

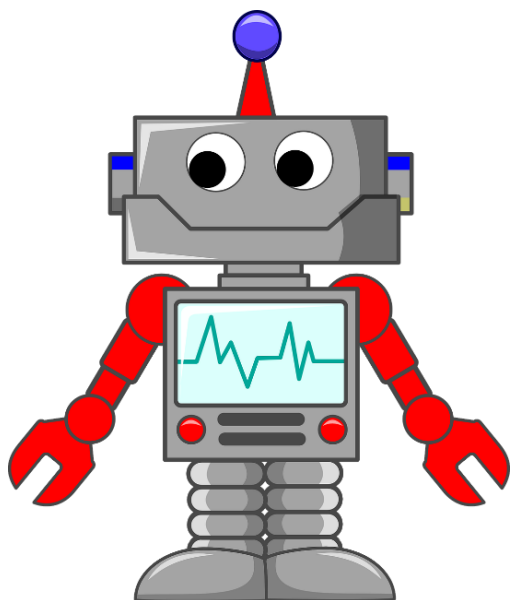


CS116 – LẬP TRÌNH PYTHON CHO MÁY HỌC

BÀI 05

ĐÁNH GIÁ MÔ HÌNH

TS. Nguyễn Vinh Tiệp





NỘI DUNG

1. TẠI SAO CẦN PHẢI ĐÁNH GIÁ MÔ HÌNH

2. ĐỘ ĐO CHO BÀI TOÁN HỒI QUY

3. ĐỘ ĐO CHO BÀI TOÁN PHÂN LỚP

4. QUY TRÌNH ĐÁNH GIÁ KHÁCH QUAN



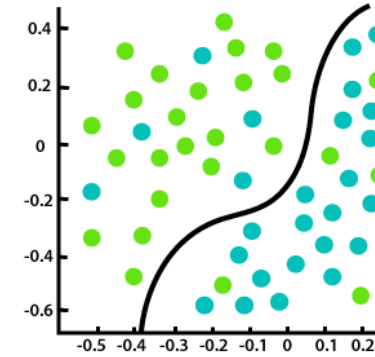
Tại sao cần đánh giá mô hình

- **Xác thực tính hiệu quả :**
 - Mô hình thực sự đoán chính xác hay chỉ nhớ dữ liệu train?
 - Có bị hiện tượng **overfitting**?
- **Chọn lựa mô hình:** cho phép so sánh các mô hình với nhau, từ đó chọn ra mô hình tốt nhất
- **Sự tin cậy:** mô hình của mình có ổn định và đáng tin cậy khi triển khai thực tế không?
- **Định hướng cải thiện mô hình:** các độ đo đánh giá giúp ta biết cần cải tiến mô hình theo hướng nào?

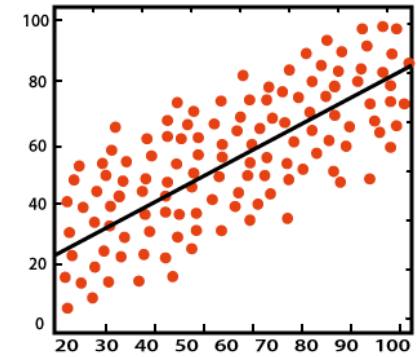


Phân biệt bài toán học máy

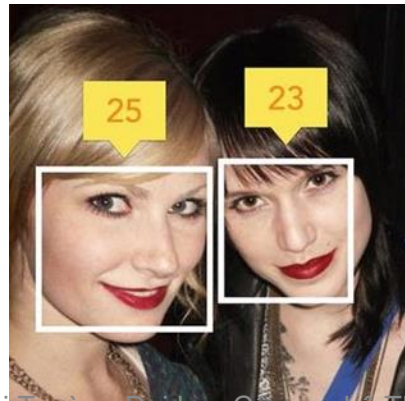
- Hai loại bài toán chính:
 - Hồi quy – Regression
 - Phân lớp – Classification
- Cách phân biệt “**đơn giản**” dựa trên tính chất của biến kết quả (output)
 - Hồi quy: đầu ra là giá trị **liên tục**
 - Phân lớp: đầu ra là giá trị **rời rạc**
- Câu hỏi:** Bài toán “dự đoán tuổi” là Hồi quy hay Phân lớp?



