

EVALUATING CLASSIFICATION MODELS PERFORMANCE

- Nguyễn Hoàng Yến Như
- Nguyễn Trần Phúc Nghi
- Nguyễn Trần Phúc An
- Nguyễn Đức Anh Phúc
- Trịnh Thị Thanh Trúc
- KS. Cao Bá Kiệt
- KS. Quan Chí Khánh An
- KS. Lê Ngọc Huy
- CN. Bùi Cao Doanh
- CN. Nguyễn Trọng Thuận
- KS. Phan Vĩnh Long
- KS. Nguyễn Cường Phát
- ThS. Nguyễn Hoàng Ngân
- KS. Hồ Thái Ngọc
- ThS. Đỗ Văn Tiến
- ThS. Nguyễn Hoàn Mỹ
- ThS. Dương Phi Long
- ThS. Trương Quốc Dũng
- ThS. Nguyễn Thành Hiệp
- ThS. Nguyễn Võ Đăng Khoa
- ThS. Võ Duy Nguyên
- TS. Nguyễn Văn Tâm
- ThS. Trần Việt Thu Phương
- TS. Nguyễn Tấn Trần Minh Khang

Đánh giá mô hình phân lớp

1. Độ chính xác – Accuracy.
2. True/False Positive/Negative
3. Confusion Matrix
4. CAP (Cumulative accuracy profile) curve

ĐỘ CHÍNH XÁC – ACCURACY

Độ chính xác – Accuracy

- Độ chính xác của mô hình trên một tập dữ liệu được tính bằng công thức:

$$Accuracy = \frac{\text{Số lượng điểm dữ liệu dự đoán chính xác}}{\text{Tổng số lượng điểm dữ liệu}}$$

- Độ chính xác là một độ đo thường được sử dụng để đánh giá các mô hình phân loại.
- Độ chính xác càng cao, mô hình càng hiệu quả.

Độ chính xác – Accuracy

— Ví dụ:

- + Cho mô hình M dự đoán trên tập dữ liệu D có 500 điểm dữ liệu.
- + Mô hình M dự đoán đúng 420 điểm dữ liệu.
- + Độ chính xác của mô hình M trên tập dữ liệu D là:

$$Accuracy = \frac{420}{500} = 0.84 = 84\%$$

Độ chính xác – Accuracy

— Nghịch lý của độ chính xác (*Accuracy Paradox*):

- + Đối với các bài toán phân loại mà số lượng điểm dữ liệu trong các lớp **mất cân bằng**, độ đo chính xác không còn hiệu quả để đánh giá mô hình nữa.
- + Nguyên nhân là nếu mô hình của ta **luôn dự đoán ra lớp mà có số lượng điểm dữ liệu lớn nhất**, thì độ chính xác luôn luôn ở mức cao.

Độ chính xác – Accuracy

— Ví dụ:

- + Cho bài toán phân loại boom dưới lòng đất.
- + Ta có 1,000,000 điểm dữ liệu trong tập dữ liệu, tuy nhiên chỉ có khoảng 10,000 điểm dữ liệu chứa dữ liệu có boom.
- + Nếu mô hình **luôn dự đoán là không có boom**, thì độ chính xác lên đến 99%.

— Do đó, độ chính xác không luôn luôn mô tả đúng hiệu quả của một mô hình.

TRUE/FALSE POSITIVE/NEGATIVE

True/False Positive/Negative

- Trong các bài toán phân loại, ta thường gặp những bài toán mà:
 - + Ta cần quan tâm đánh giá một lớp dữ liệu nào đó hơn các lớp dữ liệu còn lại.
 - + Số lượng điểm dữ liệu trong các lớp dữ liệu bị mất cân bằng.
- Ta thường áp dụng các khái niệm *True/False Positive/Negative* để đánh giá mô hình trong các bài toán trên.

True/False Positive/Negative

— Ví dụ 01:

- + Cho bài toán xác định khu vực đang chọn có boom hay không.
- + Việc bỏ sót một khu vực có boom được xem là nghiêm trọng hơn xác định nhầm rất nhiều.
- + Đồng thời, dữ liệu những khu vực có boom cũng thường khó thu thập hơn so với dữ liệu những khu vực không có boom. Nên số lượng điểm dữ liệu có boom thường ít hơn không có boom rất nhiều.

True/False Positive/Negative

— Ví dụ 02:

- + Cho bài toán xác định một email có phải là email rác hay không?
- + Việc cho nhầm email vào thùng rác nghiêm trọng hơn việc xác định một email rác là một email thông thường.
- + Dữ liệu email rác cũng khó thu thập hơn email thông thường.

True/False Positive/Negative

- Đối với các bài toán mà ta xem có một lớp quan trọng hơn các lớp còn lại. Ta có thể:
 - + Xem lớp quan trọng hơn là *Positive*.
 - + Các lớp còn lại gộp chung thành *Negative*.
- Từ đó ta định nghĩa các thuật ngữ sau:

	Dự đoán là Positive	Dự đoán là Negative
Thực sự là Positive	True Positive	False Negative
Thực sự là Negative	False Positive	True Negative

True/False Positive/Negative

— Ví dụ 01:

- + Trong bài toán xác định boom dưới lòng đất.
- + Nếu ta chọn lớp có boom là *Positive* thì:

	Dự đoán có boom	Dự đoán không có boom
Thực sự có boom	True Positive	False Negative
Thực sự không có boom	False Positive	True Negative

True/False Positive/Negative

— Ví dụ 02:

+ Trong bài toán xác định email có phải là rác hay không?

+ Nếu ta chọn lớp “email là rác” là *Positive* thì:

	Dự đoán email là rác	Dự đoán email không là rác
Thực sự là rác	True Positive	False Negative
Thực sự không là rác	False Positive	True Negative

True/False Positive/Negative

— Cho mô hình đang xét đánh giá trên một tập dữ liệu, ta gọi:

+ TP là số lượng điểm dữ liệu thuộc *True Positive*.

+ FP là số lượng điểm dữ liệu thuộc *False Positive*.

+ TN là số lượng điểm dữ liệu thuộc *True Negative*.

+ FN là số lượng điểm dữ liệu thuộc *False Negative*.

True/False Positive/Negative

—Ta định nghĩa các loại tỉ lệ sau:

$$+ \text{True Positive Rate} - TPR = \frac{TP}{TP+FN}$$

$$+ \text{False Positive Rate} - FPR = \frac{FP}{FP+TN}$$

$$+ \text{True Negative Rate} - TNR = \frac{TN}{TN+FP}$$

$$+ \text{False Negative Rate} - FNR = \frac{FN}{FN+TP}$$

True/False Positive/Negative

	Dự đoán là Positive	Dự đoán là Negative
Thực sự là Positive	$TPR = \frac{TP}{TP + FN}$	$FNR = \frac{FN}{TP + FN}$
Thực sự là Negative	$FPR = \frac{FP}{FP + TN}$	$TNR = \frac{TN}{FP + TN}$

True/False Positive/Negative

- Ta thường chỉ quan tâm đến *False Positive Rate* và *False Negative Rate*, ý nghĩa của chúng là:
 - + *False Positive Rate* có ý nghĩa là tỉ lệ xác định nhầm lớp dữ liệu ta quan tâm.
 - + *False Negative Rate* có ý nghĩa là tỉ lệ bỏ sót lớp dữ liệu ta quan tâm.

True/False Positive/Negative

— Ví dụ 01:

- + Trong bài toán xác định boom dưới lòng đất.
- + Giả sử sau khi kiểm tra mô hình trên một tập dữ liệu có 100 điểm dữ liệu, trong đó có 30 điểm dữ liệu thực sự là boom, ta thu được bảng sau:

	Dự đoán có boom	Dự đoán không có boom
Thực sự có boom	$TP = 13$	$FN = 17$
Thực sự không có boom	$FP = 20$	$TN = 50$

True/False Positive/Negative

— Ví dụ 01:

	Dự đoán có boom	Dự đoán không có boom
Thực sự có boom	$TP = 13$	$FN = 17$
Thực sự không có boom	$FP = 8$	$TN = 62$

+ Tỷ lệ xác định nhầm là $FPR = \frac{FP}{FP+TN} = \frac{8}{8+62} = \frac{8}{70} \approx 0.11$

+ Tỷ lệ bỏ sót là $FNR = \frac{FN}{FN+TP} = \frac{17}{17+13} = \frac{17}{30} \approx 0.57$

True/False Positive/Negative

— Ví dụ 02:

- + Trong bài toán xác định email có phải là rác hay không?
- + Giả sử kiểm tra mô hình với 100 email, trong đó có 40 email là rác, ta thu được bảng sau:

	Dự đoán email là rác	Dự đoán email không là rác
Email rác	$TP = 36$	$FN = 4$
Email bình thường	$FP = 40$	$TN = 20$

True/False Positive/Negative

— Ví dụ 02:

	Dự đoán email là rác	Dự đoán email không là rác
Email rác	$TP = 36$	$FN = 4$
Email bình thường	$FP = 40$	$TN = 20$

+ Tỷ lệ xác định nhầm là $FPR = \frac{FP}{FP+TN} = \frac{40}{40+20} = \frac{40}{60} \approx 0.67$

+ Tỷ lệ bỏ sót là $FNR = \frac{FN}{FN+TP} = \frac{4}{4+36} = \frac{4}{40} = 0.1$

CONFUSION MATRIX

Confusion Matrix

- *Confusion matrix* là một ma trận thường dùng để đánh giá độ hiệu quả của một mô hình.
- *Confusion matrix* của bài toán phân loại C lớp là:
 - + Một ma trận vuông có kích thước $C \times C$.
 - + Ô ở hàng thứ i , cột thứ j của ma trận có giá trị là số lượng điểm dữ liệu có lớp thật sự i và được phân loại vào lớp j .

