Logistic Regression

Ngoc Hoang Luong

University of Information Technology (UIT), VNU-HCM

May 22, 2023

Hồi quy logistic

Lư ơ ng Ngọc Hoàng

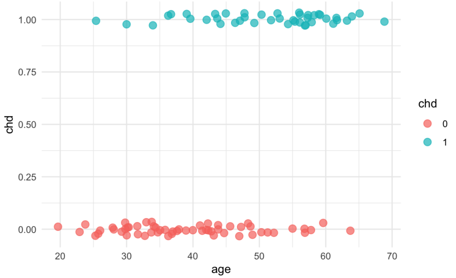
Trư ờng Đại học Công nghệ Thông tin (UIT), ĐHQG-HCM

22 Tháng Năm, 2023

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT

Motivation

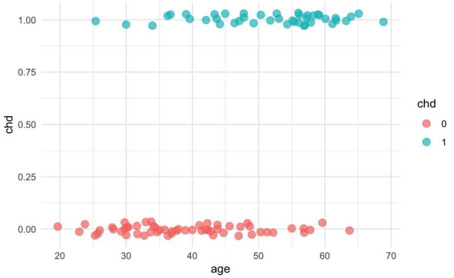
- Example: Predicting heart attack (coronary heart disease (chd)) with only one predictor: age.



- For age, we have values ranging from small x's to large x's.
- For the binary response, there are only 0's and 1's (for better visualization, there are some jitter).
- For $y_i = 0$, there are more small values x than large ones. For $y_i = 1$, there are more large values x than small ones.

Động lực

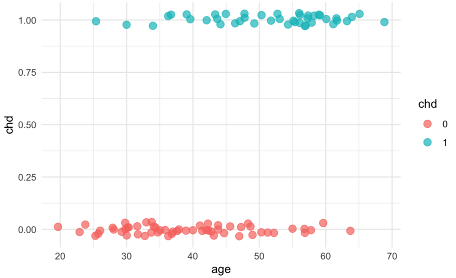
- Ví dụ: Dự đoán cơn đau tim (bệnh mạch vành (chd)) chỉ với một yếu tố dự đoán: tuổi.



- Đối với tuổi, chúng tôi có các giá trị từ x nhỏ đến x lớn.
- Đối với phản hồi nhị phân, chỉ có 0 và 1 (để tốt hơn trực quan hóa, có một số jitter).
- Với yi = 0, có nhiều giá trị x nhỏ hơn giá trị lớn. Với yi = 1, có nhiều giá trị x lớn hơn giá trị nhỏ.

Motivation

- Example: Predicting heart attack (coronary heart disease (chd)) with only one predictor: age.

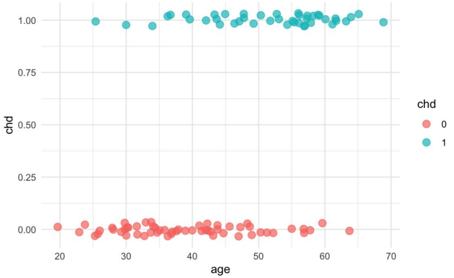


- The goal is to fit a model that predicts chd from age.

$$\text{chd} = f(\text{age}) + \varepsilon$$

Động lực

- Ví dụ: Dự đoán cơ n đau tim (bệnh mạch vành (chd)) chỉ với một yếu tố dự đoán: tuổi.



- Mục tiêu là để phù hợp với một mô hình dự đoán chd theo độ tuổi.

chd = f(tuổi) + ε

First Approach: Fitting a Line

- We could try to fit a linear model:

$$\hat{\mathbf{y}} = b_0 + b_1\mathbf{x} = \mathbf{X}\mathbf{y}$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

Cách tiếp cận đầu tiên: Lắp một dòng

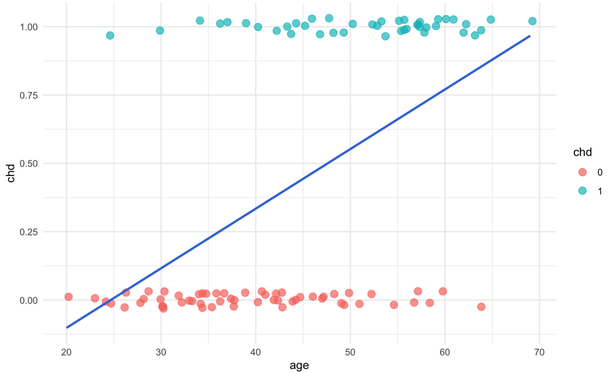
- Chúng ta có thể thử điều chỉnh một mô hình tuyến tính:

y^ = b0 + b1x = Xy

Ở đâu

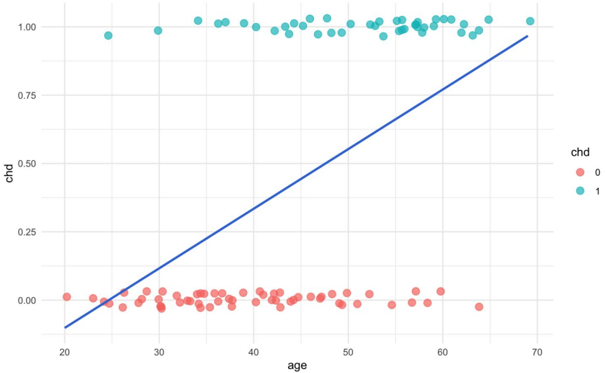
X = 1x1 _
1x2 _
; b = (b01)
1 lần

First Approach: Fitting a Line



- There seems to be some positive relation between age and `chd`.
- The regression line extends beyond the range [0,1], which does not make sense.

Cách tiếp cận đầu tiên: Lắp một dòng



- Dư ờng như có một số mối quan hệ tích cực giữa tuổi tác và chd.
- Đư ờng hồi quy mở rộng ra ngoài phạm vi [0,1], không có lý.

Second Approach: Harsh Thresholding

- We can set some threshold c and compare $\hat{y}_i = b_0 + b_1 x_i$ to c.

$$\hat{y}_i = \begin{cases} 1 & \text{if} \quad b_0 + b_1 x_i \geq c \\ 0 & \text{if} \quad b_0 + b_1 x_i < c \end{cases}$$

- We can arrange the terms to get:

$$\begin{aligned} \hat{y}_i = \hat{f}(x_i) = b_0 + b_1 x_i - c &= (b_0 - c) + b_1 x_i \\ &= b_0' + b_1 x_i = \mathbf{b}^\top \mathbf{x}_i \end{aligned}$$

- By paying attention to the sign of the signal $\mathbf{b}^\top \mathbf{x}$, we can transform our fitted model into:

$$\hat{y} = \texttt{sign}(\mathbf{b}^\top \mathbf{x})$$

Cách tiếp cận thứ hai: Ngư ỡng khắc nghiệt

- Chúng ta có thể đặt một số ngư ỡng c và so sánh yˆi = b0 + b1xi với c.

1 nếu b0 + b1xi ≥ c

yˆi = 0 nếu b0 + b1xi < c

- Ta có thể sắp xếp các điều kiện để có đư ợc:

yˆi = ˆf(xi) = b0 + b1xi c = (b0 c) + b1xi + b1xi

= b $'_0$ = b xi

- Bằng cách chú ý đến dấu của tín hiệu b x, chúng ta có thể

chuyển đổi mô hình đư ợc trang bị của chúng tôi thành:

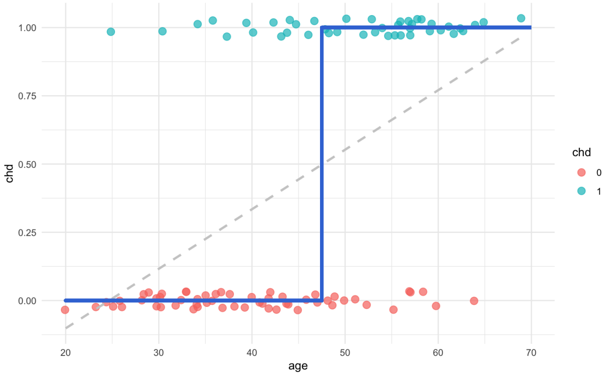
yˆ = dấu(b x)

Second Approach: Harsh Thresholding

- We have a classification rule:

$$\hat{y}_i = \begin{cases} 1 & \text{if} \quad \mathtt{sign}(b_0' + b_1 x_i) \geq 0 \\ 0 & \text{if} \quad \mathtt{sign}(b_0' + b_1 x_i) < 0 \end{cases}$$

- The signal is still linear but we apply a non-linear transformation to it: $\phi(x) = \mathtt{sign}(b_0' + b_1 x_i)$



Cách tiếp cận thứ hai: Ngư ỡng khắc nghiệt

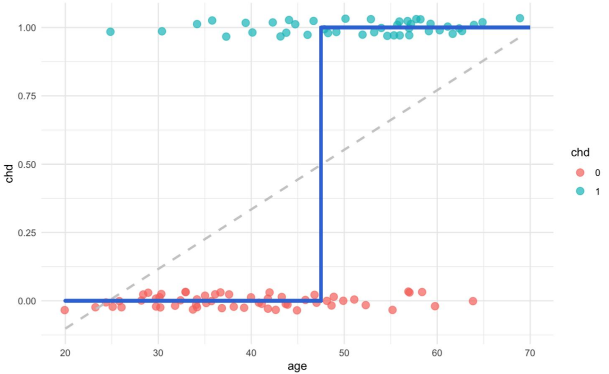
- Ta có quy tắc phân loại:

y^i

1 nếu dấu(b ′ 0 + b1xi) ≥ 0

= 0 nếu dấu(b ′ 0 + b1xi) < 0

- Tín hiệu vẫn là tuyến tính như ng chúng tôi áp dụng phép biến đổi phi tuyến tính cho nó: φ(x) = dấu(b ′ 0 + b1xi)