Classification



Regression



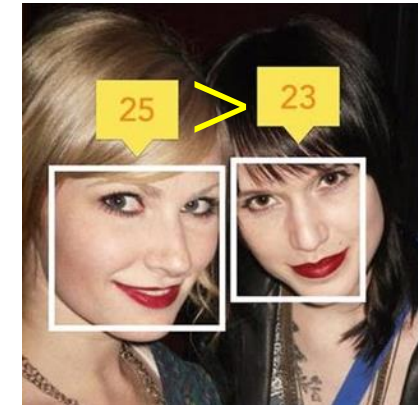
Có 2 luồng ý kiến
trái chiều



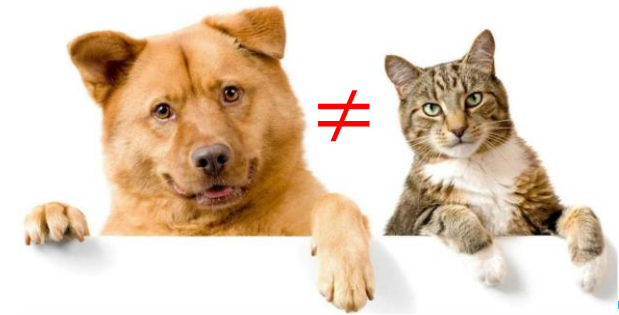
Phân biệt bài toán học máy

- Cách phân biệt “rõ ràng” hơn:
 - Hồi quy: 2 giá trị output bất kỳ có thể thực hiện tất cả so sánh $>$, $<$, $=$, \neq
 - Phân lớp: 2 giá trị output bất kỳ **chỉ có thể** thực hiện được so sánh $=$ hoặc \neq

- **Đáp án:** Bài toán “dự đoán tuổi” là **bài toán hồi quy**



- Tương tự, bài toán đoán đối tượng: “chó”, “mèo” là bài toán phân loại?





Tại sao cần Phân biệt đúng bài toán

- Để chọn lựa **hàm mục tiêu huấn luyện phù hợp**
 - Hàm loss/hàm độ lỗi đúng việc huấn luyện nhanh hơn
- Để chọn lựa **độ đo đánh giá phù hợp**
 - Đánh giá khách quan hơn



