

**KHOA CNTT & TRUYỀN THÔNG**  
**BM KHOA HỌC MÁY TÍNH**

# **Đánh giá hiệu quả của giải thuật học**

PGS. TS. Đỗ Thanh Nghị  
TS. Trần Nguyễn Minh Thư  
tnmthu@ctu.edu.vn

# Nội dung

- Nghi thức kiểm tra
- Các chỉ số đánh giá

# Nghi thức kiểm tra

- nếu dữ liệu có **1 tập học** và **1 tập kiểm tra** sẵn dùng
  - dùng dữ liệu học để xây dựng mô hình,
  - dùng tập kiểm tra để đánh giá hiệu quả của giải thuật
- nếu dữ liệu **không có 1 tập kiểm tra** sẵn?

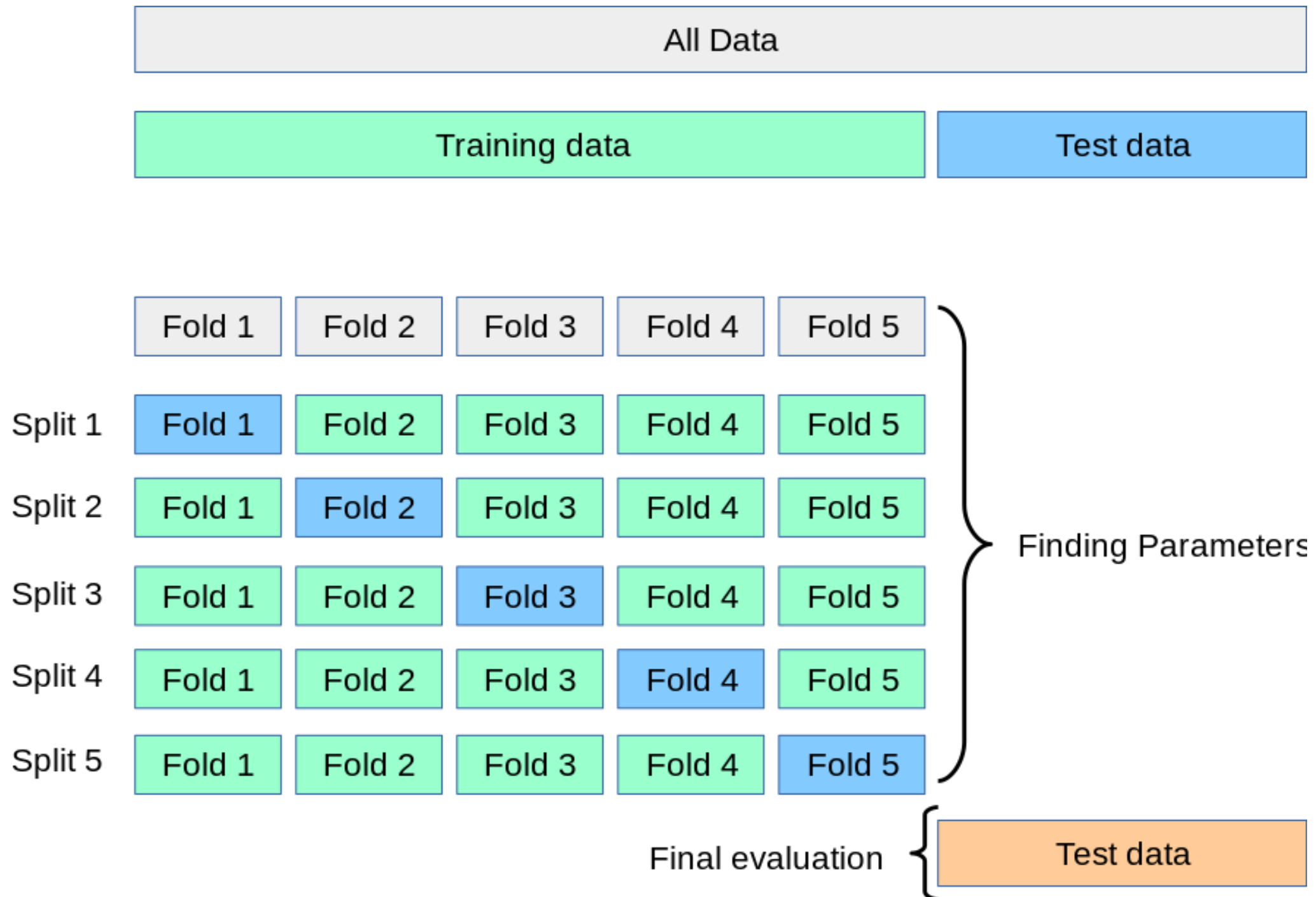
# Nghi thức kiểm tra

➤ nếu dữ liệu không có 1 tập kiểm tra sẵn

□ sử dụng nghi thức **k-fold** :

chia tập dữ liệu thành  $k$  phần (fold) bằng nhau, lặp lại  $k$  lần, mỗi lần sử dụng  $k-1$  folds để học và 1 fold để kiểm tra, sau đó tính trung bình của  $k$  lần kiểm tra

□ nghi thức **hold-out** : lấy ngẫu nhiên  $2/3$  tập dữ liệu để học và  $1/3$  tập dữ liệu còn lại dùng cho kiểm tra, có thể lặp lại quá bước này  $k$  lần rồi tính giá trị trung bình



[https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html)

# Nghi thức kiểm tra

- nếu dữ liệu có số phần tử lớn hơn 300  
sử dụng nghi thức k-fold với  $k = 10$
- nếu dữ liệu có số phần tử nhỏ hơn 300  
sử dụng nghi thức leave-1-out (k-fold với  $k = \text{số phần tử}$ )  
 $\Rightarrow$  Vd leave 1 out