Confusion Matrix

— Tính chất:

- + Tổng giá trị của các ô trong ma trận là tổng số lượng điểm dữ liệu của tập dữ liệu đang đánh giá.
- + Tổng giá trị trên đường chéo chính là tổng số điểm dữ liệu được phân loại chính xác.

Confusion Matrix

- *Confusion matrix* là trường hợp tổng quát của *True/False Positive/Negative*.
- Với bài toán phân loại 2 lớp, đặt lớp 1 là *Positive*, ta có:

	Dự đoán là 0 (Negative)	Dự đoán là 1 (Positive)
Thực sự là 0 (Negative)	True Negative	False Positive
Thực sự là 1 (Positive)	False Negative	True Positive

Confusion Matrix

- Để đánh giá độ hiệu quả của một mô hình bằng Confusion Matrix, ta thường:
 - + **Xây dựng các tỉ lệ như độ chính xác, độ lỗi, ...**
 - + Chuyển Confusion Matrix về dạng True/False Positive/Negative. Sau đó thực hiện đánh giá trên từng lớp dữ liệu.

Confusion Matrix

— Ví dụ 01:

- + Cho bài toán xác định boom dưới lòng đất.
- + Đặt 0 đại diện cho lớp không có boom, 1 cho lớp có boom.
- + Thực hiện dự đoán trên tập dữ liệu gồm 100 điểm dữ liệu, trong đó có 30 điểm dữ liệu có boom.
- + Ta thu được *confusion matrix* sau:

	0 (không có boom)	1 (có boom)
0 (không có boom)	50	20
1 (có boom)	17	13

Confusion Matrix

— Ví dụ 01:

	0 (không có boom)	1 (có boom)
0 (không có boom)	50	20
1 (có boom)	17	13

- + Số lượng điểm dữ liệu thực sự có boom và được dự đoán không có boom là 17.
- + Số lượng điểm dữ liệu thực sự không có boom và được dự đoán có boom là 20.

Confusion Matrix

— Ví dụ 01:

	0 (không có boom)	1 (có boom)
0 (không có boom)	50	20
1 (có boom)	17	13

+ Số lượng điểm dữ liệu phân loại đúng là $50 + 13 = 63$.

+ Tỷ lệ dự đoán đúng là $63/100 = 0.63$

Confusion Matrix

— Ví dụ 02:

- + Cho bài toán phân loại hình ảnh gà, chó, khỉ.
- + Đặt 0 đại diện cho lớp gà, 1 cho lớp chó và 2 cho lớp khỉ.
- + Thực hiện dự đoán trên tập dữ liệu gồm 100 điểm dữ liệu.
- + Ta thu được *confusion matrix* sau:

	0 (gà)	1 (chó)	2 (khỉ)
0 (gà)	25	2	3
1 (chó)	3	33	4
2 (khỉ)	6	8	16

Confusion Matrix

— Ví dụ 02:

	0 (gà)	1 (chó)	2 (khỉ)
0 (gà)	25	2	3
1 (chó)	3	33	4
2 (khỉ)	6	8	16

- + Lớp bị phân loại sai ít nhất là 0 (gà) với $2 + 3 = 5$ điểm dữ liệu bị phân loại sai.
- + Lớp bị phân loại sai nhiều nhất là 2 (khỉ) với $6 + 8 = 14$ điểm dữ liệu bị phân loại sai.

Confusion Matrix

— Ví dụ 02:

	0 (gà)	1 (chó)	2 (khỉ)
0 (gà)	25	2	3
1 (chó)	3	33	4
2 (khỉ)	6	8	16

+ Số lượng điểm dữ liệu phân loại sai là $100 - 25 - 33 - 16 = 26$

+ Tỷ lệ phân loại sai là $26/100 = 0.26$

Confusion Matrix

— Ví dụ 03:

- + Cho bài toán phân loại hình ảnh gà, chó, khỉ.
- + Đặt 0 đại diện cho lớp gà, 1 cho lớp chó và 2 cho lớp khỉ.
- + Thực hiện dự đoán trên tập dữ liệu gồm 100 điểm dữ liệu.
- + Ta thu được *confusion matrix* sau:

	0 (gà)	1 (chó)	2 (khỉ)
0 (gà)	25	2	3
1 (chó)	3	33	4
2 (khỉ)	6	8	16

Confusion Matrix

— Ví dụ 03:

+ Ta thực hiện đánh giá lớp dữ liệu gà. Gán lớp gà là Positive, các lớp còn lại được gộp chung thành Negative.

	Dự đoán là Positive (gà)	Dự đoán Negative (chó + khỉ)	
Thực sự là Positive (gà)	25	2	3
Thực sự là Negative (chó + khỉ)	3	33	4
	6	8	16

Confusion Matrix

— Ví dụ 03:

+ Ta thực hiện đánh giá lớp dữ liệu gà. Gán lớp gà là Positive, các lớp còn lại được gộp chung thành Negative.

	Dự đoán là Positive (gà)	Dự đoán Negative (chó + khỉ)
Thực sự là Positive (gà)	25(<i>TP</i>)	5(<i>FN</i>)
Thực sự là Negative (chó + khỉ)	9(<i>FP</i>)	61(<i>TN</i>)

Confusion Matrix

— Ví dụ 03:

	Dự đoán là Positive (gà)	Dự đoán Negative (chó + khỉ)
Thực sự là Positive (gà)	25(<i>TP</i>)	5(<i>FN</i>)
Thực sự là Negative (chó + khỉ)	9(<i>FP</i>)	61(<i>TN</i>)

+ Tỷ lệ xác định nhầm là $FPR = \frac{FP}{FP+TN} = \frac{9}{9+61} = \frac{9}{70} \approx 0.13$

Confusion Matrix

— Ví dụ 03:

	Dự đoán là Positive (gà)	Dự đoán Negative (chó + khỉ)
Thực sự là Positive (gà)	25(<i>TP</i>)	5(<i>FN</i>)
Thực sự là Negative (chó + khỉ)	9(<i>FP</i>)	61(<i>TN</i>)

+ Tỷ lệ bỏ sót là $FNR = \frac{FN}{FN+TP} = \frac{5}{5+25} = \frac{5}{30} = 0.167$

CAP CURVE

CAP Curve

- *CAP (Cumulative accuracy profile) curve* là một đường cong được vẽ khi dự đoán mô hình trên một tập dữ liệu.
- *CAP curve* là một cách đánh giá thường được sử dụng trong các bài toán phân loại.
- *CAP curve* thường dùng để đánh giá trên một lớp nhất định.