NỘI DUNG

1. TẠI SAO CẦN PHẢI ĐÁNH GIÁ MÔ HÌNH

2. ĐỘ ĐO CHO BÀI TOÁN HỒI QUY

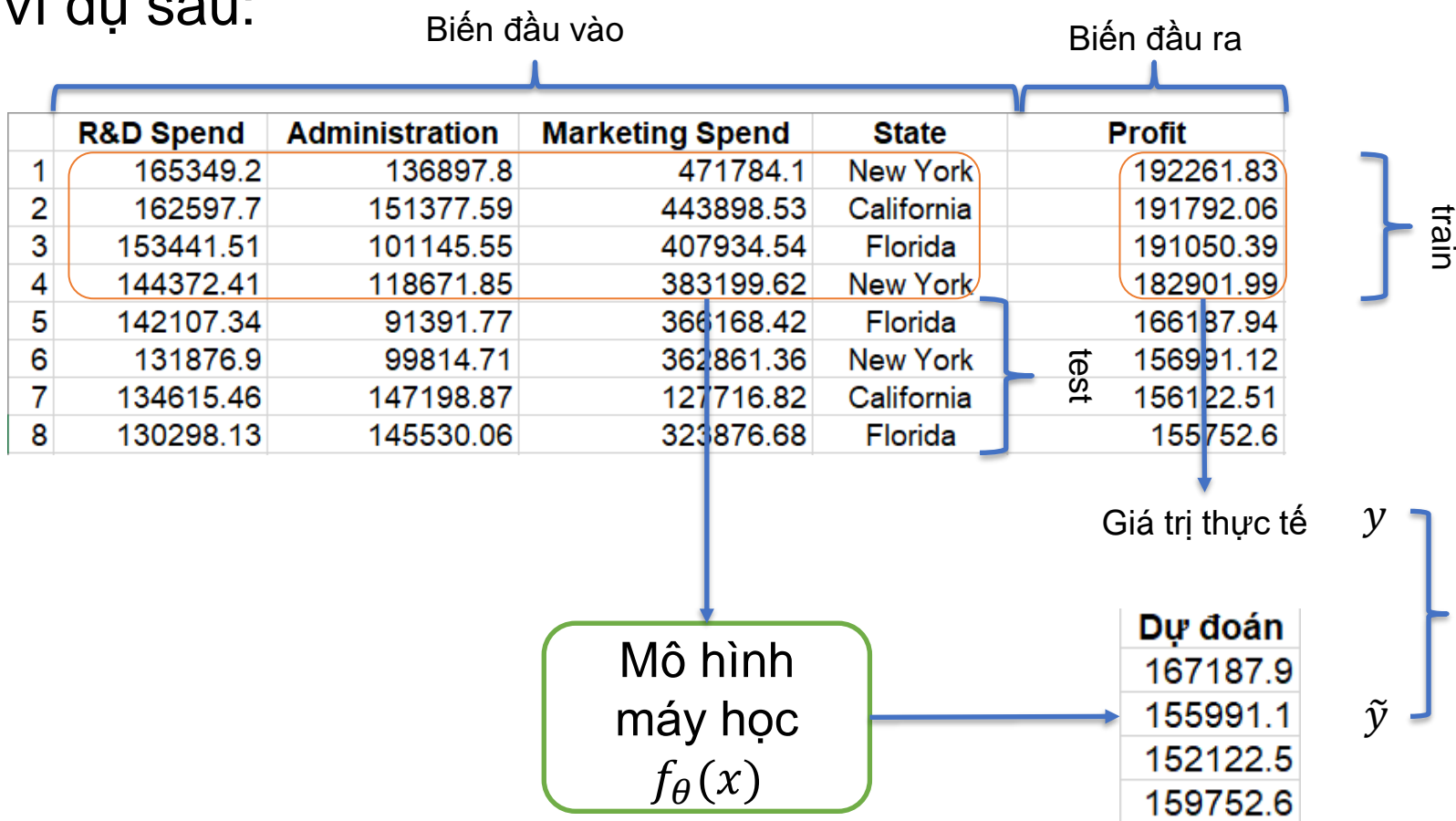
3. ĐỘ ĐO CHO BÀI TOÁN PHÂN LỚP

4. QUY TRÌNH ĐÁNH GIÁ KHÁCH QUAN



Bài toán hồi quy: chọn hàm độ lỗi (hàm loss)

- Xét ví dụ sau:





Bài toán hồi quy: độ đo đánh giá

- Độ đo đánh giá của bài toán hồi quy cũng tương tự hàm lỗi

	R&D Spend	Administration	Marketing Spend	State	Profit
1	165349.2	136897.8	471784.1	New York	192261.83
2	162597.7	151377.59	443898.53	California	191792.06
3	153441.51	101145.55	407934.54	Florida	191050.39
4	144372.41	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94
6	131876.9	99814.71	362861.36	New York	156991.12
7	134615.46	147198.87	127716.82	California	156122.51
8	130298.13	145530.06	323876.68	Florida	155752.6

train

test

Giá trị thực tế

y
 \tilde{y}

Độ đo đánh giá:

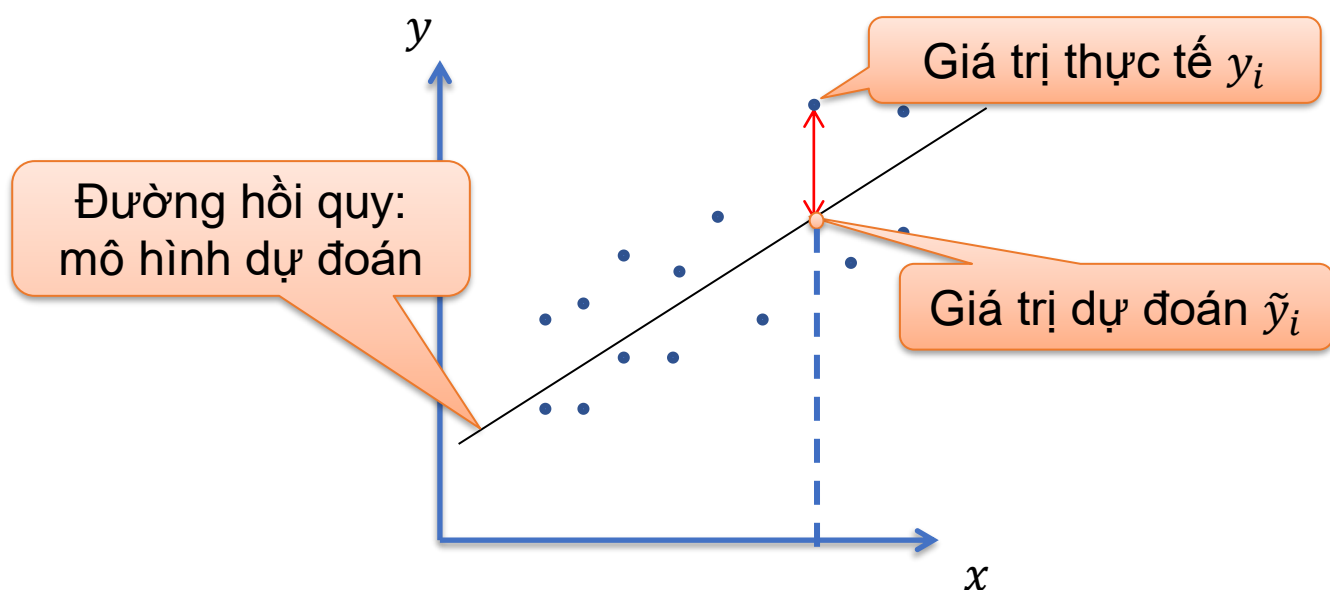
- MAE
- MSE
- RMSE
- ...





Bài toán hồi quy: độ đo đánh giá

- Các độ đo đánh giá cho bài toán hồi quy



Độ đo: đo lường sai số giữa dữ liệu thực với đường hồi quy:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$RAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - \bar{y}|}$$

