เจอมะเร็ง ⇒ ไม่ทำอะไร

	+	−
+	เจอมะเร็ง รักษา	หมอพลาด ⇒ ฟ้องหมอ
−	หัวใจพัง ⇒ โล่งใจ	ไม่มีมะเร็ง

ต้องการ sensitivity สูง เพราะว่า FN ร้ายแรง

Specificity



$$\text{Specificity} = \frac{TN}{TN + FP}$$

= ในบรรดาทั้งหมดที่เป็นลบ เค้าทายได้กี่ %

Specificity สูง ⇒ FP ต่ำ

⇒ ไม่ทายว่าบอกผิดเยอะ

Specificity

Situation where one needs high specificity?

Spam detection
- บวก = ทายว่า spam ⇒ คน
- คน = ทายว่าไม่ spam ⇒ ปล่อย

	+	−
+	spam คน	ปล่อย spam ⇒ รำคาญ
−	คนเมารัดัง ⇒ ร้ายแรง	ไม่ spam ปล่อย

ต้องการ specificity สูง

F1 Score

$$\frac{1}{\text{F1-score}} = \frac{1}{2}\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- ค่าเฉลี่ย Harmonic ของ precision & Recall
- ถ้า Precision หรือ Recall ตัวใดตัวหนึ่งต่ำ => F1 ต่ำ
- Precision และ Recall ต้องสูงทั้งคู่ F1 ถึงจะสูง

Sensitivity-Specificity Tradeoff

ตัวอย่าง (ในรพ.) ทายว่า ผู้ป่วย ท้องมั้ย?

- บวก = ท้อง ⇒ ลุกให้นั่ง
- ลบ = ไม่ท้อง ⇒ ไม่ลุก

	+	−
+	ลุกให้นั่งท้อง ⇒ ดี	ปล่อยผู้ป่วยท้องมั้น ⇒ ไม่โอเค
−	ลุกให้นั่งทั้งไม่ท้อง ⇒ โอเคน้อย	ไม่ลุกให้นั่งไม่ท้อง ⇒ โอเค

เอียงว่าท้อง

ดีขึ้น

- FN ร้ายแรงมาก
- ยอมโดนด่าฟรี ไม่ยอมปล่อยผู้ป่วยท้องมั้น
- บางที่ มีแนวโน้มจะทายว่าท้องมากขึ้น
- FN ลด , FP เพิ่ม
- Sensitivity เพิ่ม , Specificity ลด

Sensitivity-Specificity Tradeoff

ตัวอย่าง (ในรพ.) ทายว่า ผู้ป่วย ท้องมั้ย?

- บวก = ท้อง ⇒ ลูกให้นม
- ลบ = ไม่ท้อง ⇒ ไม่ลูก

	+	−
+	ลูกให้นมท้อง ⇒ ดี	ปล่อยคนท้องยืน ⇒ ไม่โอเค
−	ลูกให้นมไม่ท้อง ⇒ โคตร่า	ไม่ลูกให้นมไม่ท้อง ⇒ โอเค

เอียงว่าท้อง

คนดี

- ลูกโคตร 3 ทายว่าทุกคนท้อง

	+	−
+		0
−		0

⇒ sensitivity = 1

specificity = 0

Sensitivity-Specificity Tradeoff

ตัวอย่าง (โรค ฟฟ.) ทายว่า ผู้ป่วย ท้องยัง?

- บวก = ท้อง ⇒ ลุกให้นั่ง
- ลบ = ไม่ท้อง ⇒ ไม่ลุก

	+	−
+	ลุกให้นั่งท้อง ⇒ ดี	ปล่อยนั่งท้องยัง ⇒ ไม่โอเค
−	ลุกให้นั่งไม่ท้อง ⇒ โดนด่า	ไม่ลุกให้นั่งไม่ท้อง ⇒ โอเค

⇒ ยังไงก็ไม่ท้อง

ผลลัพธ์โดนด่า

- FP ร้ายแรงมาก
- ยอมปล่อยผู้ป่วยผิด ไม่ยอมโดนด่า
- บางรัฐ มีแนวโน้มจะทายว่า ไม่ท้อง มากขึ้น
- FP ลด , FN เพิ่ม
- Sensitivity ลด , Specificity เพิ่ม

Sensitivity-Specificity Tradeoff

ตัวอย่าง (ในรพ.) ทายว่า ผู้ป่วย ท้องมั้ย?