➤ Chỉ số đánh giá

## Confusion matrix (C) cho 2 lớp (+/-)

dự đoán =>	dương	âm
dương	TP	FN
âm	FP	TN

TP: true positive

tổng số phần tử lớp dương được giải thuật dự đoán lớp dương

TN: true negative

tổng số phần tử lớp âm được giải thuật dự đoán là lớp âm

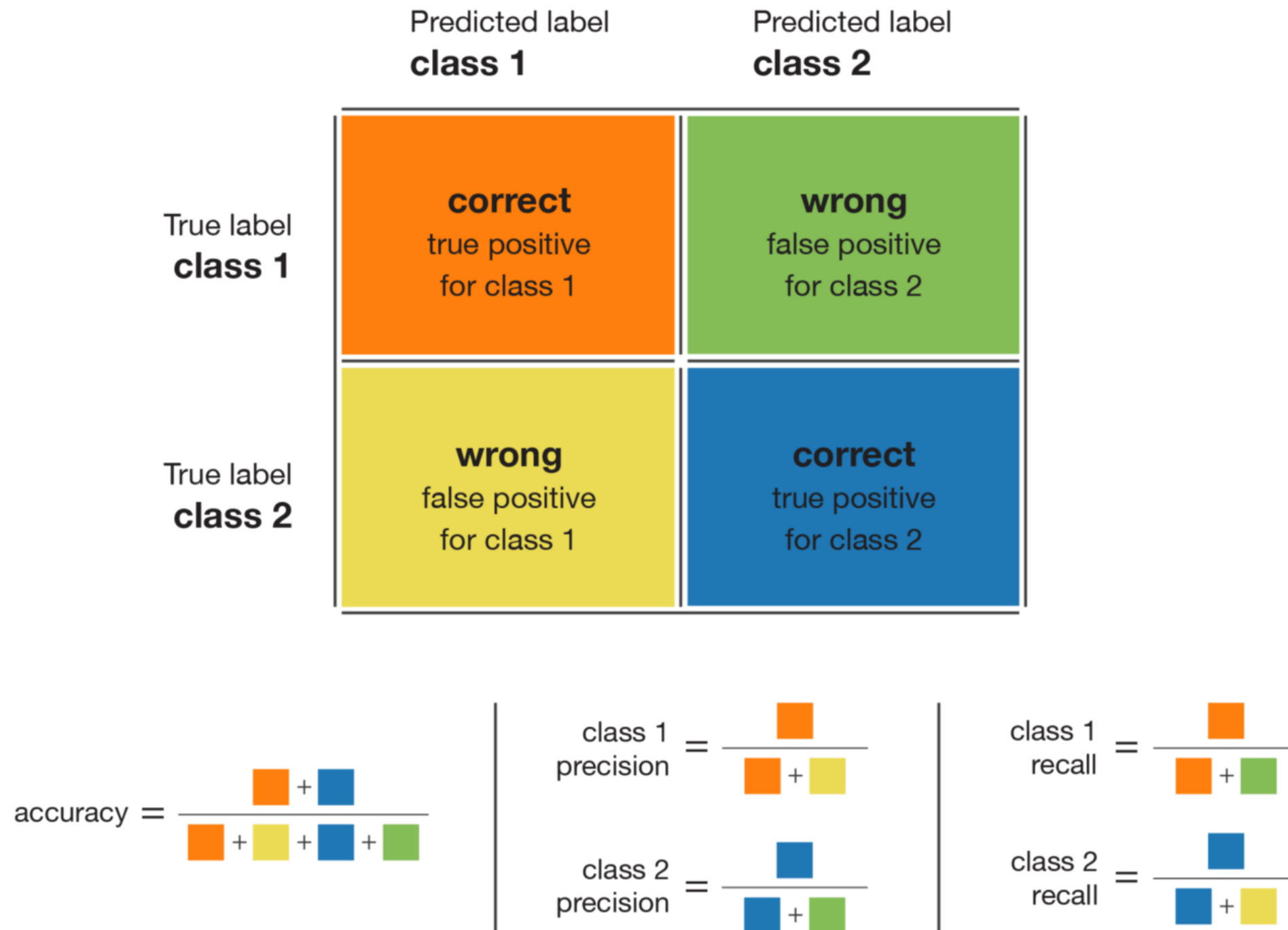
FP: false positive

tổng số phần tử lớp âm được giải thuật dự đoán là lớp dương

FN: false negative

tổng số phần tử lớp dương được dự đoán là lớp âm





# Ví dụ

	Predicted label <b>class 1</b>	Predicted label <b>class 2</b>
True label <b>class 1</b>	<b>correct</b> true positive for class 1	<b>wrong</b> false positive for class 2
True label <b>class 2</b>	<b>wrong</b> false positive for class 1	<b>correct</b> true positive for class 2

Mô hình nhận dạng ảnh con chó và con mèo. Tập kiểm tra gồm có 100 ảnh con chó và 80 ảnh con mèo. Trong 100 ảnh con chó, phần mềm nhận dạng đúng 80 ảnh. Đối với ảnh con mèo, phần mềm nhận dạng đúng 50 ảnh. Hãy xây dựng confusion matrix để đánh giá mô hình nhận dạng và tính độ chính xác (accuracy).

# Ví dụ

	Predicted label <b>class 1</b>	Predicted label <b>class 2</b>
True label <b>class 1</b>	80	?
True label <b>class 2</b>	?	50

Mô hình nhận dạng ảnh con chó và con mèo. Tập kiểm tra gồm có 100 ảnh con chó và 80 ảnh con mèo. Trong 100 ảnh con chó, phần mềm nhận dạng đúng 80 ảnh. Đối với ảnh con mèo, phần mềm nhận dạng đúng 50 ảnh. Hãy xây dựng confusion matrix để đánh giá mô hình nhận dạng.