$$RSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



NỘI DUNG

1. TẠI SAO CẦN PHẢI ĐÁNH GIÁ MÔ HÌNH

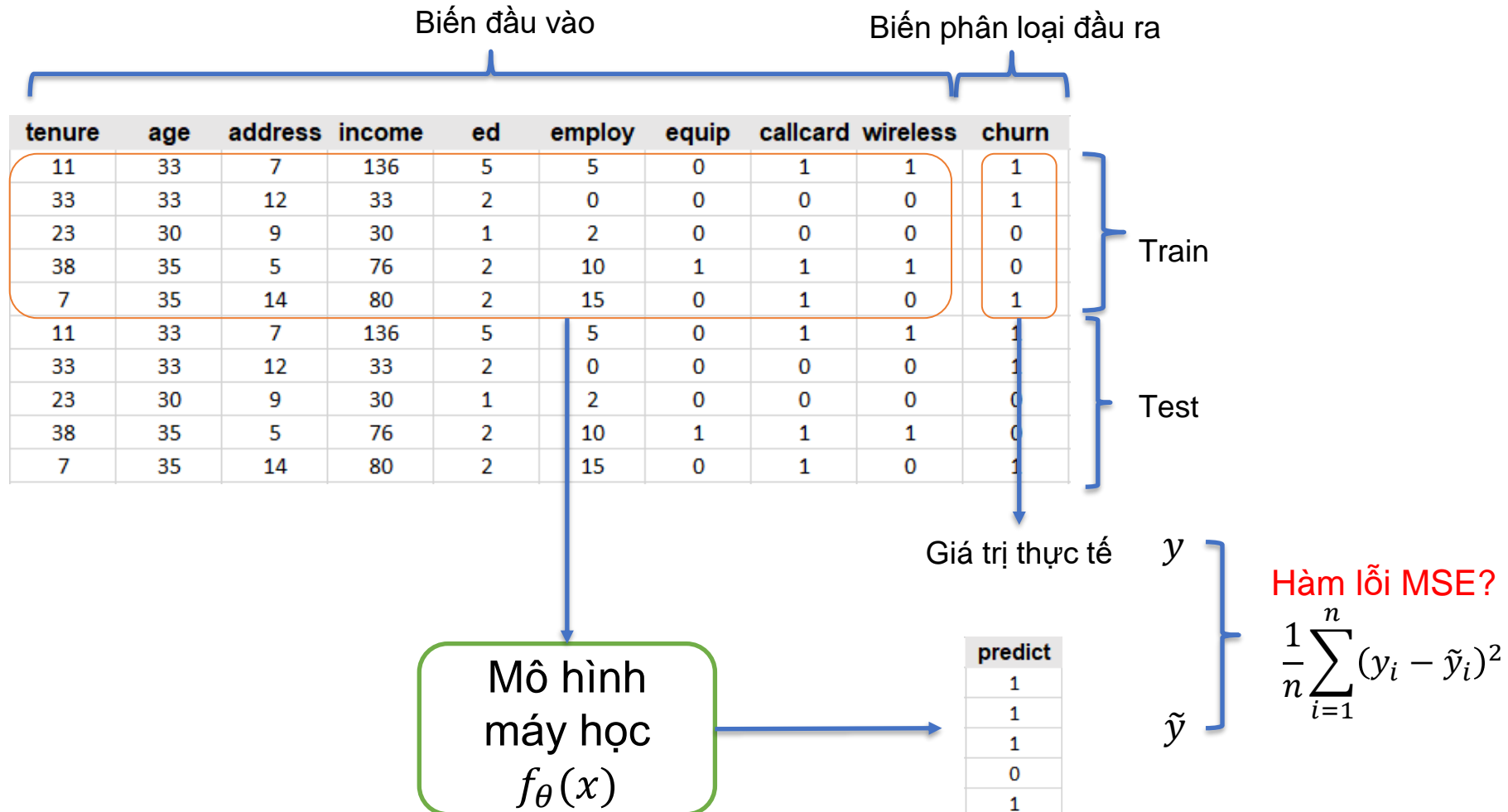
2. ĐỘ ĐO CHO BÀI TOÁN HỒI QUY

3. ĐỘ ĐO CHO BÀI TOÁN PHÂN LỚP

4. QUY TRÌNH ĐÁNH GIÁ KHÁCH QUAN

Bài toán phân lớp: chọn hàm độ lỗi (hàm loss)

- Xét ví dụ sau:



Bài toán phân lớp: hàm độ lỗi log loss

tenure	age	address	income	ed	employ	equip	callcard	wireless	churn		predict prob	Log loss
11	33	7	136	5	5	0	1	1	1	Train	0.83	0.08
33	33	12	33	2	0	0	0	0	1		0.75	0.12
23	30	9	30	1	2	0	0	0	0		0.96	1.40
38	35	5	76	2	10	1	1	1	0		0.32	0.17
7	35	14	80	2	15	0	1	0	1		0.91	0.04

Giá trị thực tế y Xác suất dự đoán \tilde{y}

Lỗi mẫu thứ i :

$$-[y \log \tilde{y} + (1 - y) \log(1 - \tilde{y})]$$

Log loss:

$$-\frac{1}{n} \sum_{i=1}^n y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)$$

Log loss: 0.0 0.1 0.2 1.0



Độ chính xác càng cao



Bài toán phân lớp: độ đo đánh giá (1)

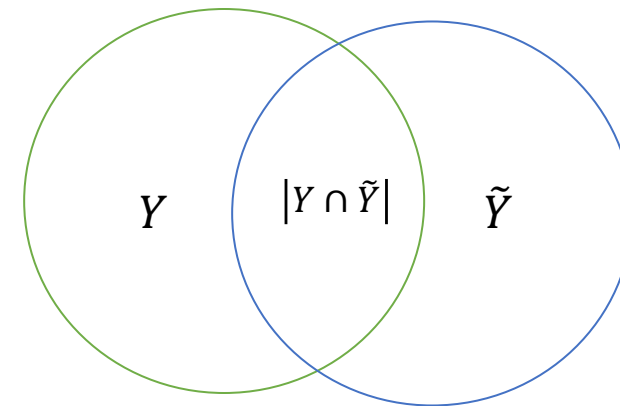
- Tại sao không sử dụng log loss làm độ đo đánh giá?
- Sử dụng Jaccard index
 - Y : Tập giá trị thực tế
 - \tilde{Y} : Tập giá trị dự đoán

$$J(Y, \tilde{Y}) = \frac{|Y \cap \tilde{Y}|}{|Y \cup \tilde{Y}|} = \frac{|Y \cap \tilde{Y}|}{|Y| + |\tilde{Y}| - |Y \cap \tilde{Y}|}$$

$$Y: [1, 1, 0, 0, 1, 0, 1]$$

$$\tilde{Y}: [1, 1, 1, 0, 1, 0, 0]$$

$$J(Y, \tilde{Y}) = \frac{5}{7 + 7 - 5} = 0.41$$





Bài toán phân lớp: độ đo đánh giá (2)