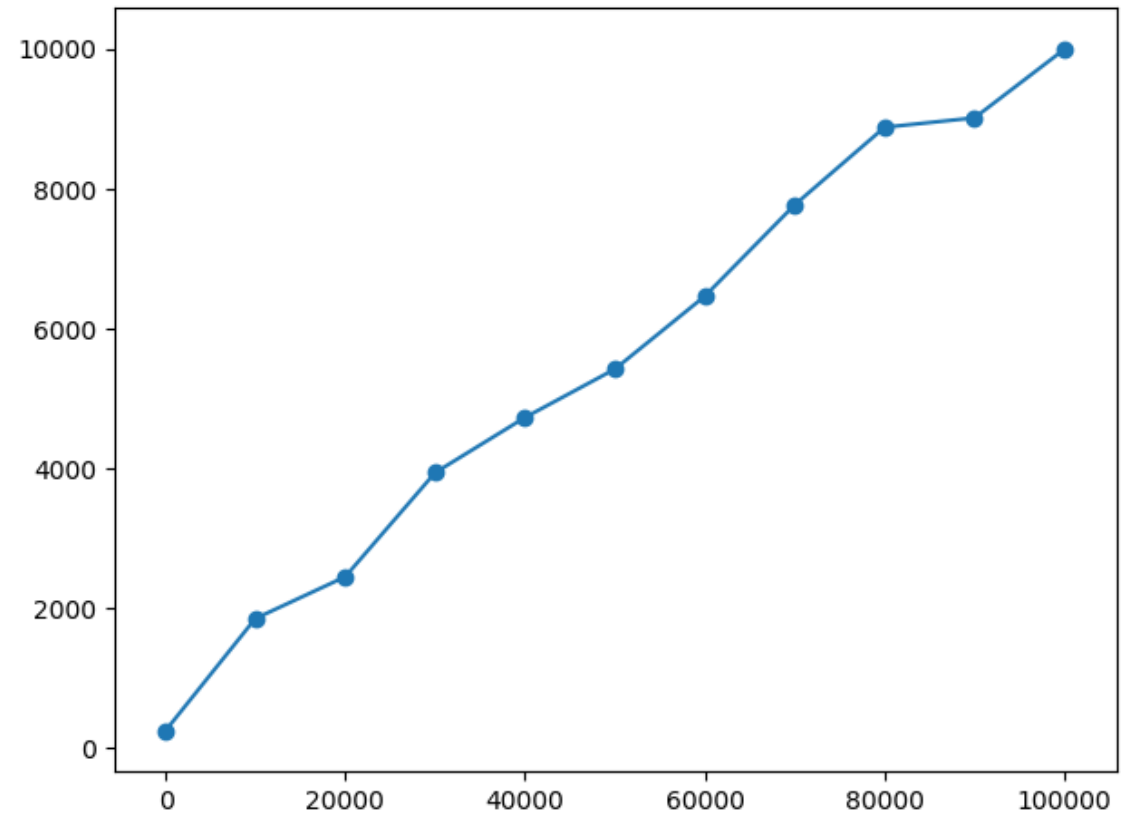
CAP Curve

- *CAP curve* biểu diễn đồ thị trong không gian 2 chiều Oxy , trong đó:
 - + Trục hoành (x) tương ứng với số lượng điểm dữ liệu được chọn bằng một cách nào đó trong tập dữ liệu.
 - + Trục tung (y) là số lượng điểm dữ liệu thuộc về lớp dữ liệu ta quan tâm trong x điểm dữ liệu đã chọn ra.

CAP Curve

— Ví dụ 01:

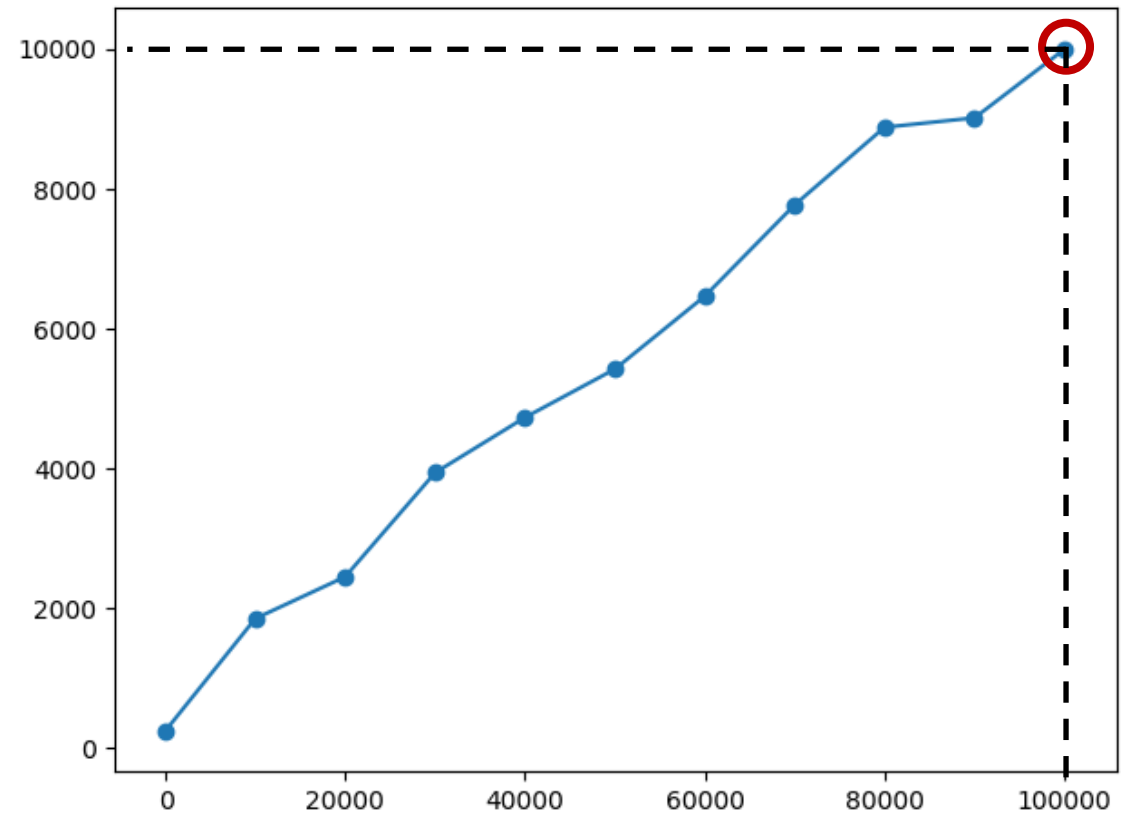
- + Trong bài toán phân loại boom, giả sử ta đang quan tâm về lớp có boom.
- + Xét tập dữ liệu D có 100,000 điểm dữ liệu. Trong đó có 10,000 điểm dữ liệu thuộc lớp có boom. Mô hình của chúng ta cho một *CAP curve* như sau.



CAP Curve

— Ví dụ 01:

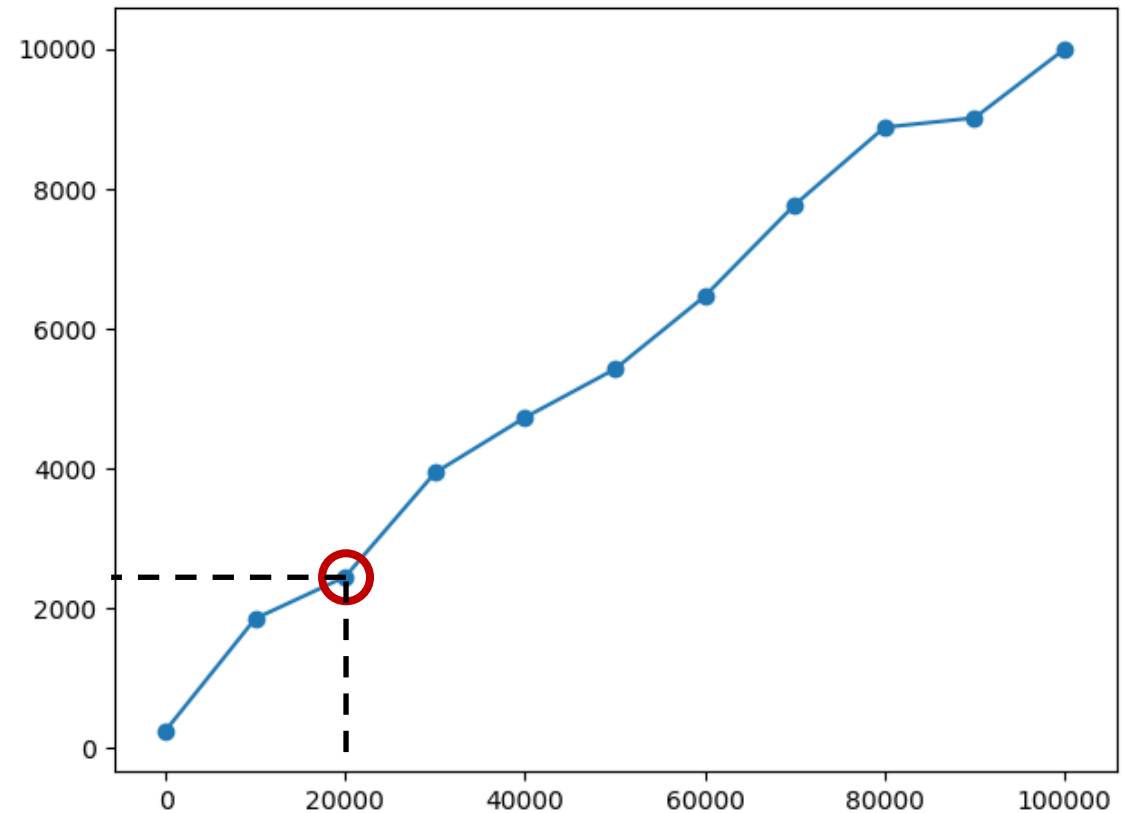
+ Điểm trên cùng có nghĩa là nếu ta chọn 100,000 điểm dữ liệu trong tập dữ liệu (tức toàn bộ tập dữ liệu), ta nhận được tương ứng 10,000 điểm dữ liệu có boom.



CAP Curve

— Ví dụ 01:

+ Điểm dữ liệu thứ 3 (tương ứng với $x = 20,000$) có ý nghĩa là nếu ta được phép chọn 20,000 điểm dữ liệu trong số 100,000 điểm, ta thu được (khoảng) 2,400 điểm dữ liệu có boom.



