CHƯƠNG 4: BÀI TOÁN XẤP XỈ HÀM SỐ (PHẦN 2: XẤP XỈ TRUNG BÌNH PHƯƠNG/ LEAST SQUARE APPROXIMATION)

Tài liệu:

1. Giải Tích Số, Phạm Kỳ Anh
2. Numerical methods in Engineering with Python 3, Kiusalaas
3. Elementary Numerical Analysis, Atkinson and Han

Tác giả: TS. Hà Phi
Khoa Toán – Cơ - Tin học
ĐHKHTN, ĐHQGHN

VÍ DỤ MÔ HÌNH HỒI QUY

Bài toán hồi quy đơn biến (simple regression)

Ví dụ 1: Mức chi tiêu sinh viên = f (Thu nhập gia đình sinh viên đó)

- Thu nhập gia đình của sv ảnh hưởng đến chi tiêu.

Dự đoán mức độ chi tiêu của một tân sinh viên?

Bài toán hồi quy đa biến (multiple regression)

Ví dụ 2: Mức chi tiêu hộ gia đình = f (Thu nhập, Địa điểm, Số thành viên)

- Thu nhập ảnh hưởng đến chi tiêu.
- Địa điểm sinh sống ảnh hưởng đến chi tiêu.
- Số thành viên gia đình ảnh hưởng đến chi tiêu.

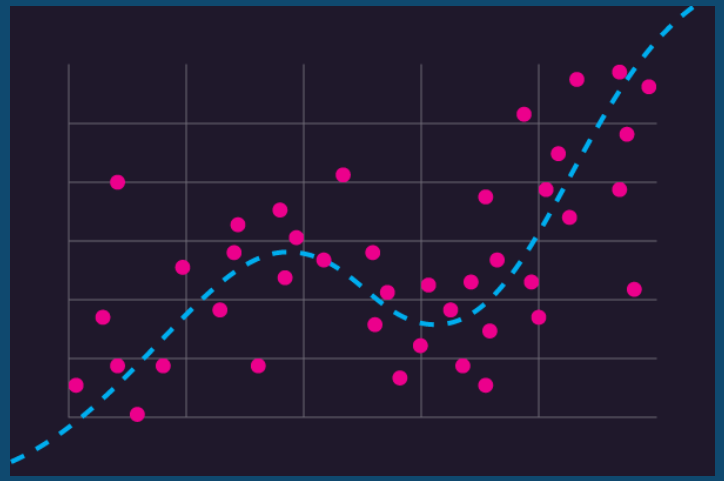
Dự đoán mức độ chi tiêu của một hộ gia đình mới chuyển đến sống?

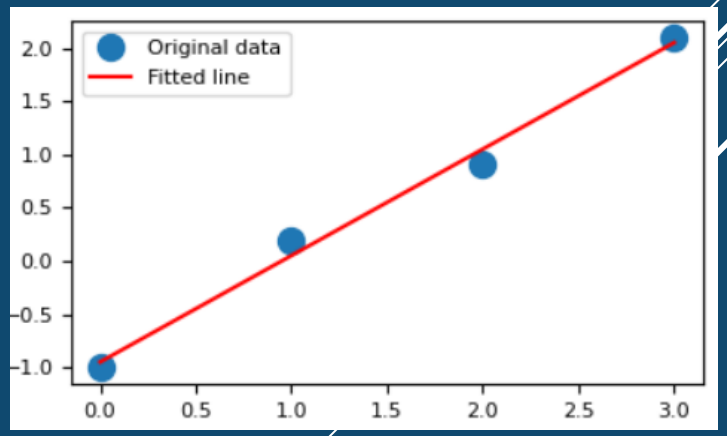
BÀI TOÁN HỒI QUY ĐƠN BIẾN (CURVE FITTING)

- BÀI TOÁN. Thiết kế mô hình xấp xỉ cho phù hợp dữ liệu đã có (CURVE FITTING)