Third Approach: Conditional Means

- For example, we consider a patient $x = 24$ years old, and we count the relative frequency of chd cases. We compute the conditional mean: $avg(y_i \mid x_i = 24)$
- Similarly, we could compute all conditional means for all age values:

$$(\bar{y} \mid x_i = 25), \quad (\bar{y} \mid x_i = 26), \quad \ldots (\bar{y} \mid x_i = 70),$$

- If we do not have data points for a specific x-value, we can use groups of ages. For example, we compute the proportion of chd cases in the group of ages $20 - 29$ years:

$$avg(y_i \mid x_i \in \{20 - 29 \text{ years}\})$$

Cách tiếp cận thứ ba: Phư ơ ng tiện có điều kiện

- Ví dụ: chúng tôi xem xét một bệnh nhân x = 24 tuổi và chúng tôi tính tần suất tư ơ ng đối của các trư ờng hợp mắc bệnh chd. Chúng tôi tính giá trị trung bình có điều kiện: avg(yi xi = 24)
- Tư ơ ng tự, chúng ta có thể tính tất cả các phư ơ ng tiện có điều kiện cho tất cả các giá trị tuổi:

(y¯ xi = 25), (y¯ xi = 26), . . . (y¯ xi = 70),

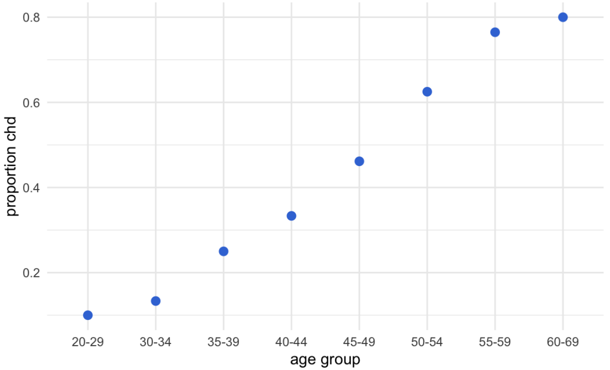
- Nếu chúng tôi không có điểm dữ liệu cho một giá trị x cụ thể, chúng tôi có thể sử dụng các nhóm tuổi. Ví dụ: chúng tôi tính tỷ lệ các trư ờng hợp mắc bệnh chd trong nhóm tuổi 20 - 29 tuổi:

avg(yi xi {20 29 năm})

Third Approach: Conditional Means

- For each age group, we calculate the proportion of chd cases:

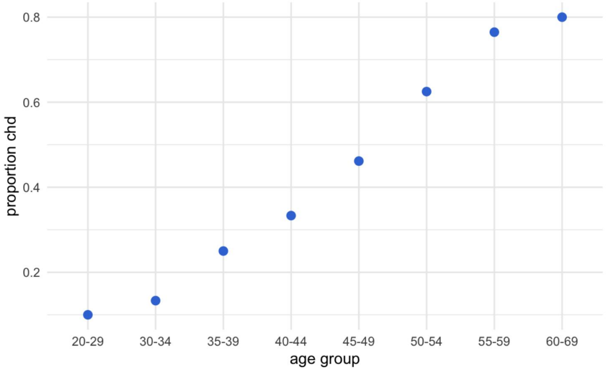
$$avg(y_i \mid x_i \in \text{ age group}\})$$



Cách tiếp cận thứ ba: Phư ơ ng tiện có điều kiện

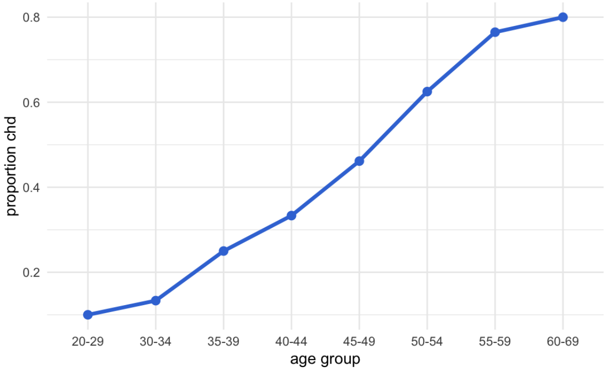
- Với mỗi nhóm tuổi, chúng tôi tính tỷ lệ mắc bệnh chd:

avg(yi xi nhóm tuổi})