- **P**recision = $TP / (TP + FP)$
- **R**ecall = $TP / (TP + FN)$
- **F1**-score = $2 (Pr \times Rec) / (Pr + Rec)$

F1-score: 0.0 0.1 0.2 1.0



Độ chính xác càng cao

	Precision	Recall	F1-score
Churn=0			
Churn=1			

Trung bình: ?

Confusion matrix

Nhãn thực tế y	Churn=1	6 TP	9 FN
	Churn=0	1 FP	24 TN
		Churn=1	Churn=0
		Nhãn dự đoán ŷ	



Nhận xét

- Các độ đo đánh giá, hàm độ lỗi trên là cách tính trung bình, thực tế có thể có **tính thiên lệch**
- Một số độ đo đánh giá bắt nguồn **yếu tố thực tiễn**: doanh thu, lợi nhuận, lợi nhuận biên, thiệt hại kinh tế,...
- Một số hàm độ lỗi lấy trực tiếp từ độ đo đánh giá



NỘI DUNG

1. TẠI SAO CẦN PHẢI ĐÁNH GIÁ MÔ HÌNH

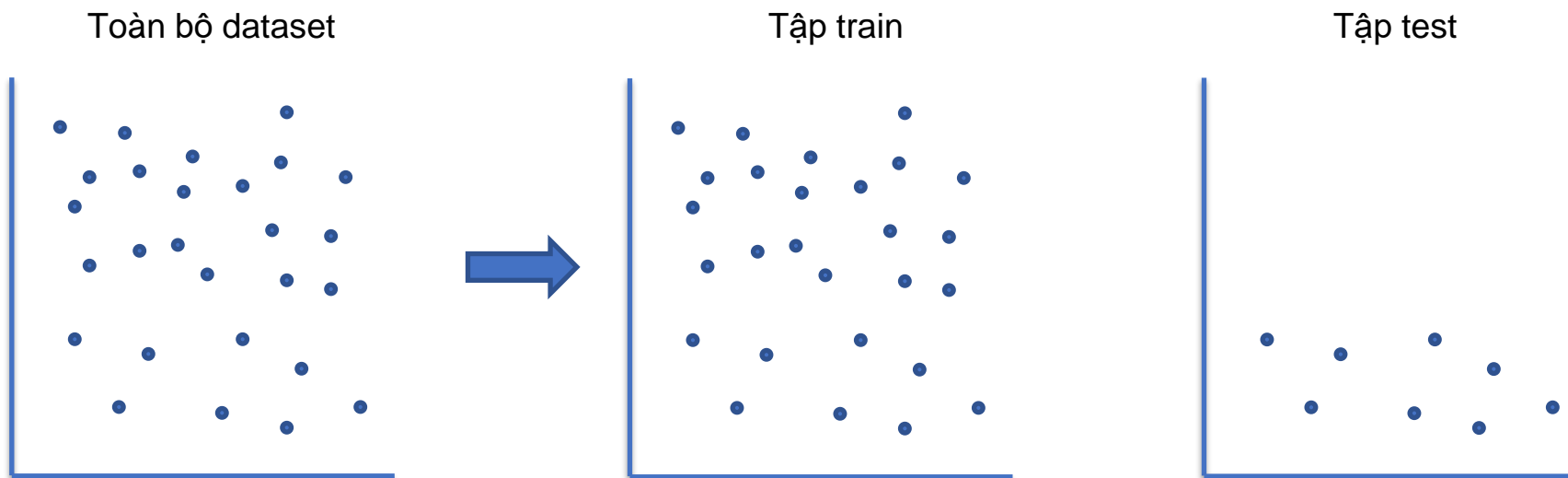
2. ĐỘ ĐO CHO BÀI TOÁN HỒI QUY

3. ĐỘ ĐO CHO BÀI TOÁN PHÂN LỚP

4. QUY TRÌNH ĐÁNH GIÁ KHÁCH QUAN

Quy trình 1: Tập test là một phần tập train

- Sử dụng một tập con của tập train làm dữ liệu test:



Độ chính xác **cao** trên tập train (và test)
Độ chính xác **thấp** trên tập ngoài mẫu
(out-of-sample dataset)



Quy trình 1: Tập test là một phần tập train

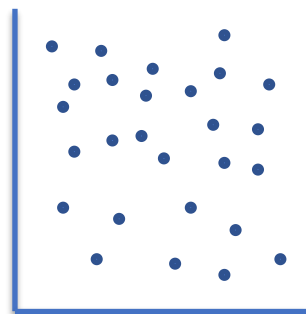
- Độ chính xác trên tập train
 - Độ chính xác cao không phản ánh đúng hiệu quả của mô hình
 - Là kết quả của hiện tượng over-fitting (quá khớp): mô hình quá đà trên tập dữ liệu dẫn đến học cả nhiễu và không có tính tổng quát
- Độ chính xác trên tập ngoài mẫu (out-of-sample)
 - Độ chính xác trên tập ngoài mẫu phản ánh hiệu quả khi áp dụng thực tế
 - Làm sao để cải thiện độ chính xác trên tập ngoài mẫu?



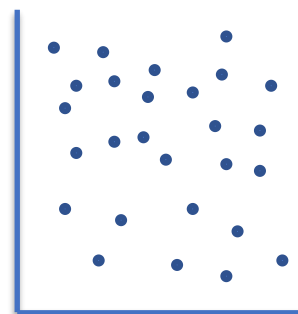
Quy trình 2: Train-Test tách biệt (train-test split)