CAP Curve

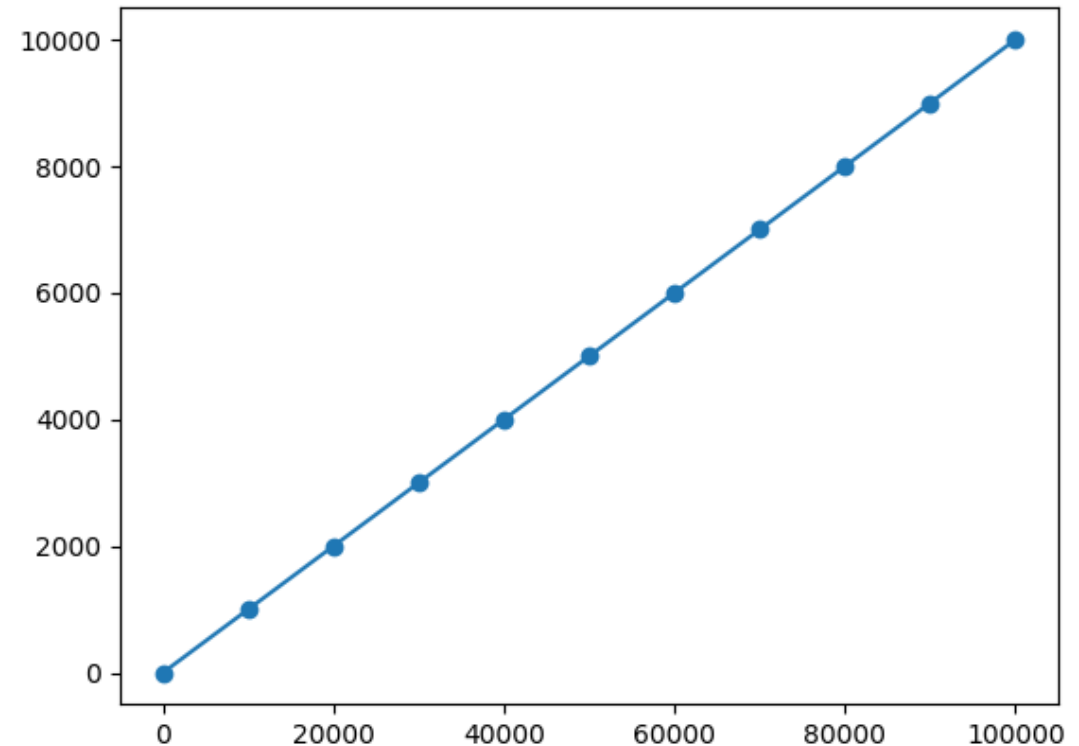
— Định nghĩa *Random CAP Curve*:

- + *Random CAP Curve* được xây dựng bằng cách chọn các điểm dữ liệu hoàn toàn ngẫu nhiên.
- + Nếu số lượng điểm dữ liệu đủ lớn, đồ thị *Random CAP Curve* là một đường thẳng.

CAP Curve

— Ví dụ 02:

- + Trong bài toán phân loại boom, giả sử ta đang quan tâm về lớp có boom.
- + Xét tập dữ liệu D có 100,000 điểm dữ liệu. Trong đó có 10,000 điểm dữ liệu thuộc lớp có boom.
- + *Random CAP curve* có dạng như sau.

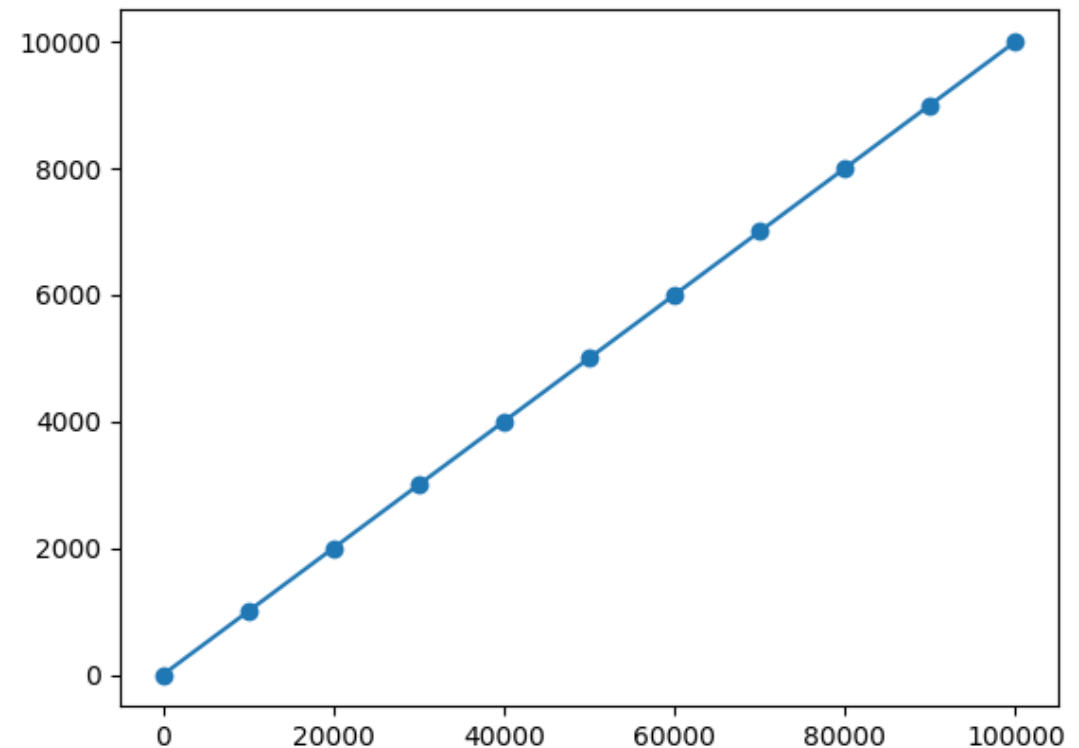


CAP Curve

— Ví dụ 02:

+ Tần suất xuất hiện điểm dữ liệu thuộc lớp có boom trong tập dữ liệu này là $\frac{10,000}{100,000} = 10\%$.

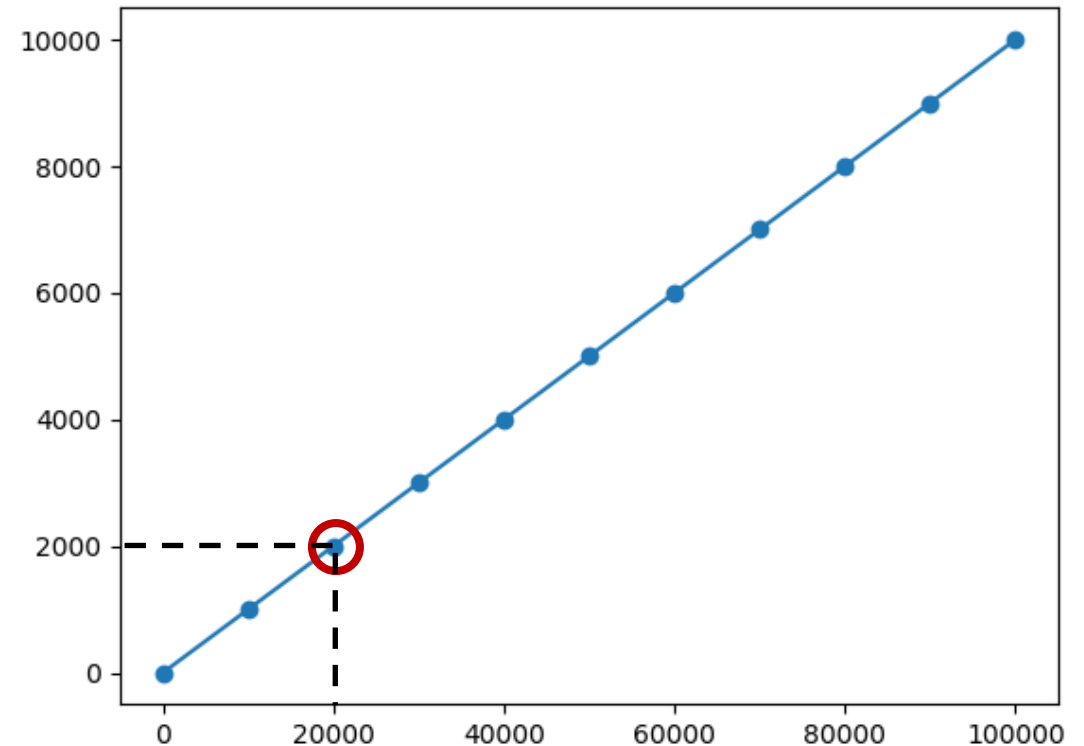
+ Tức nếu chọn ngẫu nhiên một số lượng bất kỳ các điểm dữ liệu, sẽ có khoảng 10% điểm dữ liệu thuộc lớp có boom.



CAP Curve

— Ví dụ 02:

+ Nếu chọn ngẫu nhiên 20,000 điểm dữ liệu trong tập dữ liệu, thì có khoảng 2,000 điểm dữ liệu có boom.



