Tối ưu lồi

Đại học Khoa học Tự Nhiên, ĐHQG Thành phố Hồ Chí Minh

Ngày 17 tháng 11 năm 2021

1. Bài toán bình phương tối tiểu

2. Least squares data fitting

Outlines

Bài toán bình phương tối tiểu

Định nghĩa

Cho $A \in \mathbb{R}^{m \times n}$ và $\mathrm{b} \in \mathbb{R}^m$, bài toán tìm $\mathrm{x} \in \mathbb{R}^n$ để

$$\text{minimize} \;\; \|A\mathrm{x} - \mathrm{b}\|^2$$

được gọi là bài toán bình phương tối tiểu. Vector $\mathrm{r} = A\mathrm{x} - \mathrm{b} \in \mathrm{R}^m$ được gọi là phần dư (residual) và $\hat{\mathrm{x}}$ được gọi là nghiệm của bài toán, trong đó

$$\begin{aligned} \|A\mathrm{x} - \mathrm{b}\|^2 &= \|\mathrm{r}\|^2 = r_1^2 + r_2^2 + \cdots + r_m^2 \\ &= \left(a_1^T\mathrm{x} - b_1\right)^2 + \cdots + \left(a_m^T\mathrm{x} - b_m\right)^2 \end{aligned}$$

với a_i^T là dòng thứ i của ma trận A,

Bài toán bình phương tối tiểu

Mệnh đề

Với bài toán tìm nghiệm $\hat{\mathrm{x}}$ cho điều kiện minimize $\|A\mathrm{x}-\mathrm{b}\|^2$, nếu các cột của A độc lập tuyến tính (A có hạng là n) thì

$$\hat{\mathrm{x}} = \left(A^T A\right)^{-1} A^T \mathrm{b}$$

là nghiệm duy nhất của bài toán.

Chứng minh nghiệm bài toán bình phương tối tiểu

Với ma trận $A \in {}^{m\times n}$, $n \leq m$, $rank\ A = n$ và vector $b \in {}^{m}$. Xét hàm số $f : \mathbb{R}^n \to \mathbb{R}$ xác định bởi $f(x) = \|Ax - b\|^2$, ta có

$$\begin{aligned} f(\mathrm{x}) &= \langle A\mathrm{x} - \mathrm{b}, A\mathrm{x} - \mathrm{b} \rangle \\ &= (A\mathrm{x} - \mathrm{b})^T (A\mathrm{x} - \mathrm{b}) = \left(\mathrm{x}^T A^T - \mathrm{b}^T\right)(A\mathrm{x} - \mathrm{b}) \\ &= \mathrm{x}^T A^T A\mathrm{x} - \mathrm{b}^T A\mathrm{x} - \mathrm{x}^T A^T \mathrm{b} + \mathrm{b}^T \mathrm{b} \\ &= \mathrm{x}^T \left(A^T A\right)\mathrm{x} - \left(2\mathrm{b}^T A\right)\mathrm{x} + \mathrm{b}^T \mathrm{b}, \end{aligned}$$

Vì $\mathrm{b}^T A\mathrm{x} = \mathrm{x}^T A^T \mathrm{b}$, nên

$$\nabla f(\mathrm{x}) = 2\left(A^T A\right)\mathrm{x} - 2A^T\mathrm{b} \text{ và } \nabla^2 f(\mathrm{x}) = 2\left(A^T A\right)$$

Do với mọi $\mathrm{h} \in \mathbb{R}^m$,

$$\mathrm{h}^T \nabla^2 f(\mathrm{x})\,\mathrm{h} = 2\left[\mathrm{h}^T \left(A^T A\right)\mathrm{h}\right] = 2\left[\left(\mathrm{h}^T A^T\right)(A\mathrm{h})\right] = 2\|A\mathrm{h}\|^2 \geq 0$$

Chứng minh nghiệm bài toán bình phương tối tiểu

Nên $\nabla^2 f(\mathrm{x})$ là ma trận xác định dương. Ngoài ra vì $rank(A^T A) = n$ nên $A^T A$ là ma trận khả nghịch.
Suy ra nghiệm $\mathrm{x} = (A^T A)^{-1} A^T \mathrm{b}$ của phương trình $\nabla f(\mathrm{x}) = 0$ là cực tiểu toàn cục của f.