Train và test trên cùng dataset

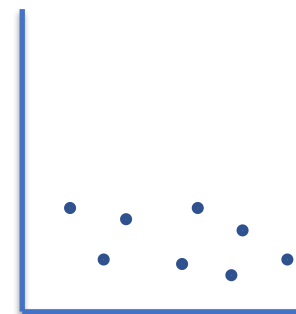
Toàn bộ dataset



Tập train

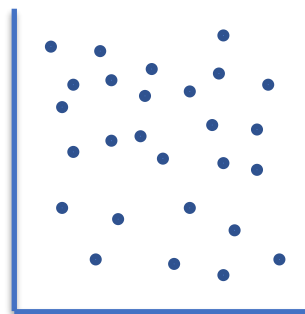


Tập test

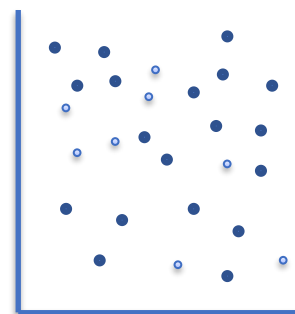


Train và test riêng biệt

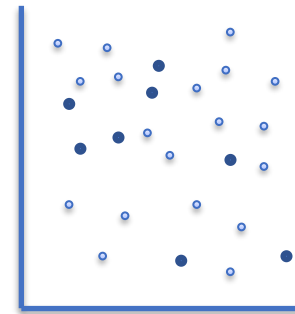
Toàn bộ dataset



Tập train

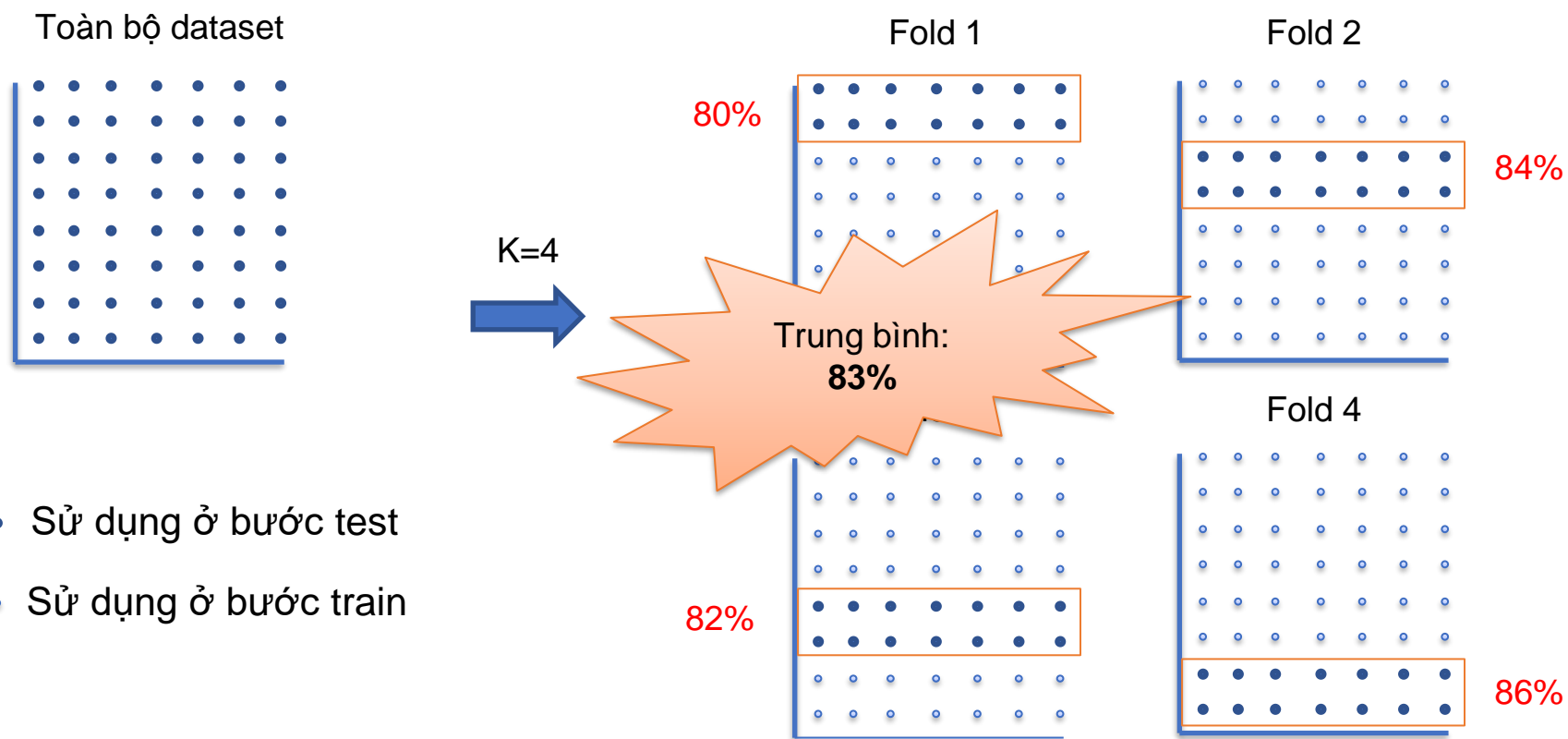


Tập test





Quy trình 3: Kiểm định chéo – K-Fold cross validation



- Sử dụng ở bước test
- Sử dụng ở bước train



So sánh các quy trình đánh giá

	Test là một phần của Train	Train-Test split	K-Fold Cross Validation
Ưu điểm	<ul style="list-style-type: none">- Đơn giản: Không cần thêm quy trình đánh giá phức tạp- Nhanh, do tận dụng chính tập train làm tập test	<ul style="list-style-type: none">- Đơn giản, dễ thực hiện- Ít tốn kém chi phí tính toán so với Cross Validation- Phù hợp khi tập dữ liệu lớn	<ul style="list-style-type: none">- Đánh giá toàn diện về tính tổng quát của mô hình- Phù hợp khi tập dữ liệu ít hoặc huấn luyện nhanh
Khuyết điểm	<ul style="list-style-type: none">- Đánh giá không khách quan- Tính tổng quát của mô hình không được đảm bảo	<ul style="list-style-type: none">- Tốn thêm thời gian- Không khách quan so với cross validation	<ul style="list-style-type: none">- Chi phí tính toán cao do huấn luyện nhiều lần



BÀI QUIZ VÀ HỎI ĐÁP