CAP Curve

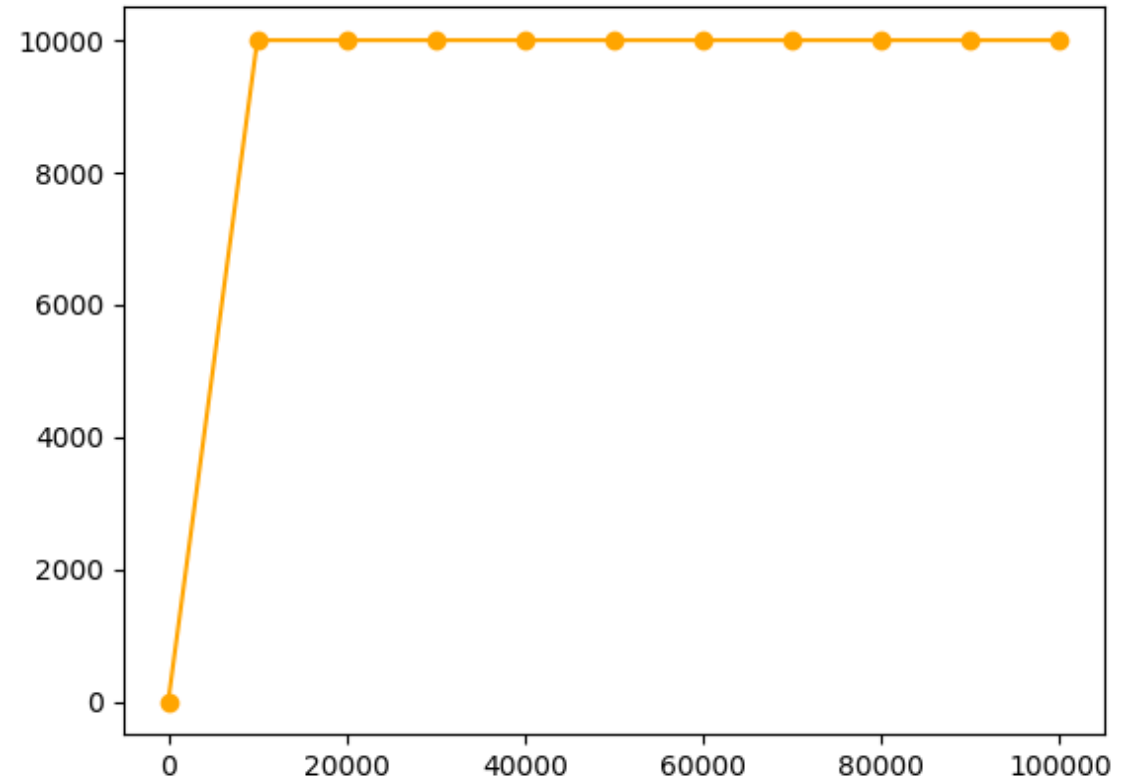
— Định nghĩa *Perfect CAP Curve*:

+ *Perfect CAP Curve* được xây dựng bằng cách ưu tiên chọn các điểm thuộc lớp ta quan tâm.

CAP Curve

— Ví dụ 03:

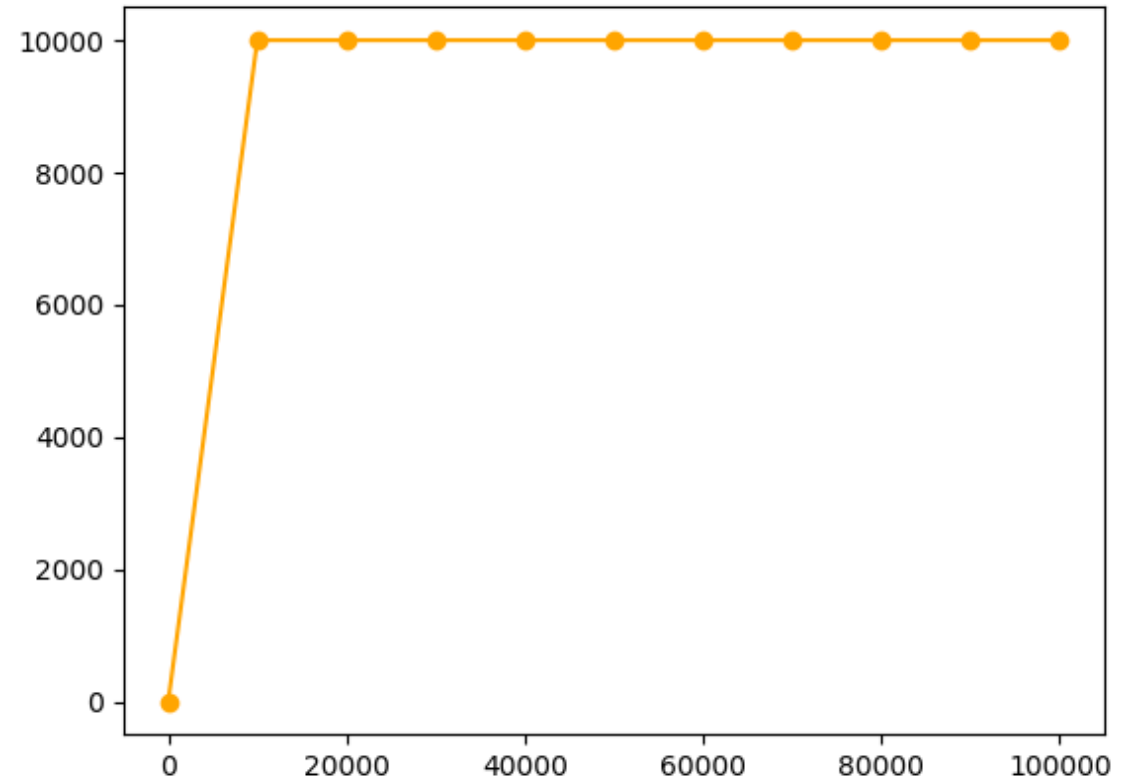
- + Trong bài toán phân loại boom, giả sử ta đang quan tâm về lớp có boom.
- + Xét tập dữ liệu D có 100,000 điểm dữ liệu. Trong đó có 10,000 điểm dữ liệu thuộc lớp có boom.
- + *Perfect CAP curve* có dạng như sau.



CAP Curve

— Ví dụ 03:

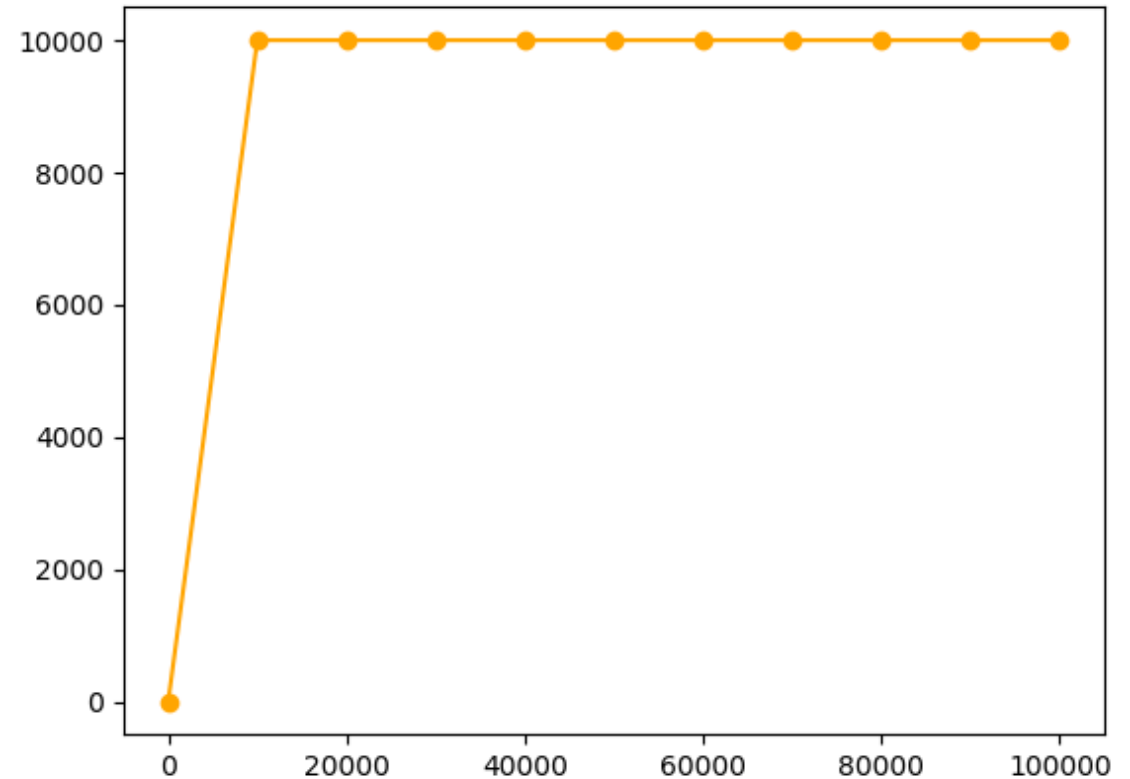
+ Khi ưu tiên chọn các điểm dữ liệu thuộc lớp có boom, nếu số lượng điểm dữ liệu cần chọn nhỏ hơn hoặc bằng 10,000, số lượng điểm có boom vừa đúng bằng với số lượng điểm dữ liệu mà ta chọn.