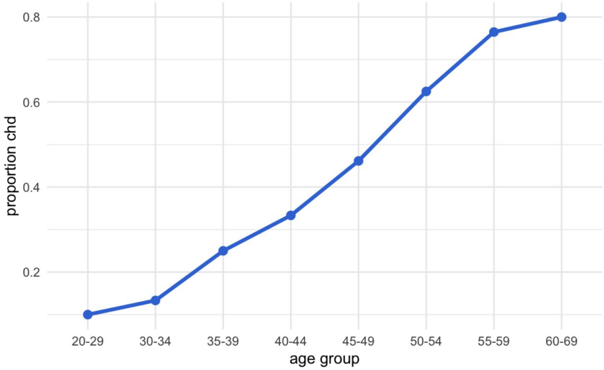
Third Approach: Conditional Means

- Theoretical we are modeling the conditional expectations: $\mathbb{E}(y \mid x)$. This is the regression function.
- Connecting the averages, we have a sigmoid pattern:



Cách tiếp cận thứ ba: Phư ơ ng tiện có điều kiện

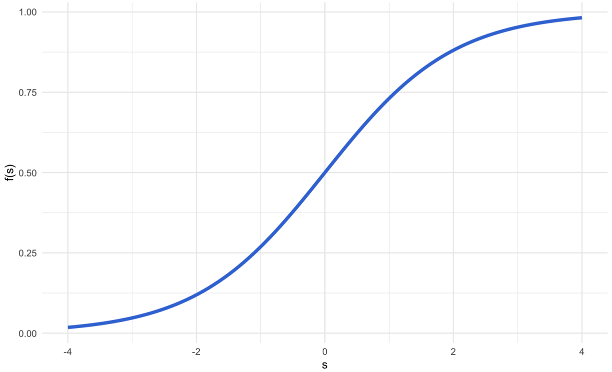
- Về mặt lý thuyết, chúng tôi đang mô hình hóa các kỳ vọng có điều kiện: E(y x). Đây là hàm hồi quy.
- Nối các giá trị trung bình, chúng ta có một mẫu sigmoid:



Third Approach: Conditional Means

- The pattern can be approximated by some mathematical functions.
- The most popular is the logistic function:

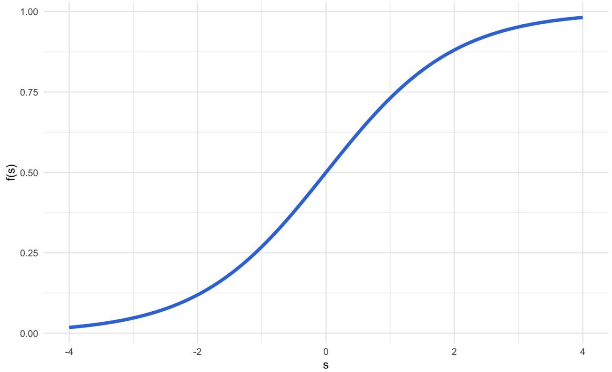
$$f(s) = \frac{e^s}{1 + e^s} = \frac{1}{1 + e^{-s}}$$



Cách tiếp cận thứ ba: Phư ơ ng tiện có điều kiện

• Mẫu có thể đư ợc tính gần đúng bằng một số hàm toán học. • Phổ biến nhất là chức năng logistic:

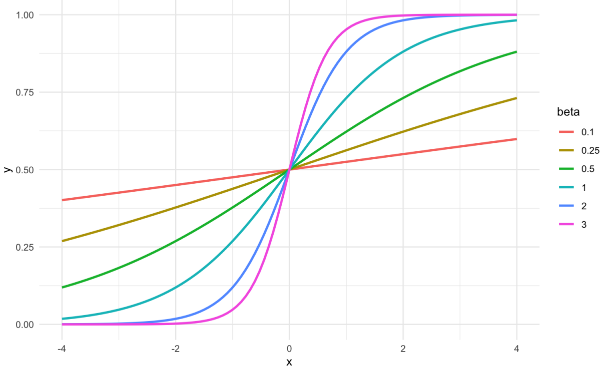
$$f(s) = \frac{s\ e}{1 + e^S} = \frac{1}{1 + e\ s}$$