# Actual Values

1

0

Predicted Values

1



TYPE 1 ERROR

0



TYPE 2 ERROR



## Confusion matrix (C) cho k lớp

dự đoán =>	1	...	k
1			
...			
k			

- $C[i, j]$ : số phần tử lớp **i** (**dòng**) được giải thuật dự đoán là lớp **j** (**cột**)
- $C[i, i]$ : số phần tử phân lớp đúng
- Độ chính xác lớp **i**:  $C[i, i] / C[i, ]$
- Độ chính xác tổng thể:  $\sum C[i, i] / C$

## Confusion matrix (C) cho k lớp

dự đoán =>	Setosa	vesicolor	virginica
Setosa	15	0	0
vesicolor	0	16	2
virginica	0	3	14

❑ Độ chính xác lớp **i**:  $C[i,i] / C[i,]$

❑ Setosa = ?

❑ Vesicolor = ?

❑ Virginica = ?

❑ Độ chính xác tổng thể:  $\sum C[i,i] / C = ?$

# Confusion matrix (C) cho k lớp

dự đoán =>	Setosa	vesicolor	virginica
Setosa	15	0	0
vesicolor	0	16	2
virginica	0	3	14

- $C[i, j]$ : số phần tử lớp **i** (dòng) được giải thuật dự đoán là lớp **j** (**cột**)
- $C[i, i]$ : số phần tử phân lớp đúng
- Độ chính xác lớp **i**:  $C[i, i] / C[i, ]$ 
  - Setosa = 15/15
  - Vesicolor = 16/18
  - Virginica = 14/17
- Độ chính xác tổng thể:  $\sum C[i, i] / C = 45/50$

Giả sử ta cần dự đoán kết quả xét nghiệm của 1005 bệnh nhân xem họ có bị ung thư hay không. Dưới đây là những gì mô hình của chúng ta dự đoán:

- 90 bệnh nhân bị ung thư và tất cả dự đoán này của chúng ta đều đúng.
- 915 bệnh nhân không bị ung thư nhưng thật ra có tới 910 người lại bị trong thực tế.

Anh/chị hãy xây dựng confusion matrix và cho biết độ chính xác (Accuracy) của mô hình.

Accuracy = ...

Hãy nhận xét về Accuracy trong trường hợp này

<https://math2it.com/hieu-confusion-matrix/>

		Mô hình dự đoán	
		Ung thư	Không ung thư
Thực tế	Ung thư	90	?
	Không ung thư	?	?



- 90 bệnh nhân bị ung thư và tất cả dự đoán này của chúng ta đều đúng.
- 915 bệnh nhân không bị ung thư nhưng thật ra có tới 910 người lại bị trong thực tế.

	Thực tế (có)	Thực tế (không)
Dự đoán (có)	90 (True Positive)	0 (False Positive)
Dự đoán (không)	910 (False Negative)	5 (True Negative)

<https://math2it.com/hieu-confusion-matrix/>

# Dữ liệu không cân bằng

- nếu dữ liệu không cân bằng
  - lớp positive có số lượng rất nhỏ so với lớp negative
  - ví dụ : positive = 5%, negative = 95%
  - một giải thuật học có thể cho kết quả 95% độ chính xác khi phân loại, nhưng chúng ta có thể mất hoàn toàn lớp positive
- khả năng tách lớp positive từ lớp negative

# Cần chọn chỉ số đánh giá phù hợp

- **Precision**
- **Recall**
- **Accuracy**
- **F1**

dự đoán =>	dương	âm
dương	TP	FN
âm	FP	TN

$$prec = \frac{tp}{tp + fp}$$

$$rec = \frac{tp}{tp + fn}$$

$$acc = \frac{tp + tn}{tp + fn + tn + fp}$$

$$F1 = \frac{2 \times prec \times rec}{prec + rec}$$

dự đoán =>	dương	âm
dương	10 (TP)	5 (FN)
âm	8 (FP)	22 (TN)