CAP Curve

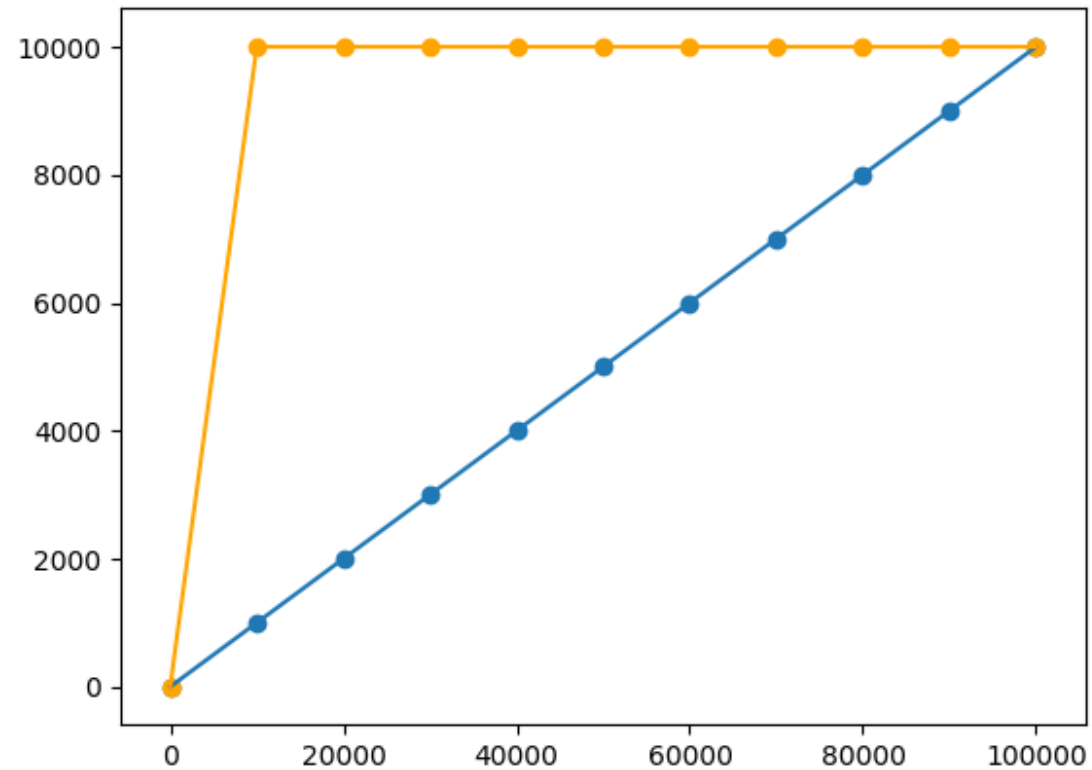
— Ví dụ 03:

+ Nếu số lượng điểm dữ liệu cần chọn lớn hơn 10,000, số lượng điểm dữ liệu có boom sẽ không tăng thêm, vì trong tập dữ liệu này, chỉ có tối đa 10,000 điểm dữ liệu có boom.



CAP Curve

— *Random CAP Curve* và *Perfect CAP Curve*.



CAP Curve

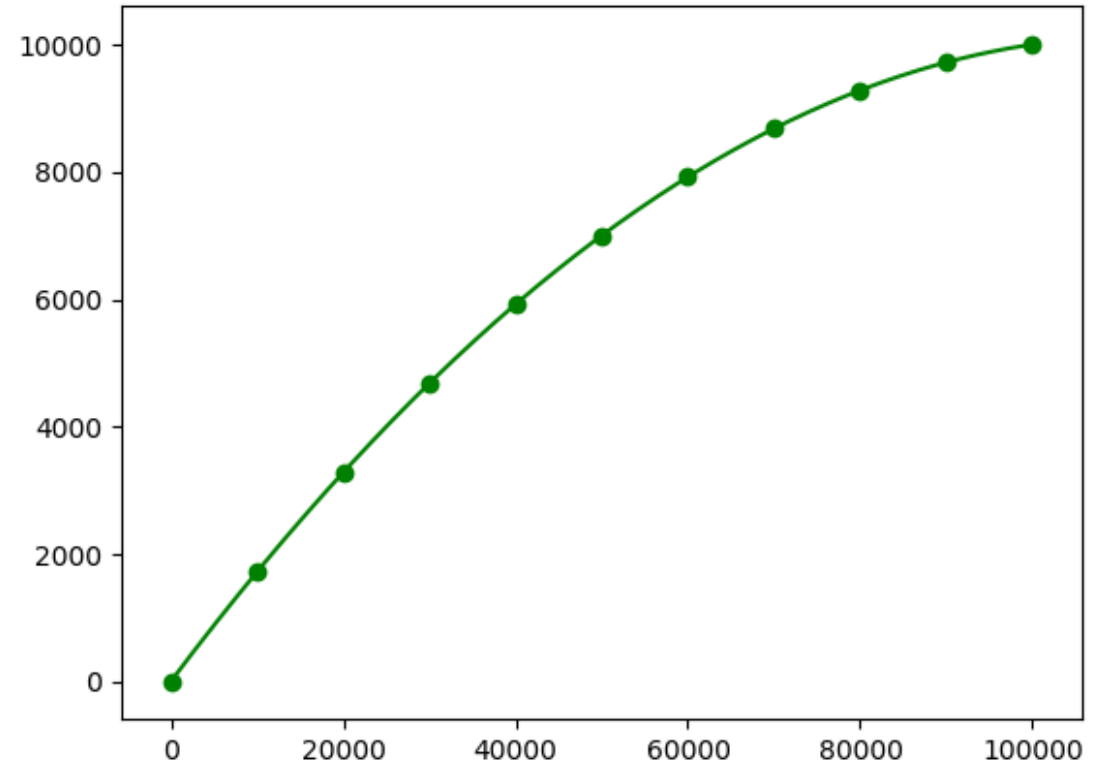
— Định nghĩa *Model CAP Curve*:

+ *Model CAP Curve* được xây dựng bằng cách ưu tiên chọn các điểm dữ liệu mà **mô hình dự đoán** là thuộc về lớp ta quan tâm.

CAP Curve

— Ví dụ 04:

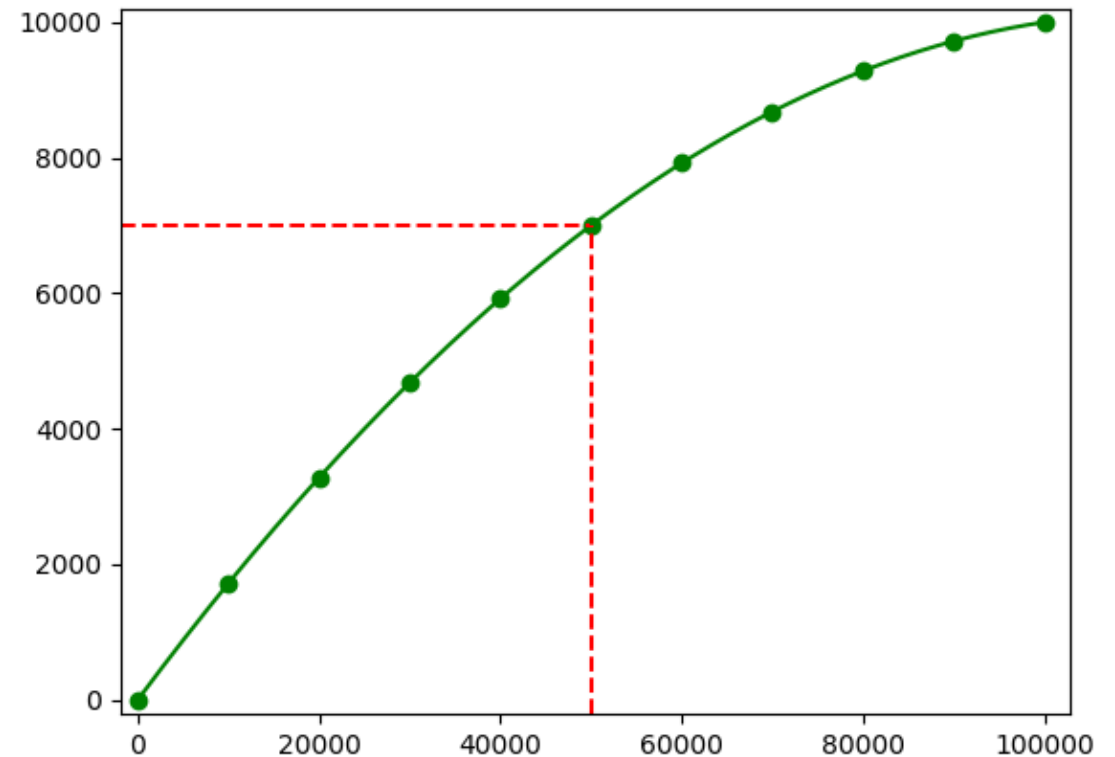
- + Trong bài toán phân loại boom, giả sử ta đang quan tâm về lớp có boom.
- + Xét tập dữ liệu D có 100,000 điểm dữ liệu. Trong đó có 10,000 điểm dữ liệu thuộc lớp có boom.
- + Mô hình của chúng ta cho một *CAP curve* như sau.