Third Approach: Conditional Means

- Replacing the signal s by a linear model $\beta_0 + \beta_1 x$, we have:

$$\hat{f}(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}}$$

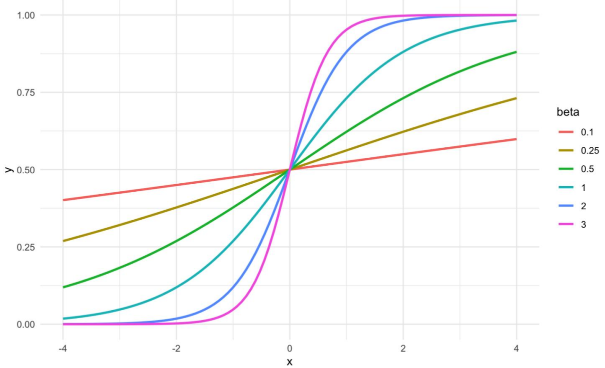


- Since probability values range inside [0,1], instead of using a line to approximate these values, we should use a more adequate curve.

Cách tiếp cận thứ ba: Phư ơ ng tiện có điều kiện • Thay tín

hiệu s bằng mô hình tuyến tính β0 + β1x, ta có:

f(x) = e β0+β1x / 1 + eβ0+β1x = / 1 + e 1 (β0+β1x)



• Vì các giá trị xác suất nằm trong khoảng [0,1] nên thay vì sử dụng một đư ờng để tính gần đúng các giá trị này, chúng ta nên sử dụng một đư ờng cong phù hợp hơ n.

Logistic Regression Model

- We consider the model:

$$Prob(y \mid \mathbf{x}; \mathbf{b}) = f(\mathbf{x})$$

- However, we don't observe the true probability $\in [0, 1]$. We only observe the noisy targets $y_i \in \{0, 1\}$ that is generated by the probability $f(\mathbf{x})$.
- We model the probability by using a mathematical function that has the sigmoid shape, and the most popular function is the logistic function.

$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \frac{e^{\mathbf{b}^\top \mathbf{x}_i}}{1 + e^{\mathbf{b}^\top \mathbf{x}_i}}$$

where $\mathbf{x}_i$ represents the vector of features of individual i.

Mô hình hồi quy logistic

- Ta xét mô hình:

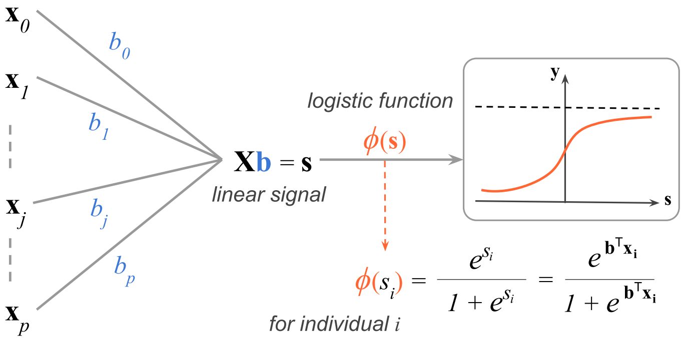
P cư ớp(y x;b) = f(x)

- Tuy nhiên, chúng tôi không quan sát thấy xác suất thực [0, 1]. Chúng tôi chỉ quan sát các mục tiêu nhiễu yi {0, 1} đư ợc tạo bởi xác suất f(x).
- Chúng tôi lập mô hình xác suất bằng cách sử dụng hàm toán học có dạng sigmoid và hàm phổ biến nhất là hàm logistic.

P cư ớp(yi xi ;b) = e b xi / 1 + e b xi

trong đó xi đại diện cho vectơ đặc trư ng của cá nhân i.

Logistic Regression Model

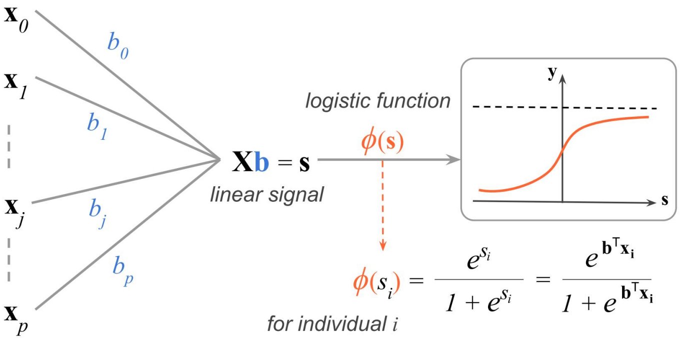


$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \begin{cases} h(\mathbf{x}_i) & \text{for } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{for } y_i = 0 \end{cases}$$

where $h()$ denotes the logistic function $\phi()$:

$$h(\mathbf{x}) = \phi(\mathbf{b}^\top \mathbf{x})$$

Mô hình hồi quy logistic



h(xi) với yi = 1

P cư ớp(yi xi ;b)