- บวก = ท้อง ⇒ ลูกให้นม
- ลบ = ไม่ท้อง ⇒ ไม่ลูก

	+	−
+	ลูกให้นมท้อง ⇒ ดี	ปล่อยคนท้องยื่น ⇒ ไม่โอเค
−	ลูกให้นมไม่ท้อง ⇒ โดนด่า	ไม่ลูกให้นมไม่ท้อง ⇒ โอเค

⇒ แย่งว่าไม่ท้อง

คนหัวโดนด่า

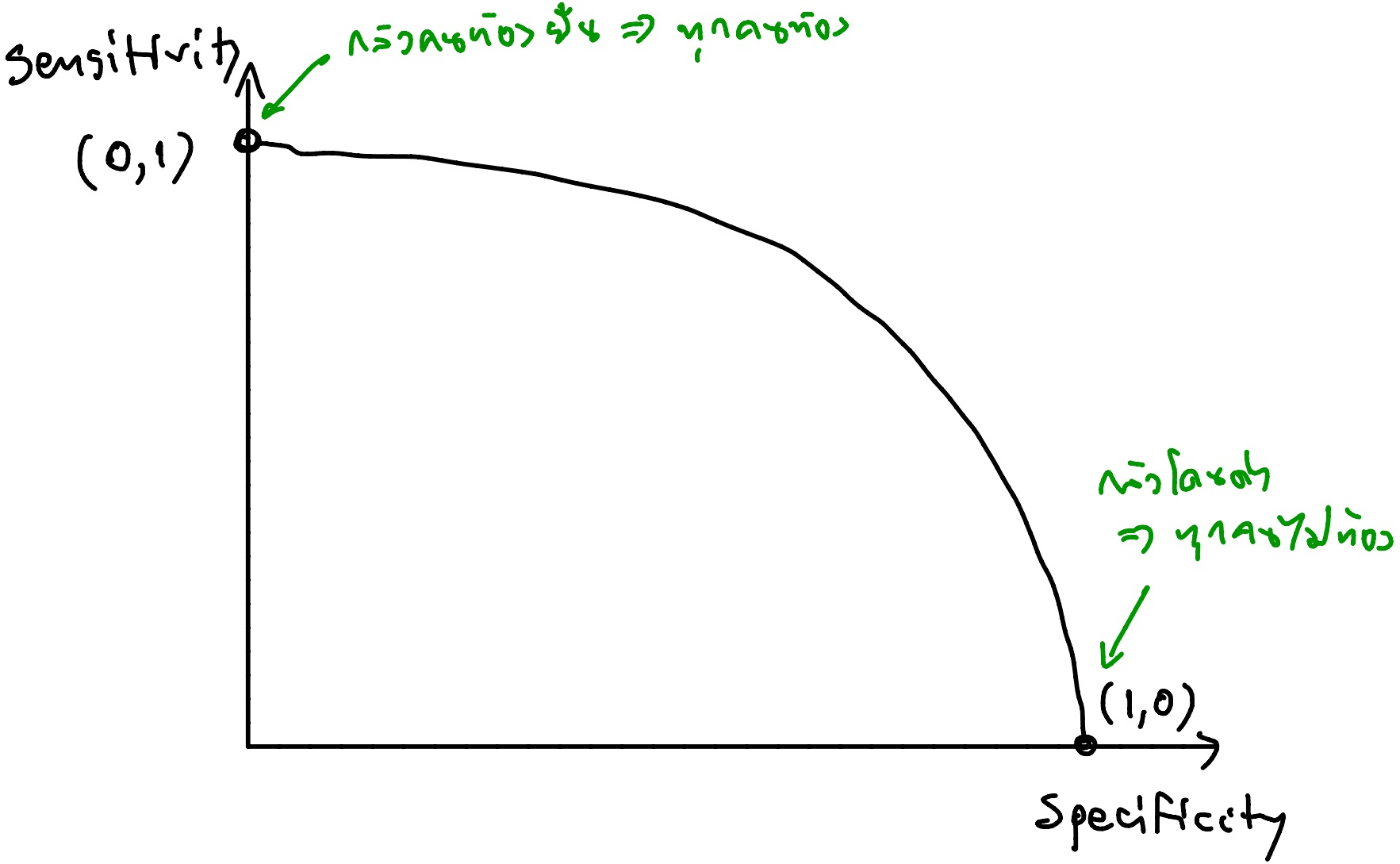
- ลูกโดนด่า ; ทำนายว่า ทุกคนไม่ท้อง

	0	
	0	

⇒ sensitivity = 0

specificity = 1

Sensitivity-Specificity Tradeoff



Receiver Operating Characteristic (ROC) Curve