CAP Curve

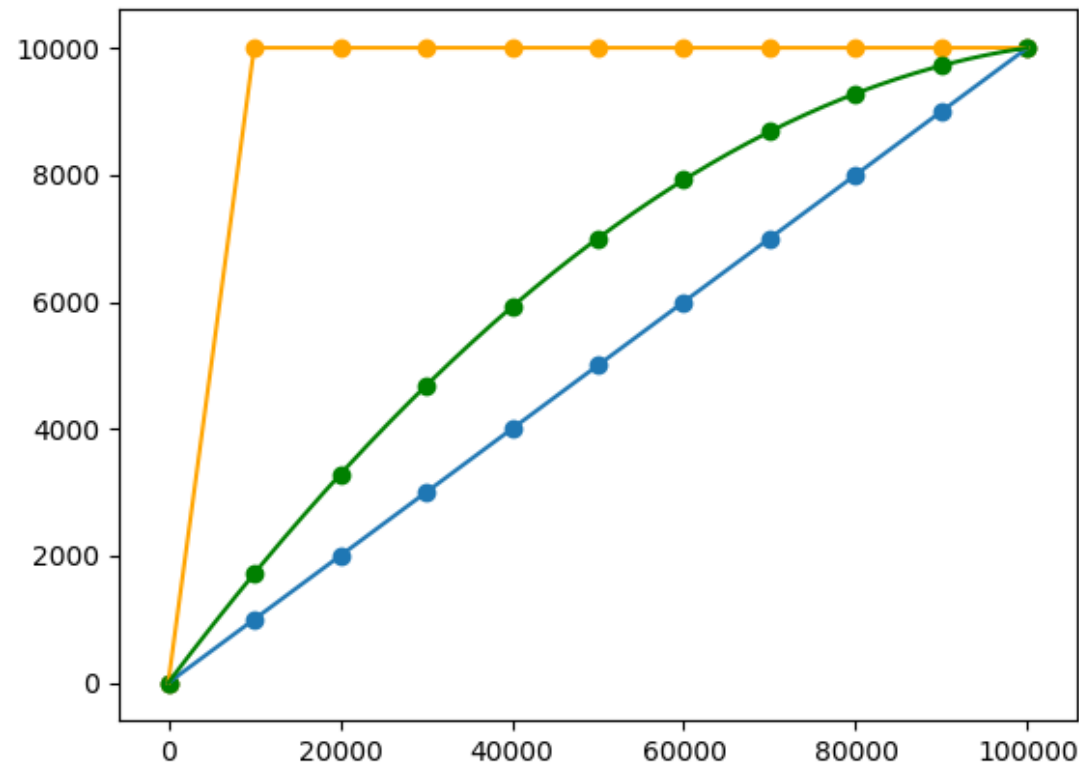
— Ví dụ 04:

+ Tại vị trí $x = 50,000$, tức khi chọn 50,000 điểm dữ liệu, trong đó ta ưu tiên chọn các điểm dữ liệu mà mô hình dự đoán là có boom, ta đã thu được 7,000 điểm dữ liệu thực sự có boom.



CAP Curve

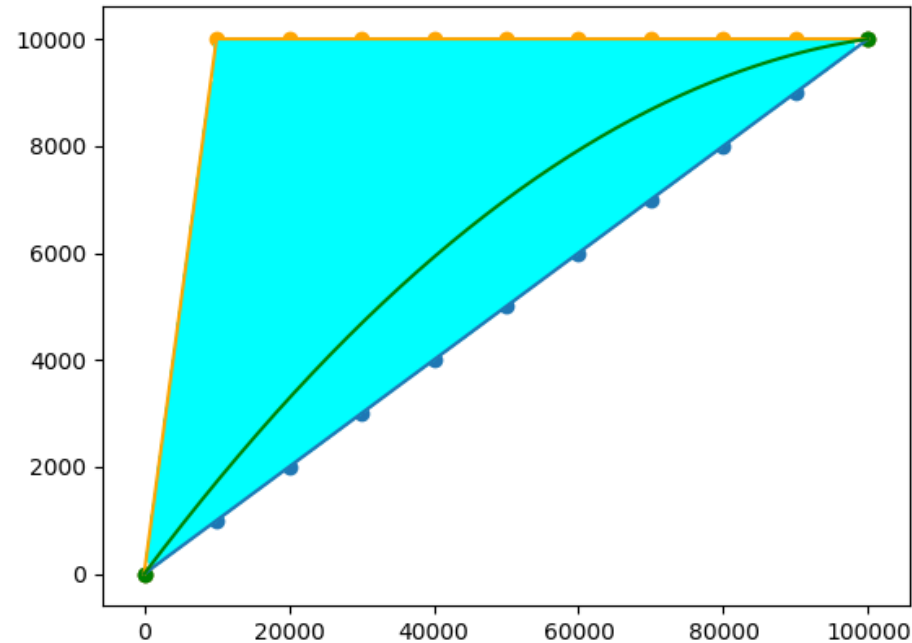
— *Random CAP Curve, Perfect CAP Curve và Model CAP Curve.*



CAP Curve

— Đánh giá mô hình dựa vào *CAP Curve*:

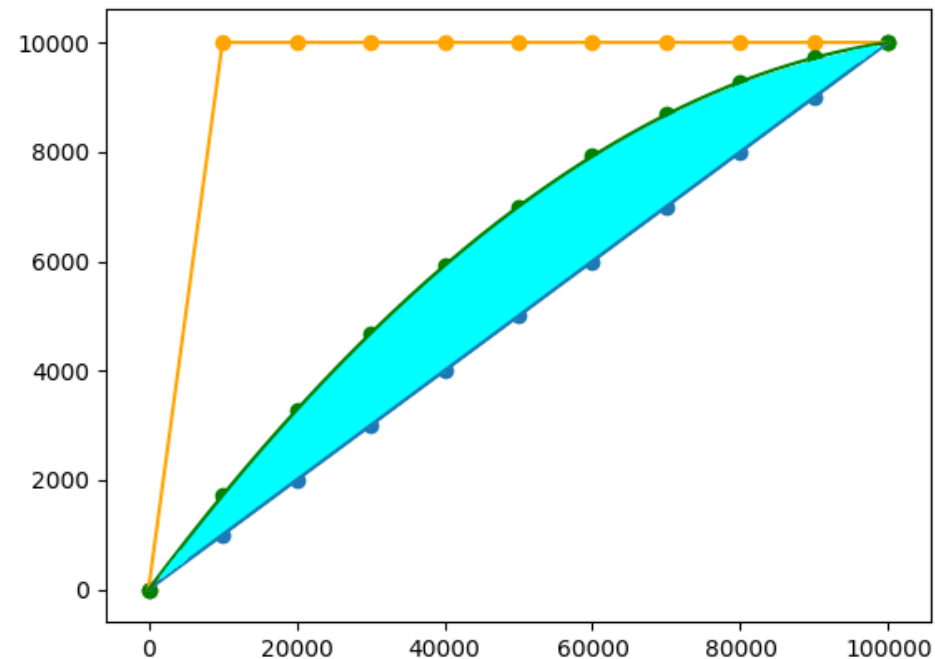
+ Đặt a_P là diện tích giữa hai đường *Perfect CAP Curve* và *Random CAP Curve*.



CAP Curve

— Đánh giá mô hình dựa vào *CAP Curve*:

+ Đặt a_R là diện tích giữa hai đường *Model CAP Curve* và *Random CAP Curve*.



CAP Curve

— Đánh giá mô hình dựa vào *CAP Curve*:

+ Đặt $AR = \frac{a_R}{a_P}$.

+ Khi đó, AR có giá trị trong đoạn $[0,1]$.

+ Giá trị AR càng lớn, độ hiệu quả của mô hình càng cao.

CAP Curve

- Tuy nhiên, việc tính diện tích của các đường cong trên là rất khó khăn.
- Chúng ta có một cách đánh giá khác tốt hơn mà vẫn dựa trên *CAP Curve*.