= 1 h(xi) với yi = 0

trong đó h() biểu thị hàm logistic φ(): h(x) = φ(b

x)

The Criterion Being Optimized

$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \begin{cases} h(\mathbf{x}_i) & \text{for } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{for } y_i = 0 \end{cases}$$

- Assuming that our model is true, i.e., $h(x) = f(x)$, we ask "how likely is it that we observe the data we already observed (y_i)?"

$$Prob(\mathbf{y} \mid x_1, x_2, \ldots, x_p; \mathbf{b}) = \prod_{i=1}^{n} P(y_i \mid \mathbf{b}^\top \mathbf{x}_i)$$
$$= \prod_{i=1}^{n} h(\mathbf{b}^\top \mathbf{x}_i)^{y_i} [1 - h(\mathbf{b}^\top \mathbf{x}_i)]^{1-y_i}$$

Tiêu chí đư ợc tối ư u hóa

h(xi) với yi = 1
P cư ớp(yi xi ;b) =
1 h(xi) với yi = 0

- Giả sử rằng mô hình của chúng ta là đúng, tức là h(x) = f(x), chúng ta hỏi "khả năng chúng ta quan sát dữ liệu mà chúng ta đã quan sát (yi) là bao nhiêu?"

N
P cư ớp(y x1, x2, . . , xp;b) = P(yi b xi)
tôi = 1
N
= h(b xi) yi [1 h(b xi)]1 yi
tôi = 1

The Criterion Being Optimized

- Take the logarithm (log-likelihood)

$$
\begin{aligned}
l(\mathbf{b}) &= \ln\left[L(\mathbf{b})\right] \\
&= \sum_{i=1}^{n} \ln\left[P(y_i \mid \mathbf{b}^\top \mathbf{x}_i)\right] \\
&= \sum_{i=1}^{n} \{y_i \ln\left[h(\mathbf{b}^\top \mathbf{x}_i)\right] + (1 - y_i)\ln\left[1 - h(\mathbf{b}^\top \mathbf{x}_i)\right]\} \\
&= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{e^{\mathbf{b}^\top \mathbf{x}_i}}{1 + e^{\mathbf{b}^\top \mathbf{x}_i}}\right) + (1 - y_i)\ln\left(1 - \frac{e^{\mathbf{b}^\top \mathbf{x}_i}}{1 + e^{\mathbf{b}^\top \mathbf{x}_i}}\right)\right] \\
&= \sum_{i=1}^{n} \left[y_i \mathbf{b}^\top \mathbf{x}_i - \ln\left(1 + e^{\mathbf{b}^\top \mathbf{x}_i}\right)\right]
\end{aligned}
$$

- Differentiating and setting to 0 yields an equation for which no closed-form solution exists.

Tiêu chí đư ợc tối ư u hóa

- Lấy logarit (log-likelihood)

l(b) = ln [L(b)]

= N ln [P(yi b xi)] tôi = 1

= N {yi ln [h(b xi)] + (1 yi) ln [1 h(b xi)]} tôi = 1

= N i=1 [yi ln (b xi / 1 + eb xi) + (1 yi) ln (1 eb xi / 1 + eb xi)]

= N [yib xi ln (1 + e b xi)] tôi = 1

- Vi phân và đặt thành 0 sẽ tạo ra một phư ơ ng trình không tồn tại nghiệm dạng đóng.

The Criterion Being Optimized

- We need to use gradient ascent.
- The gradient of the likelihood function is:

$$\nabla l(\mathbf{b}) = \sum_{i=1}^{n}\left[y_i\mathbf{x}_i - \left(\frac{e^{\mathbf{b}^\top\mathbf{x}_i}}{1+e^{\mathbf{b}^\top\mathbf{x}_i}}\right)\mathbf{x}_i\right]$$

$$= \sum_{i=1}^{n}[y_i\mathbf{x}_i - \phi(\mathbf{b}^\top\mathbf{x}_i)\mathbf{x}_i]$$

$$= \sum_{i=1}^{n}[y_i - \phi(\mathbf{b}^\top\mathbf{x}_i)]\mathbf{x}_i$$

- In gradient ascent, we would update the parameters $\mathbf{b}$:

$$\mathbf{b}^{(s+1)} = \mathbf{b}^{(s)} + \alpha\nabla l(\mathbf{b}^{(s)})$$

Tiêu chí được tối ư u hóa

• Chúng ta cần sử dụng độ dốc

tăng dần. • Độ dốc của hàm khả năng là:

N b xi
l(b) = i=1 [yixi (1 e+ eb xi)xi]