## Confusion matrix (C) cho 2 lớp (+/-)

dự đoán =>	dương	âm
dương	10 (TP)	5 (FN)
âm	8 (FP)	22 (TN)

$$\text{Precision} = 10/(10+8) = 0.56$$

$$\text{Recall} = 10/(10+5) = 0.67$$

$$\begin{aligned}\text{Accuracy} &= (10+22)/10+5+8+22 \\ &= 32/45 = 0.71\end{aligned}$$

$$\begin{aligned}\text{F1} &= 2 \times \text{precision} \times \text{recall} / (\text{prec} + \text{recall}) \\ &= 0.75/1.23 = 0.61\end{aligned}$$

$$\text{prec} = \frac{tp}{tp + fp}$$

$$\text{rec} = \frac{tp}{tp + fn}$$

$$\text{acc} = \frac{tp + tn}{tp + fn + tn + fp}$$

$$\text{F1} = \frac{2 \times \text{prec} \times \text{rec}}{\text{prec} + \text{rec}}$$

# Bài tập

Giả sử tập dữ liệu có 40000 mẫu tin trong đó có 8 mẫu tin thuộc lớp dương (+1) và 39992 mẫu tin thuộc lớp âm (-1), có hai mô hình phân lớp M1 và M2 cho kết quả tương ứng trong bảng 1, 2 như bên dưới. Anh/chị hãy cho biết mô hình nào thích hợp để xử lý tập dữ liệu trên? Hãy giải thích lý do cho lựa chọn đó.

Ma trận confusion thu được từ mô hình M1 (bảng bên trái) và M2 (bảng bên phải)

<u>dự báo =&gt;</u>	<u>dương</u>	<u>âm</u>
<u>dương</u>	1	7
<u>âm</u>	1	39991

<u>dự báo =&gt;</u>	<u>dương</u>	<u>âm</u>
<u>dương</u>	8	0
<u>âm</u>	32	39960

# Bài tập

Ma trận confusion thu được từ mô hình M1 (bảng bên trái) và M2 (bảng bên phải)

<u>dự báo =&gt;</u>	<u>dương</u>	<u>âm</u>
<u>dương</u>	1	7
<u>âm</u>	1	39991

M1:

$$\text{recall} = 1/8 = 12.5\%,$$

$$\text{precision} = 1/2 = 50\%,$$

$$F1 = 2 \times 12.5 \times 50 / 62.5 = 20\%$$

<u>dự báo =&gt;</u>	<u>dương</u>	<u>âm</u>
<u>dương</u>	8	0
<u>âm</u>	32	39960

M2:

$$\text{recall} = 8/8 = 100\%,$$

$$\text{precision} = 8/40 = 20\%,$$

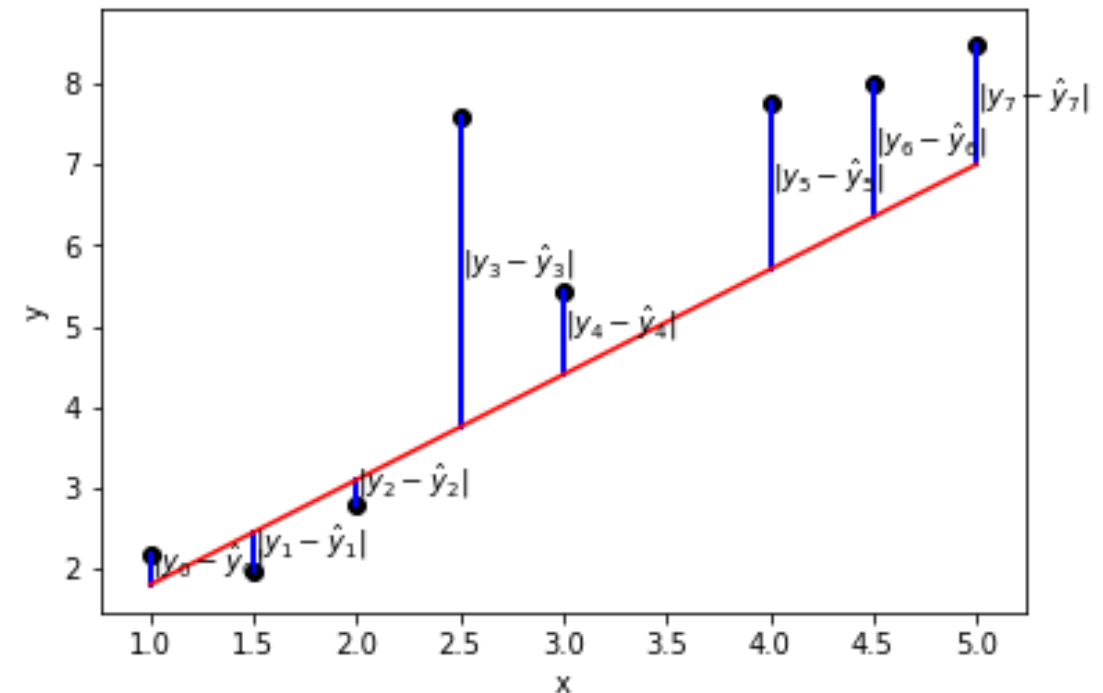
$$F1 = 2 \times 100 \times 20 / (120) \\ = 33.33\%$$