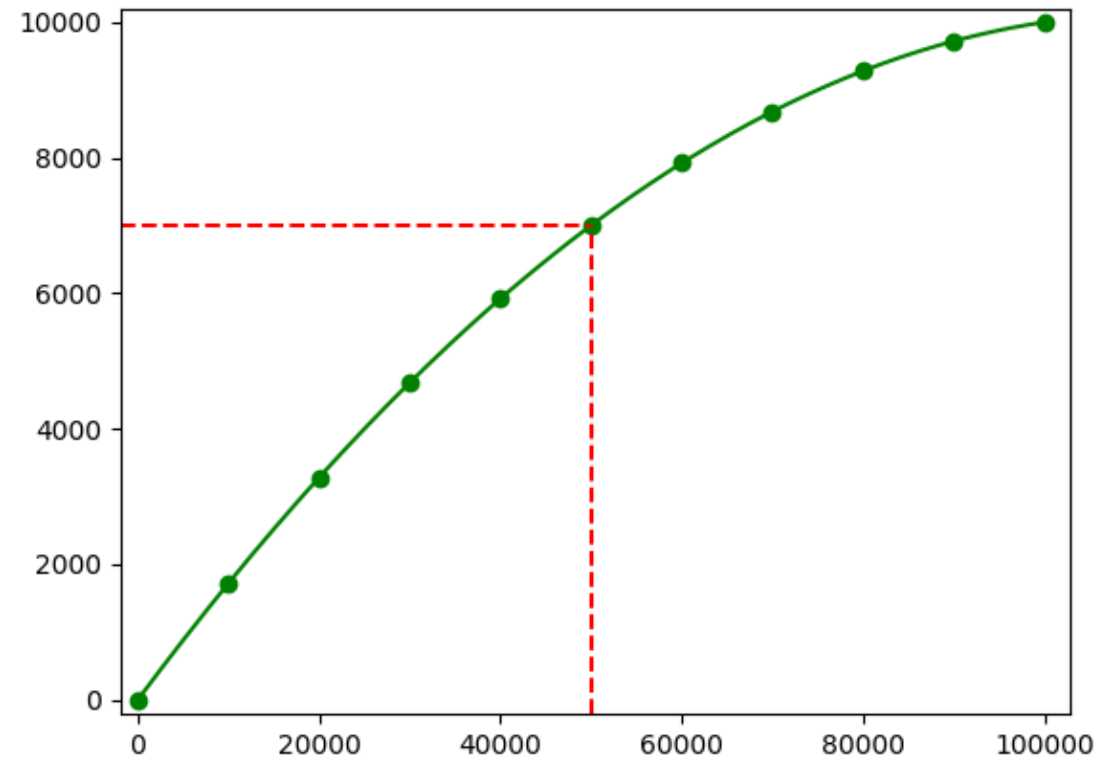
CAP Curve

- Nhìn vào vị trí chính giữa của trục hoành, tức x bằng 50% giá trị tối đa. Hay nói cách khác là chọn một nửa số lượng điểm dữ liệu trong tập dữ liệu.
- Kiểm tra xem giá trị tương ứng của y và tính toán nó đạt tỉ lệ bao nhiêu so với giá trị y tối đa.

CAP Curve

— Ví dụ 05:

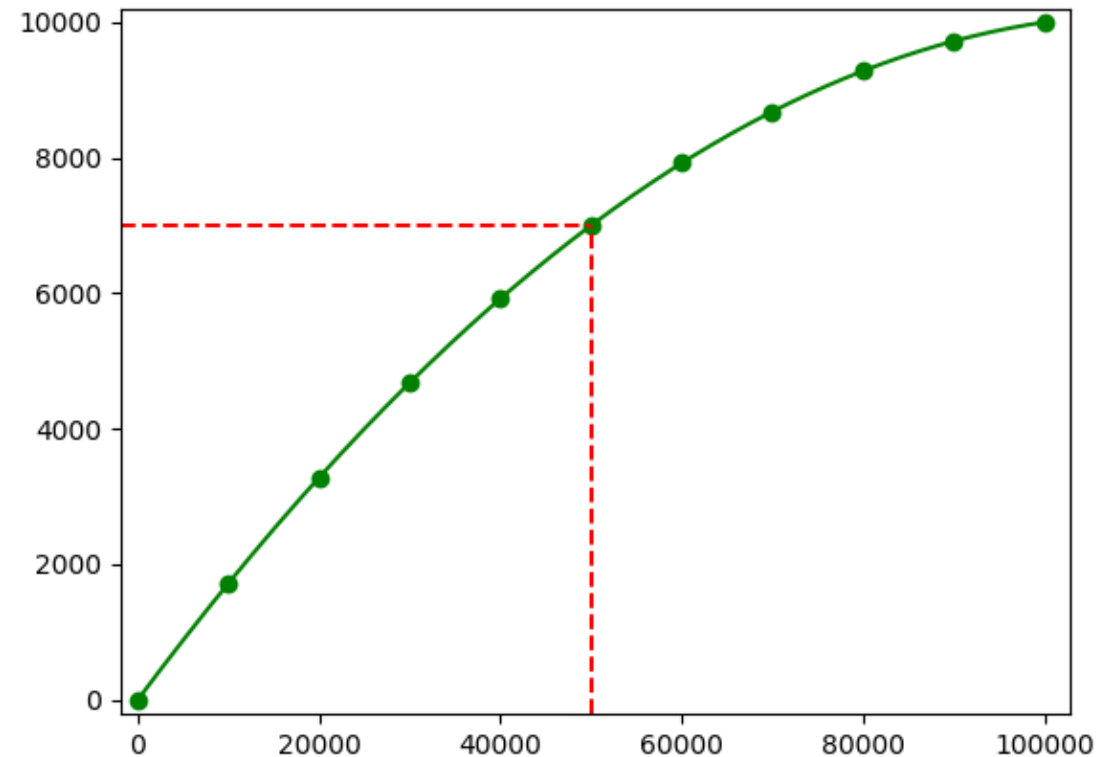
- + Trong bài toán phân loại boom, giả sử ta đang quan tâm về lớp có boom.
- + Xét tập dữ liệu D có 100,000 điểm dữ liệu. Trong đó có 10,000 điểm dữ liệu thuộc lớp có boom.



CAP Curve

— Ví dụ 05:

+ Theo hình bên, tại vị trí $x = 50,000$ thì $y = 7,000$. Tức y đạt 70% giá trị tối đa của nó.



CAP Curve

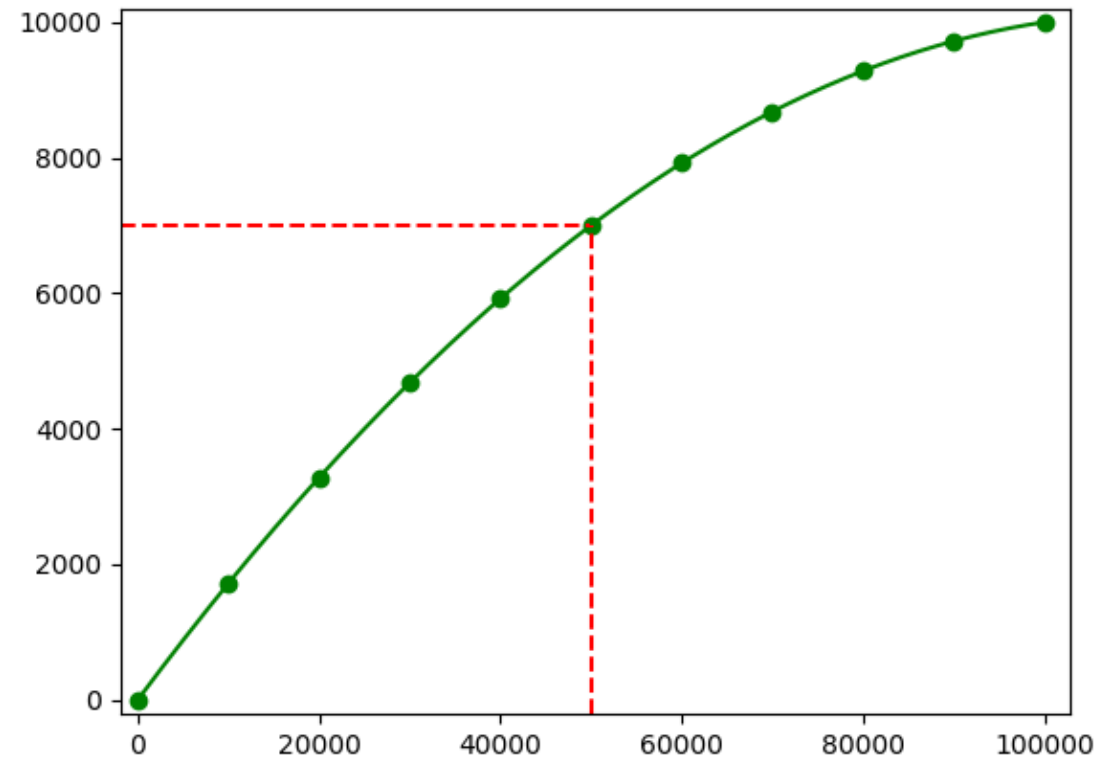
- Dựa trên tỉ lệ của y so với giá trị tối đa của nó, khi chọn vị trí x bằng một nửa số lượng điểm dữ liệu trong tập dữ liệu, ta xây dựng bảng đánh giá sau:

Tỉ lệ của y	Độ hiệu quả
90% – 100%	Tuyệt vời
80% – 90%	Rất tốt
70% – 80%	Tốt
60% – 70%	Tệ
< 60%	Rất tệ

CAP Curve

— Ví dụ 05:

+ Theo bảng trên, độ hiệu quả của mô hình này ở mức tốt (tương ứng với y đạt 70%).



Chúc các bạn học tốt
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN TP.HCM

Nhóm UIT-Together
Nguyễn Tấn Trần Minh Khang