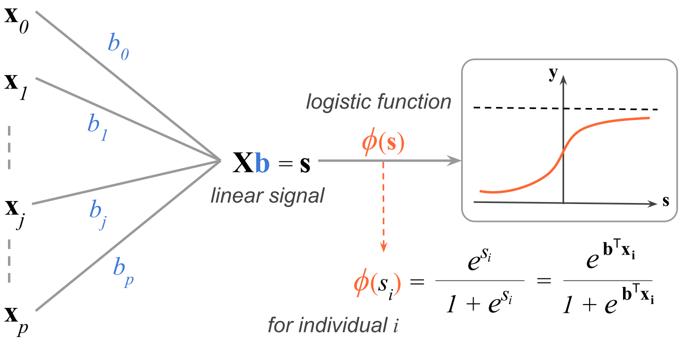
N
= [yixi φ(b xi)xi]
tôi = 1

N
= [yi φ(b xi)]xi
tôi = 1

• Khi tăng dần độ dốc, chúng tôi sẽ cập nhật các tham số b:

b (s+1) = b (s) + α l(b (s))

Another Way to Solve Logistic Regression

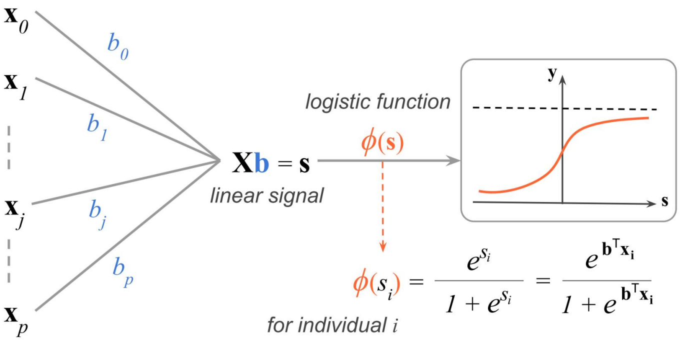


What it used to be $y_i = 0$, let's encode it as $y_i = -1$

$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \begin{cases} h(\mathbf{x}_i) & \text{for } y_i = 1 \\ 1 - h(\mathbf{x}_i) & \text{for } y_i = -1 \end{cases}$$

where $h()$ denotes the logistic function $\phi()$: $h(\mathbf{x}) = \phi(\mathbf{b}^\top \mathbf{x})$

Một cách khác để giải quyết hồi quy logistic



Nó từng là yi = 0, hãy mã hóa nó thành yi = 1

h(xi) với yi = 1

P cư ớp(yi xi ;b) =

1 h(xi) với yi = 1

trong đó h() biểu thị hàm logistic φ(): h(x) = φ(b x)

Another Way to Solve Logistic Regression

- The logistic function has the property:

$$\phi(-s) = \frac{e^{-s}}{1+e^{-s}} = \frac{1-1+e^{-s}}{1+e^{-s}} = 1 - \frac{1}{1+e^{-s}} = 1 - \phi(s)$$

- We can update the expression for the conditional probability:

$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \begin{cases} h(\mathbf{x}_i) & = \phi(y_i \mathbf{b}^\top \mathbf{x}) \text{ for } y_i = 1 \\ 1 - h(\mathbf{x}_i) & = \phi(y_i \mathbf{b}^\top \mathbf{x}) \text{ for } y_i = -1 \end{cases}$$

- This implies that we only need one case regardless of y_i:

$$Prob(y_i \mid \mathbf{x}_i; \mathbf{b}) = \phi(y_i \mathbf{b}^\top \mathbf{x})$$

Một cách khác để giải quyết hồi quy logistic

- Hàm hậu cần có thuộc tính:

φ(s) = e s / 1 + e s = 1 1 + e s / 1 + e s = 1 1 / 1 + e s = 1 φ(s)

- Chúng ta có thể cập nhật biểu thức cho xác suất có điều kiện:

P cư ớp(yi xi ;b) = h(xi) = φ(yib x) với yi = 1
1 h(xi) = φ(yib x) với yi = 1

- Điều này ngụ ý rằng chúng ta chỉ cần một trư ờng hợp bất kể yi :

P cư ớp(yi xi ;b) = φ(yib x)

Another Way to Solve Logistic Regression

- We can compute the likelihood as previously:

$$Prob(\mathbf{y} \mid x_1, x_2, \ldots, x_p; \mathbf{b}) = \prod_{i=1}^{n} P(y_i \mid \mathbf{b}^\top \mathbf{x}_i) = \prod_{i=1}^{n} \phi(y_i \mathbf{b}^\top \mathbf{x}_i)$$

- then log-likelihood:

$$l(\mathbf{b}) = \ln[L(\mathbf{b})] = \sum_{i=1}^{n} \ln[\phi(y_i \mathbf{b}^\top \mathbf{x}_i)] \Rightarrow \frac{1}{n}\sum_{i=1}^{n} \ln[\phi(y_i \mathbf{b}^\top \mathbf{x}_i)]$$