# Chỉ số đánh giá cho bài toán hồi quy

Đánh giá độ chính xác của các giá trị dự đoán

➤ Các chỉ số thường dùng:

- MAE – Mean Absolute Error
- MSE – Mean Square Error
- RMSE – Root Mean Square Error



- Đo lường mức độ sai số của các dự đoán. Các giá trị đo lường này bằng 0 khi hệ thống đạt được hiệu quả tốt nhất. Giá trị này càng cao thì hiệu quả của hệ thống càng thấp.
- MAE là chỉ số được sử dụng nhiều nhất vì khả năng giải thích trực tiếp của nó

# Chỉ số đánh giá cho bài toán hồi quy

## Đánh giá độ chính xác của các dự đoán

- MAE – Mean Absolute (sai số trung bình tuyệt đối) tính toán độ lệch giữa dự đoán xếp hạng và xếp hạng thực tế
- Tính chính xác của dự đoán được đo trên **n** quan sát trong đó **p<sub>i</sub>** là giá trị dự đoán đánh giá của item **i**,

**r<sub>i</sub>** là giá trị đánh giá thực tế của item **i**

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i|$$



# Chỉ số đánh giá cho bài toán hồi quy

## Đánh giá độ chính xác của các dự đoán

- Mean Square Error ( $MSE$ ) (sai số bình phương trung bình).....

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2$$

- Root Mean Square Error ( $RMSE$ ) (sai số trung bình toàn phương) tương tự như MAE nhưng chú trọng tới những giá trị có độ lệch lớn