Ví dụ

Cho ma trận $A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix}$, $\mathrm{b} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, hãy tìm nghiệm của bài toán minimize $\|A\mathrm{x} - \mathrm{b}\|^2$.

Bài toán bình phương tối tiểu

Ta có

$$A^TA = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

$$(A^TA)^{-1} = \begin{pmatrix} 5/24 & 1/24 \\ 1/24 & 5/24 \end{pmatrix}$$

$$(A^TA)^{-1}A^T = \begin{pmatrix} 5/12 & -1/6 & 1/12 \\ 1/12 & 1/6 & 5/12 \end{pmatrix}$$

Vậy nghiệm của bài toán

$$\hat{\mathrm{x}} = (A^TA)^{-1}A^T\mathrm{b} = \begin{pmatrix} 1/3 \\ -1/3 \end{pmatrix}$$

Outlines

Least squares data fitting

Một trong những bài toán căn bản trong các ngành khoa học ứng dụng là khảo sát sự liên hệ giữa một đại lượng, ký hiệu y, theo một số đại lượng khác, ký hiệu $x_1, x_2, \cdots, x_n$. Tiếp cận đơn giản nhất là tìm một hàm số $f : {}^n \to$ sao cho $y \approx f(x)$.

Trong đó $\mathrm{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ được gọi là vectơ các biến độc lập (independent variables), và $y \in$ được gọi là biến phụ thuộc. (response variable).

Một cách tự nhiên, việc đầu tiên mà người ta cần làm là thu thập số liệu của các biến độc lập

$$\mathrm{x}^{(1)} = \left(x_1^{(1)}, \cdots, x_n^{(1)}\right), \cdots, \mathrm{x}^{(N)} = \left(x_1^{(N)}, \cdots, x_n^{(N)}\right)$$

và của biến phụ thuộc tương ứng, $y^{(1)}, \cdots, y^{(N)}$

Ta nhận được bộ dữ liệu (data) cho các biến, có thể sắp xếp dưới dạng ma trận như sau.

Least squares data fitting

	x_1	x_2	$\dots$	x_n	y
Số liệu(data) 1	$x_1^{(1)}$	$x_2^{(1)}$	$\dots$	$x_n^{(1)}$	$y^{(1)}$
Số liệu(data) 2	$x_1^{(2)}$	$x_2^{(2)}$	$\dots$	$x_n^{(2)}$	$y^{(2)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
Số liệu(data) N	$x_1^{(N)}$	$x_2^{(N)}$	$\dots$	$x_n^{(N)}$	$y^{(N)}$

Một cách tiếp cận cho bài toán khảo sát là đưa ra một mô hình (model) cho mối quan hệ giữa biến phụ thuộc và các biến độc lập thông qua một hàm $\hat{f} : {}^n \to$ sao cho $y \approx \hat{f}(x)$ và $\hat{y} = \hat{f}(x)$ được gọi là giá trị dự báo(prediction) của biến phụ thuộc y, khi biết giá trị của các biến độc lập $x_1, x_2, \cdots, x_n$.
Hàm $\hat{f}$ được gọi là hàm dự báo (prediction function), hay vắn tắt là mô hình (model) của bài toán.

Trong phần này, ta tập trung khảo sát mô hình tuyến tính theo tham số (linear in the parameters model) cho lớp hàm dự báo,

$$\hat{f}(\mathrm{x}) = \theta_1 f_1(\mathrm{x}) + \cdots + \theta_p f_p(\mathrm{x}) \tag{1}$$

Trong đó $f_i : \mathbb{R}^n \to \mathbb{R}$ là các hàm số cơ sở (basis functions) được xác định trước, các θ_i là các tham số mô hình (model parameters) cần tìm sao cho hàm dự báo tương ứng là phù hợp (consistent) với dữ liệu (data) thu thập được.