- Instead of maximizing the log-likelihood, we can minimize the negative log-likelihood

$$\min_{\mathbf{b}} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \ln[\phi(y_i \mathbf{b}^\top \mathbf{x}_i)] \right\}$$

Một cách khác để giải quyết hồi quy logistic

- Chúng ta có thể tính toán khả năng xảy ra như trư ớc đây:

P cư ớp(y x1, x2, . . , xp;b) = N P(yi b xi) tôi = 1

= N φ(yib xi) tôi = 1

- thì log-likelihood:

l(b) = ln [L(b)] = N ln [φ(yib xi]) tôi = 1

1/n N i=1 ln [φ(yib xi)]

- Thay vì tối đa hóa khả năng đăng nhập, chúng ta có thể giảm thiểu khả năng đăng nhập tiêu cực

tối thiểu b { 1N N tôi = 1 ln [φ(yib xi)]}

Another Way to Solve Logistic Regression

$$\min_{\mathbf{b}} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \ln\left[\phi(y_i \mathbf{b}^\top \mathbf{x}_i)\right] \right\} \Leftrightarrow \min_{\mathbf{b}} \left\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} \underbrace{\ln\left(1+e^{-y_i\mathbf{b}^\top\mathbf{x}_i}\right)}_{\text{pointwise error}}}_{E_{in}(\mathbf{b})} \right\}$$

- Focus on the product term between the response and the linear signal: $y_i \mathbf{b}^\top \mathbf{x}_i = y_i \mathbf{s}$ where $\mathbf{s}$ represents "signal".
- A small signal means that the probability $\phi(\mathbf{b}^\top \mathbf{x}_i)$ will be small.
- A large signal means that the probability $\phi(\mathbf{b}^\top \mathbf{x}_i)$ will be large.
- The term y_i can be either -1 or +1.

Một cách khác để giải quyết hồi quy logistic

tối thiểu N 1 N yib
b { 1 — n i=1 ln [φ(yib xi)]} phút b — n i=1 xi ln (1 + e)
lỗi theo chiều kim
Ein(b)

- Tập trung vào thuật ngữ tích giữa phản hồi và tín hiệu tuyến tính: yib xi = yis trong đó s đại diện cho "tín hiệu". • Tín hiệu nhỏ nghĩa là xác suất φ(b xi) sẽ nhỏ. • Tín hiệu lớn có nghĩa là xác suất φ(b xi) sẽ lớn. • Số hạng yi có thể là -1 hoặc +1.

VIET NAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY
VNUHCM - UIT
21 / 22

Another Way to Solve Logistic Regression

$$\min_{\mathbf{b}} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \ln\left[\phi(y_i \mathbf{b}^\top \mathbf{x}_i)\right] \right\} \Leftrightarrow \min_{\mathbf{b}} \left\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} \underbrace{\ln\left(1+e^{-y_i\mathbf{b}^\top\mathbf{x}_i}\right)}_{\text{cross-entropy error}}}_{E_{in}(\mathbf{b})} \right\}$$

- With correct predictions, $e^{-y_i\mathbf{b}^\top\mathbf{x}_i}$ will be small, and will give a small error.
- With incorrect predictions, $e^{-y_i\mathbf{b}^\top\mathbf{x}_i}$ will be large, and will give a large error.
- We need to find $\mathbf{b}$ to minimize $E_{in}(\mathbf{b})$. We need to use gradient descent with the gradient w.r.t. $\mathbf{b}$ can be computed as:

$$\nabla E_{in}(\mathbf{b}) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{e^{-y_i\mathbf{b}^\top\mathbf{x}_i}}{1+e^{-y_i\mathbf{b}^\top\mathbf{x}_i}}\right) y_i\mathbf{x}_i = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1+e^{y_i\mathbf{b}^\top\mathbf{x}_i}}\right) y_i\mathbf{x}_i$$

Một cách khác để giải quyết hồi quy logistic

tối thiểu
N
b { 1n i=1 ln [φ(yib xi)]} phút b
1 N
N xi ln (1 + e yib)
tôi = 1
lỗi entropy chéo
Ein(b)

Với dự đoán đúng, e sai yib xi sẽ nhỏ và sẽ cho • số nhỏ.

đoán sai, e yib xi sẽ lớn và sẽ cho kết quả • Với dự sai số lớn.

- Ta cần tìm b để Ein(b) nhỏ nhất. Chúng ta cần sử dụng giảm độ dốc với độ dốc wrt b có thể được tính như sau:

Ein(b) = 1 N yib xi 1 N 1
N i=1 (1 + e yib xi) yixi = n i=1 (1 + eyib xi) yix