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

- Biến độc lập x ở đây có thể là vector 1 chiều (hồi quy đơn biến/simple regression) hoặc 1 ma trận (hồi quy đa biến/multiple regression). Dưới đây ta chỉ xét mô hình hồi quy đơn biến.
- Nếu dùng nội suy thì có thể bậc của đa thức nội suy rất lớn (nếu n rất lớn), chi phí & thời gian tính toán lớn.
- Thay vào đó người ta sẽ dùng phương pháp bình phương tối thiểu để đi tìm 1 đường cong mà khi thay dữ liệu vào ta thấy hợp lý nhất, ví dụ như ...
- Bài toán này còn có tên là Bài toán hồi quy (Regression) gặp rất nhiều trong Machine Learning. Các mô hình hồi quy rất đa dạng (tuyến tính, đa tuyến tính, logistic,).





HƯỚNG GIẢI QUYẾT

- Bài toán: Tìm mô hình nào diễn tả tốt nhất bảng dữ liệu .

x	x_1	x_2	...	x_m
y	y_1	y_2	...	y_m

- Đơn giản nhất: Hồi quy đa thức
- Ví dụ: Hồi quy tuyến tính (linear regression) $y = c_0 + c_1 x$
- Mô hình đa thức bậc 2 $y = c_0 + c_1 x + c_2 x^2$
- Mô hình đa thức bậc 3 $y = c_0 + c_1 x + c_2 x^2 + c_3 x^3$
- Tổng quát: Đa thức bậc n-1. $y = c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-1} x^{n-1}$

- Lập hệ phương trình
- Chú ý hệ này có thể 0 có nghiệm do số liệu đo đạc có sai số => Tìm nghiệm tốt nhất theo một nghĩa nào đó phù hợp các bài toán trong thực tế

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_m & x_m^2 & \ldots & x_m^{n-1} \end{bmatrix}}_{A \in \mathbb{R}^{m,n}} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}}_{c \in \mathbb{R}^n} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}}_{b \in \mathbb{R}^m}$$

VÍ DỤ: GIẢM LỖI ĐO LƯỜNG

- Giả sử một người khảo sát xác định độ cao của ba ngọn đồi ở trên một số điểm tham chiếu là x1 = 123,7 m, x2 = 194,1m và x3 = 241,7 m và để xác nhận các phép đo này, người khảo sát leo lên đỉnh của ngọn đồi đầu tiên và đo chiều cao của ngọn đồi thứ hai nằm trên ngọn đồi thứ nhất, x2 = x1 + 71,1 và ngọn đồi thứ ba ở trên ngọn đồi thứ nhất là x3 = x1 + 117,7. Tương tự, người khảo sát leo lên ngọn đồi thứ hai và đo chiều cao của ngọn thứ ba trên ngọn đồi thứ hai là x3 = x2 + 47,5.

- Các phương trình này có thể được viết dưới dạng ma trận
- Hệ có nhiều phương trình hơn ẩn số được gọi là **quá xác định (over-determined)**. Nó có thể 0 có nghiệm.
- Người khảo sát nên cung cấp giá trị nào cho độ cao của các ngọn đồi?

$$A * x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \approx \begin{bmatrix} 1237 \\ 1941 \\ 2417 \\ 711 \\ 1177 \\ 475 \end{bmatrix} = b$$

HỆ TỔNG QUÁT (QUÁ/DƯỚI XÁC ĐỊNH) (OVER-/UNDER- DETERMINED SYSTEMS) Nghiệm bình phương tối thiểu

Ax = b

- Nếu A là ma trận m × n thì nói chung, một vectơ m × 1 b có thể không nằm trong không gian cột của A (range(A)). Do đó Ax = b có thể không có nghiệm chính xác.
- Định nghĩa: Vector dư là r = b - Ax.
- Nghiệm bình phương tối thiểu (nhỏ nhất) là vector x sao cho chuẩn 2 (2-norm) của vector dư r là nhỏ nhất.