Để đánh giá sự phù hợp của mô hình (1) với dữ liệu, ta tính các giá trị dự báo của mô hình tương ứng với giá trị của các biến độc lập của dữ liệu,

$$\hat{\mathrm{y}^{(i)}} = \hat{f}\left(\mathrm{x}^{(i)}\right) = \hat{f}\left(x_1^{(i)}, \cdots, x_n^{(i)}\right) \text{ với } i = 1, \cdots, N$$

và đánh giá các sai số dự báo (prediction error) hay phần dư (residual),

$$r^{(i)} = \hat{y}^{(i)} - y^{(i)} \text{ với } i = 1, \cdots, N \tag{2}$$

Phương pháp bình phương nhỏ nhất tìm cách xác định các tham số θ_i của mô hình (1) sao cho vectơ các phần dư

$$\mathrm{r} = \left(r^{(1)}, \cdots, r^{(N)}\right) \in {}^N \tag{3}$$

có chuẩn nhỏ nhất.
Từ (1), (2), ta có

$$\begin{aligned}
\mathrm{r}^{(i)} &= \theta_1 f_1\left(\mathrm{x}^{(i)}\right) + \cdots + \theta_p f_p\left(\mathrm{x}^{(i)}\right) - y^{(i)} \\
&= \begin{bmatrix} f_1\left(\mathrm{x}^{(i)}\right) & \cdots & f_p\left(\mathrm{x}^{(i)}\right) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} - y^{(i)} \text{ với } i = 1, \cdots, N
\end{aligned}$$

Bằng cách đặt

$$A = \begin{bmatrix} f_1\left(\mathrm{x}^{(1)}\right) & \cdots & f_p\left(\mathrm{x}^{(1)}\right) \\ \vdots & & \vdots \\ f_1\left(\mathrm{x}^{(N)}\right) & \cdots & f_p\left(\mathrm{x}^{(N)}\right) \end{bmatrix} \in \mathbb{R}^{N\times p},$$

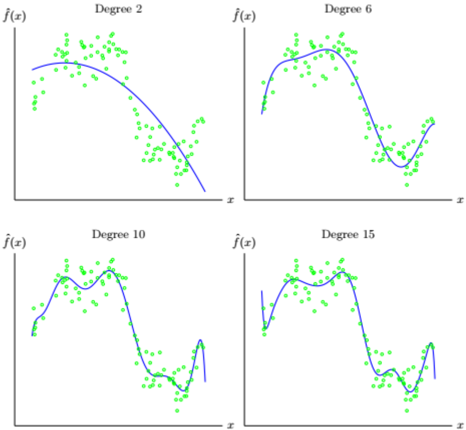
$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} \in \mathbb{R}^p, \mathrm{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix} \in \mathbb{R}^N$$

Ta có $\mathrm{r} = A\theta - \mathrm{y}$. Do áp dụng kết quả nghiệm của bài toán bình phương tối tiểu, ta có

$$\theta = \left(A^T A\right)^{-1} A^T \mathrm{y}$$

Least squares data fitting

Đồ thị sau thể hiện việc khớp dữ liệu bằng các hàm đa thức ở các bậc khác nhau, trong đó với đa thức bậc 15, các điểm dữ liệu được khớp rất tốt.



Least squares data fitting

Ví dụ

Cho bộ dữ liệu trong bảng sau

Điểm i	x_i	y_i
1	-2	9
2	-1	-1
3	0	1
4	1	3
5	2	17

a. Hãy fitting dữ liệu trên bằng hàm $y = \theta_1 + \theta_2 x$
b. Hãy fitting dữ liệu trên bằng hàm $y = \theta_1 + \theta_2 x^2$
c. Hãy fitting dữ liệu trên bằng hàm $y = \theta_1 + \theta_2 x + \theta_3 x^2$

Giải câu a)

Ta có $f = \theta_1 + \theta_2 x$,

$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}, \mathrm{y} = \begin{bmatrix} 9 \\ -1 \\ 1 \\ 3 \\ 17 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}, (A^T A)^{-1} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}$$