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2}$$

*The End*

# Nghi thức đánh giá

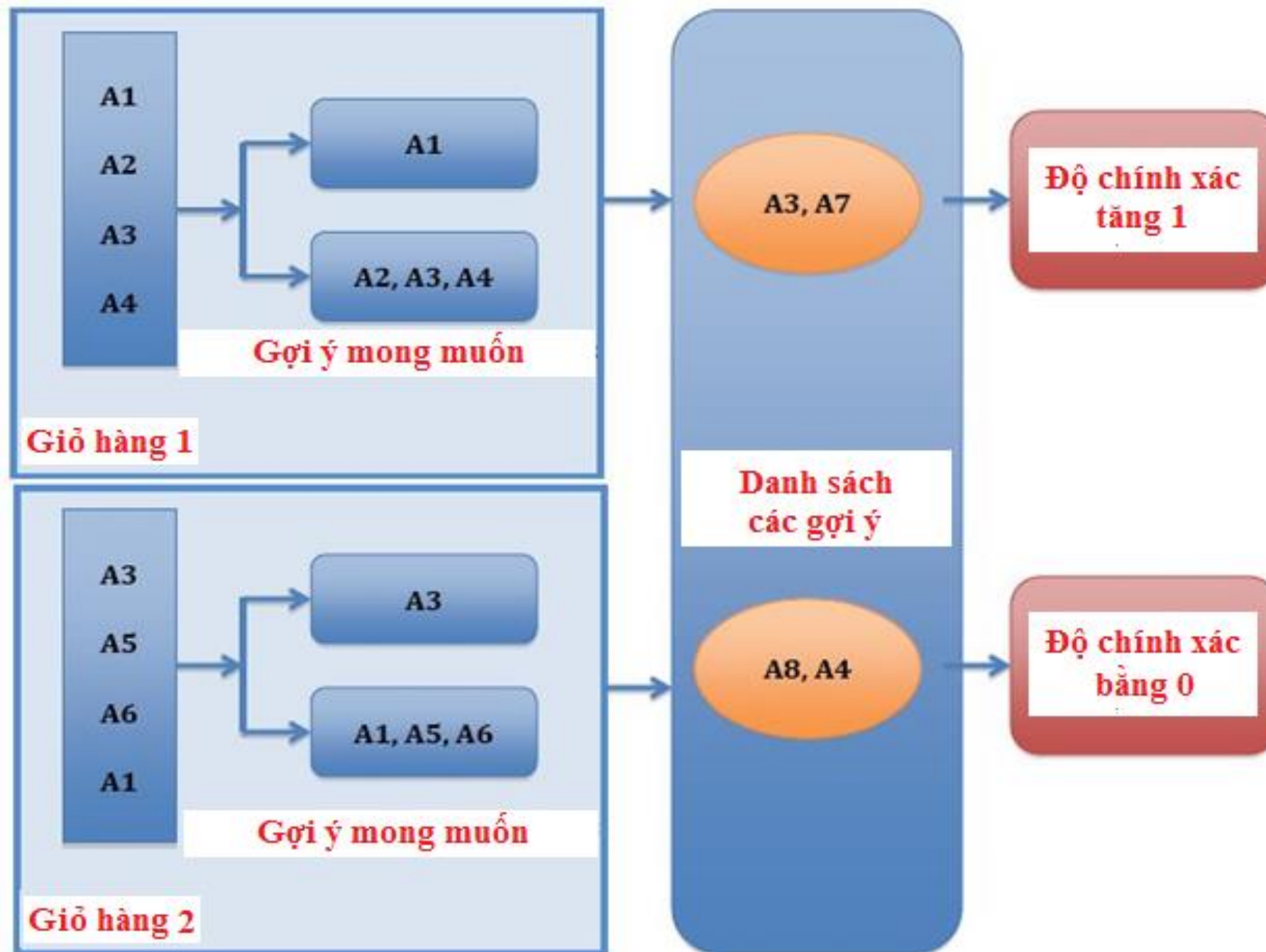
## ➤ Given-N [J.S Breese98]

- Thường được sử dụng để đánh giá trong các lĩnh vực **thương mại điện tử**
- Là một mở rộng của **k-fold** nhưng thực hiện **trên từng giao dịch** thay vì toàn bộ dữ liệu
- Giao dịch sử dụng để đánh giá phải có ít nhất **N+1** mục dữ liệu

## ➤ Phương pháp

- Chia danh sách các sản phẩm trong giỏ hàng thành 2 phần: tập được gọi là « **Given** » và 1 tập « **Test** »
- So sánh các gợi ý thực tế (**Test**) và những sản phẩm **gợi ý đề nghị bởi hệ thống**, độ chính xác của hệ thống sẽ tăng lên 1 đơn vị hay bằng 0

# Phương pháp đánh giá



# Phương pháp đánh giá

**“All But One”** trường hợp đặc biệt của **Given-N**

- Tập “**given**” : số lượng các sản phẩm của giỏ hàng - 1 (ít nhất 1 sản phẩm)
- Tập « **test** » : luôn luôn bằng 1
- Ưu điểm của phương pháp “ **All But One** ” cho phép đánh giá các giỏ hàng có kích thước lớn hơn 1 sản phẩm