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{bmatrix}}_{A \in \mathbb{R}^{m,n}} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}}_{c \in \mathbb{R}^n} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}}_{b \in \mathbb{R}^m}$$

$$y = c_0 + c_1 x + c_2 x^2 + \dots + c_{n-1} x^{n-1}$$

PHƯƠNG PHÁP 1: PHƯƠNG TRÌNH CHÍNH TẮC

- Lập phương trình chính tắc $A^T A\, x = A^T b$
- Giải phương trình chính tắc bằng LU hoặc Cholesky.
- Ví dụ 1. Giải hệ phương trình

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix}$$

Chú ý: Thực tế cần kiểm tra tính giải được & số điều kiện của phương trình chính tắc

Consider

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}$$

where $0 < \epsilon < \sqrt{\epsilon_{mach}}$. The normal equations for this system is given by

$$A^T A = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Normal equations

Writing $r = (b - Ax)$ and substituting, we want to find an x that minimizes the following function

$$\phi(x) = \|r\|_2^2 = r^T r = (b - Ax)^T (b - Ax) = b^T b - 2x^T A^T b + x^T A^T A x$$

From calculus we know that the minimizer occurs where $\nabla\phi(x) = 0$.

The derivative is given by

$$\nabla\phi(x) = -2A^T b + 2A^T A x = 0$$

Definition

The system of **normal equations** is given by

$$A^T A x = A^T b.$$

The normal equations has a unique solution if $rank(A) = n$.

Normal equations: conditioning

The normal equations tend to worsen the condition of the matrix.
Since we defined the condition number for a square matrix only we will have to extend this definition for $A_{m\times n}$.

Definition

Let $A_{m\times n}$ have $rank(A) = n$. Then we define the pseudo-inverse A^+ of A as $A^+ = (A^TA)^{-1}A^T$
and we define the condition number of A as,
$cond(A) = \|A\|_2\|A^+\|_2$

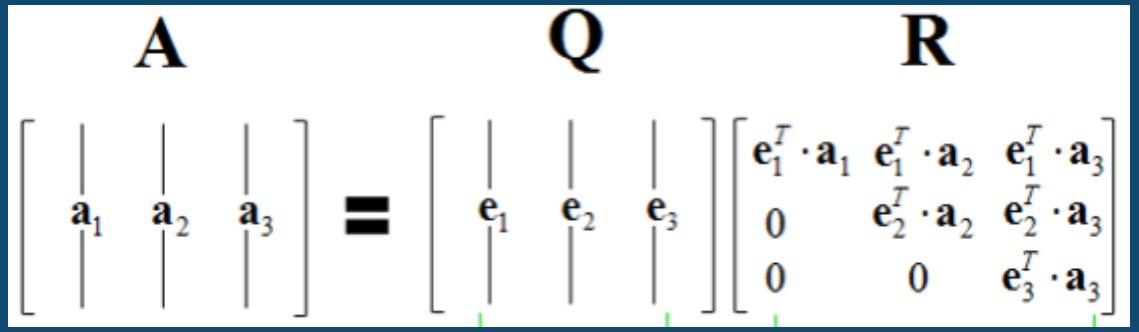
Theorem

$$cond(A^TA) = (cond(A))^2$$

PHƯƠNG PHÁP 2: PHÂN TÍCH QR

- Sử dụng phân tích QR (trực chuẩn hóa Gram-Schmidt) (yêu cầu khi đi thi)

Trực chuẩn hóa Gram-Schmidt các cột của A. (ĐSTT dạy các em như thế nào) ta được phân tích QR trong đó Q là ma trận trực giao, R là ma trận tam giác trên.



- Khi đó, thay vì giải Ax=b ta đi giải $Rx = Q^T b$
- Ví dụ 2. Giải lại hệ trong VD1 bằng phương pháp QR.

Bình luận:
1. Thực tế Matlab hay Python dùng cách 2 để giải các bài toán lớn.
2. Nhưng đối với sv tính tay thì cách 1 tốt hơn.
3. Đi thi yêu cầu làm cách 2, làm cách 1 được 1/3 số điểm.