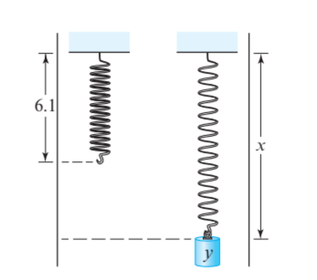
nên ta có các tham số cần tìm cho hàm dự báo là

$$\hat{\theta} = (A^T A)^{-1} A^T \mathrm{y} = \begin{bmatrix} 5.8 \\ 2 \end{bmatrix}$$

Vậy đa thức bậc nhất dùng để dự báo là $y = 5.8 + 2x$

Ứng dụng bài toán data fitting

Theo định luật Hook trong Vật lý, chiều dài của một lò xo x là một hàm tuyến tính theo trọng lượng treo vào lò xo y. Nếu biểu diễn sự liên hệ thông qua dạng hàm $y = a + bx$ thì b được gọi là hằng số co dãn của lò xo. Giả sử một lò xo có chiều dài tự nhiên là 6.1 inches như trong hình bên dưới, ngoài ra các thí nghiệm với trọng lượng treo và chiều dài lò xo tương ứng được ghi lại.



Ứng dụng bài toán data fitting

Weight y (lb)	0	2	4	6
Length x (in)	6.1	7.6	8.7	10.4

Hãy dùng phương pháp bình phương tối tiểu để tìm đường thẳng có dạng $y = a + bx$ để khớp dữ liệu và tìm ước lượng cho hằng số co dãn của lò xo ở thí nghiệm trên.

Ứng dụng bài toán data fitting - Tính toán trên Python

Một cửa hàng mô hình chi phí bán hàng (1000\$) y theo biến độc lập doanh số bán hàng(1000\$) x theo mô hình sau

$$y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Tìm mô hình khớp dữ liệu của bài toán với chi phí và doanh số được ghi lại như sau

$$\begin{array}{cccc} (4, 1.58) & (6, 2.08) & (8, 2.5) & (10, 2.8) \\ (12, 3.1) & (14, 3.4) & (16, 3.8) & (18, 4.32) \end{array}$$

Ứng dụng bài toán data fitting - Python

Hàm lượng phân rã của 2 chất phóng xạ A, B tỷ lệ lần lượt với 0.02 và 0.07. Một hỗn hợp 2 chất A, B tại thời điểm 0 chứa M_A gram A và M_B gram B, tổng khối lượng y gram của 2 chất A, B theo thời gian t được mô hình như sau.

$$y = M_A e^{-0.02t} + M_B e^{-0.07t}$$

Nếu khối lượng ban đầu của A, B chưa biết, một nhà khoa học đã đo tổng khối lượng của hỗn hợp A,B theo thời gian và nhận được số liệu như sau

$$(10, 21.34) \quad (11, 20.68) \quad (12, 20.05) \quad (14, 18.87) \quad (15, 18.30)$$

Tìm mô hình phù hợp để ước lượng M_A, M_B.

Ứng dụng bài toán data fitting - Python

Huyết áp của trẻ em(mgHg) p và cân nặng của trẻ(pound) w có ảnh hưởng đến nhau và được mô hình bởi phương trình

$$p = \beta_0 + \beta_1 \ln w$$

Sử dụng dữ liệu đã được ghi lại ở bảng sau để ước lượng huyết áp của một trẻ em có cân nặng là 100 pounds.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

Ứng dụng bài toán data fitting - Python

Để kiểm tra quá trình cất cánh của máy bay, khoảng cách theo chiều ngang từ khi máy bay cất cánh được ghi lại mỗi giây từ $t = 0$ đến $t = 12$. Vị trí của máy bay (feet), y tương ứng như sau

0	8.8	29.9	62.0	104.7	159.1	222.0
294.5	380.4	471.1	571.7	686.8	809.2	

a. Hãy tìm mô hình dự báo của y theo thời gian t
$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3^3$,

b. Dùng mô hình đã tìm được ở trên, tìm vận tốc máy bay tại $t = 4.54$ giây.