Using QR factorization for least squares

Now that we know orthogonal matrices preserve the euclidean norm, we can apply orthogonal matrices to the residual vector without changing the norm of the residual. Note that A is $m \; x \; n$, Q is $m \; x \; m$, R' is $n \; x \; n$, x is $n \; x \; 1$ and b is $m \; x \; 1$.

$$\|r\|_2^2 = \|b - Ax\|_2^2 = \left\| b - Q \begin{bmatrix} R' \\ 0 \end{bmatrix} x \right\|_2^2 = \left\| Q^T b - Q^T Q \begin{bmatrix} R' \\ 0 \end{bmatrix} x \right\|_2^2 = \left\| Q^T b - \begin{bmatrix} R' \\ 0 \end{bmatrix} x \right\|_2^2$$

If $Q^T b = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ where c_1 is an $n \; x \; 1$ vector then

$$\left\| Q^T b - \begin{bmatrix} R' \\ 0 \end{bmatrix} x \right\|_2^2 = \left\| \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} R'x \\ 0 \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} c_1 - R'x \\ c_2 \end{bmatrix} \right\|_2^2 = \|c_1 - R'x\|_2^2 + \|c_2\|_2^2$$

Hence the least squares solution is given by solving $R'x = c_1$. We can solve $R'x = c_1$ using back substitution and the residual is $\|r\|_2 = \|c_2\|_2$.

PHẦN NÂNG CAO: PHÂN TÍCH SVD

Using SVD for least squares

Recall that a singular value decomposition is given by

$$A = \begin{bmatrix} \vdots & \vdots & \vdots \\ u_1 & \dots & u_m \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \ddots \\ & & & & 0 \end{bmatrix} \begin{bmatrix} \dots & v_1^T & \dots \\ \dots & \vdots & \dots \\ \dots & v_n^T & \dots \end{bmatrix}$$

where σ_i are the singular values.

Assume that A has rank k (and hence k nonzero singular values σ_i) and recall that we want to minimize

$$||r||_2^2 = ||b - Ax||_2^2.$$

Substituting the SVD for A we find that

$$||r||_2^2 = ||b - Ax||_2^2 = ||b - USV^T x||_2^2$$

where U and V are orthogonal and S is diagonal with k nonzero singular values.

$$||b - USV^T x||_2^2 = ||U^T b - U^T USV^T x||_2^2 = ||U^T b - SV^T x||_2^2$$

Let $c = U^T b$ and $y = V^T x$ (and hence $x = Vy$) in $||U^T b - SV^T x||_2^2$. We now have

$$||r||_2^2 = ||c - Sy||_2^2$$

Since S has only k nonzero diagonal elements, we have

$$||r||_2^2 = \sum_{i=1}^{k} (c_i - \sigma_i y_i)^2 + \sum_{i=k+1}^{m} c_i^2$$

which is minimized when $y_i = \frac{c_i}{\sigma_i}$ for $1 \leqslant i \leqslant k$.

Theorem

Let A be an $m \times n$ matrix of rank r and let $A = USV^T$, the singular value decomposition. The least squares solution of the system $Ax = b$ is

$$x = \sum_{i=1}^{r} (\sigma_i^{-1} c_i) v_i$$

where $c_i = u_i^T b$.

Weighting of Data

There are occasions when our confidence in the accuracy of data varies from point to point. For example, the instrument taking the measurements may be more sensitive in a certain range of data. Sometimes the data represent the results of several experiments, each carried out under different conditions. Under these circumstances we may want to assign a confidence factor, or *weight*, to each data point and minimize the sum of the squares of the *weighted residuals* $r_i = W_i\left[y_i - f(x_i)\right]$, where W_i are the weights. Hence the function to be minimized is

$$S(a_0, a_1, \ldots, a_m) = \sum_{i=0}^{n} W_i^2\left[y_i - f(x_i)\right]^2 \tag{3.24}$$

This procedure forces the fitting function $f(x)$ closer to the data points that have higher weights.

Weighted linear regression. If the fitting function is the straight line $f(x) = a + bx$, Eq. (3.24) becomes

$$S(a, b) = \sum_{i=0}^{n} W_i^2(y_i - a - bx_i)^2 \tag{3.25}$$

Fitting exponential functions. A special application of weighted linear regression arises in fitting various exponential functions to data. Consider as an example the fitting function

$$f(x) = ae^{bx}$$

Normally, the least-squares fit would lead to equations that are nonlinear in a and b.

Other examples that also benefit from the weights $W_i = y_i$ are given in Table 3.4.

$f(x)$	$F(x)$	Data to be fitted by $F(x)$
axe^{bx}	$\ln[f(x)/x] = \ln a + bx$	$[x_i, \ln(y_i/x_i)]$
ax^b	$\ln f(x) = \ln a + b\ln(x)$	$(\ln x_i, \ln y_i)$

Table 3.4. Fitting exponential functions.

fit to the original data. The residuals of the logarithmic fit are

$$R_i = \ln y_i - F(x_i) = \ln y_i - (\ln a + bx_i) \tag{3.29a}$$

whereas the residuals used in fitting the original data are

$$r_i = y_i - f(x_i) = y_i - ae^{bx_i} \tag{3.29b}$$

LỰA CHỌN ĐA THỨC NỘI SUY NÀO CHO PHÙ HỢP BẢNG DỮ LIỆU

Đánh giá sai số của đa thức nội suy bằng độ lệch chuẩn: càng bé càng tốt

The spread of the data about the fitting curve is quantified by the *standard deviation*, defined as

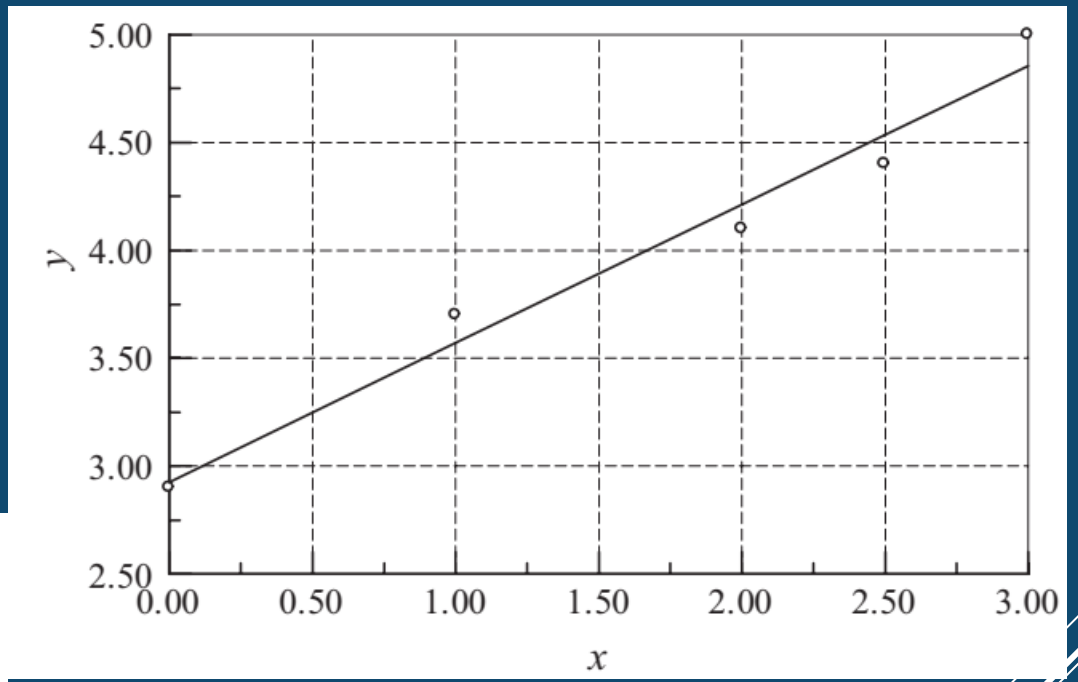
$$\sigma = \sqrt{\frac{S}{n-m}} \tag{3.15}$$

EXAMPLE 3.10

Fit a straight line to the data shown and compute the standard deviation.

x	0.0	1.0	2.0	2.5	3.0
y	2.9	3.7	4.1	4.4	5.0

Lời giải: Đáp án là f(x) = 2.927 + 0.6431x



We start the evaluation of the standard deviation by computing the residuals:

x	0.000	1.000	2.000	2.500	3.000
y	2.900	3.700	4.100	4.400	5.000
$f(x)$	2.927	3.570	4.213	4.535	4.856
$y - f(x)$	−0.027	0.130	−0.113	−0.135	0.144

The sum of the squares of the residuals is

$$S = \sum [y_i - f(x_i)]^2$$

$$= (-0.027)^2 + (0.130)^2 + (-0.113)^2 + (-0.135)^2 + (0.144)^2 = 0.06936$$

so that the standard deviation in Eq. (3.15) becomes

$$\sigma = \sqrt{\frac{S}{5-2}} = \sqrt{\frac{0.06936}{3}} = 0.1520$$

BÀI TOÁN HỒI QUY ĐA TUYẾN TÍNH (MULTIPLE LINEAR REGRESSION)

Xét một dữ liệu có biến phụ thuộc là Y và n biến độc lập là $X_1, X_2, ..., X_n$. Mô hình hồi quy tuyến tính mô tả mối quan hệ giữa Y và X_i được trình bày dưới dạng:

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n + e = b_0 + \sum_{i=1}^{n} b_iX_i + e \quad (1)$$

trong đó :

- b_i là các hệ số của phương trình hồi quy. Về mặt ý nghĩa, b_i là độ tăng trung bình của Y khi ta tăng X_i một đơn vị trong khi giữ tất cả các biến độc lập khác không đổi.
- e là sai lệch (hoặc sai số) giữa mô hình và số liệu thực tế. e cũng còn được gọi là phần dư (residual).

Ta đã biết giá trị của Y và X_i cho tất cả các phần tử từ dữ liệu. Công việc quan trọng cần làm là xác định giá trị các hệ số b_i.

- Dạng ma trận của mô hình xấp xỉ $Y = X\,b$
- Dạng ma trận của mô hình thực tế $Y_{data} = X_{data}\,b + e$

trong đó

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1n} \\ 1 & X_{21} & \cdots & X_{2n} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nn} \end{pmatrix}$$

Ta đi giải bài toán bình phương tối thiểu

$$Min \rightarrow \|e\|_2 = \sum_{i=1}^{m} e_i^2$$

- Về mặt lý thuyết, phương pháp thường dùng để xác định các hệ số của phương trình hồi quy cho trường hợp hồi quy đa biến là bình phương cực tiểu (least square) có nguyên tắc tương tự như trường hợp một biến độc lập.
- Như vậy ta có 2 cách tiếp cận:
  - Phương trình chính tắc (normal equation)

$$b = (X^T X)^{-1}(X^T Y)$$

  - Bình phương tối thiểu bằng phương pháp QR hoặc SVD.

$$X = QR$$

$$Rb = Q^T X \quad (Unknown: b)$$

hoặc

$$X = USV^T$$

$$SV^T b = U^T Y$$

NHỮNG GÌ THẦY CHƯA NÓI TRONG CHƯƠNG NÀY (MÀ CÓ THỂ GẶP TRONG THỰC TẾ)

- Bài toán hồi quy tuyến tính với các hàm mục tiêu khác, v.d. hồi quy Ridge, hồi quy Lasso, hồi quy logistic
- Hồi quy với biến giả (dummy variables) trong trường hợp dữ liệu có tính chất mùa vụ.
- Phương pháp CG (gradient liên hợp), phương pháp giảm gradient (gradient descent/steepest descent)