

Ôn tập Đại số tuyến tính

Theo blog: <http://machinelearningcoban.com>

Vũ Hữu Tiệp, 2017

1 Lưu ý về ký hiệu

Trong các bài viết của tôi, các số vô hướng được biểu diễn bởi các chữ cái viết ở dạng không in đậm, có thể viết hoa, ví dụ x_1, N, y, k . Các vector được biểu diễn bằng các chữ cái thường in đậm, ví dụ \mathbf{y}, \mathbf{x}_1 . Nếu không giải thích gì thêm, các vector được mặc định hiểu là các vector cột. Các ma trận được biểu diễn bởi các chữ viết hoa in đậm, ví dụ $\mathbf{X}, \mathbf{Y}, \mathbf{W}$.

Đối với vector, $\mathbf{x} = [x_1, x_2, \dots, x_n]$ được hiểu là một vector hàng. Trong khi $\mathbf{x} = [x_1; x_2; \dots; x_n]$ được hiểu là vector cột. Chú ý sự khác nhau giữa dấu phẩy (,) và dấu chấm phẩy (;). Đây chính là ký hiệu được Matlab sử dụng.

Tương tự, trong ma trận, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ được hiểu là các vector cột \mathbf{x}_j được đặt cạnh nhau theo thứ tự từ trái qua phải để tạo ra ma trận \mathbf{X} . Trong khi $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_m]$ được hiểu là các vector \mathbf{x}_i được đặt chồng lên nhau theo thứ tự từ trên xuống dưới để tạo ra ma trận \mathbf{X} . Các vector được ngầm hiểu là có kích thước phù hợp để có thể xếp cạnh hoặc xếp chồng lên nhau.

Cho một ma trận \mathbf{W} , nếu không giải thích gì thêm, chúng ta hiểu rằng \mathbf{w}_i là **vector cột** thứ i của ma trận đó. Chú ý sự tương ứng giữa ký tự viết hoa và viết thường.

2 Norms (chuẩn)

Trong không gian một chiều, việc đo khoảng cách giữa hai điểm đã rất quen thuộc: lấy trị tuyệt đối của hiệu giữa hai giá trị đó. Trong không gian hai chiều, tức mặt phẳng, chúng ta thường dùng khoảng cách Eclid để đo khoảng cách giữa hai điểm. Khoảng cách này chính là cái chúng ta thường nói bằng ngôn ngữ thông thường là *đường chim bay*. Đôi khi, để đi từ một điểm này tới một điểm kia, con người chúng ta không thể đi bằng đường chim bay được mà còn phụ thuộc vào việc đường đi nối giữa hai điểm có dạng như thế nào nữa.

Việc đo khoảng cách giữa hai điểm dữ liệu nhiều chiều, tức hai vector, là rất cần thiết trong Machine Learning. Chúng ta cần đánh giá xem điểm nào là điểm gần nhất của một điểm khác; chúng ta cũng cần đánh giá xem độ chính xác của việc ước lượng; và trong rất nhiều ví dụ khác nữa.

Và đó chính là lý do mà khái niệm norm (chuẩn) ra đời. Có nhiều loại norm khác nhau mà

các bạn sẽ thấy ở dưới đây:

Để xác định khoảng cách giữa hai vector \mathbf{y} và \mathbf{z} , người ta thường áp dụng một hàm số lên vector hiệu $\mathbf{x} = \mathbf{y} - \mathbf{z}$. Một hàm số được dùng để đo các vector cần có một vài tính chất đặc biệt.

2.1 Định nghĩa

Một hàm số $f()$ ánh xạ một điểm \mathbf{x} từ không gian n chiều sang tập số thực một chiều được gọi là norm nếu nó thỏa mãn ba điều kiện sau đây:

1. $f(\mathbf{x}) \geq 0$. Dấu bằng xảy ra $\Leftrightarrow \mathbf{x} = \mathbf{0}$.
2. $f(\alpha\mathbf{x}) = \|\alpha\|f(\mathbf{x})$, $\forall \alpha \in \mathbb{R}$
3. $f(\mathbf{x}_1) + f(\mathbf{x}_2) \geq f(\mathbf{x}_1 + \mathbf{x}_2)$, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^n$

Điều kiện thứ nhất là dễ hiểu vì khoảng cách không thể là một số âm. Hơn nữa, khoảng cách giữa hai điểm \mathbf{y} và \mathbf{z} bằng 0 nếu và chỉ nếu hai điểm nó trùng nhau, tức $\mathbf{x} = \mathbf{y} - \mathbf{z} = \mathbf{0}$.

Điều kiện thứ hai cũng có thể được lý giải như sau. Nếu ba điểm \mathbf{y} , \mathbf{v} và \mathbf{z} thẳng hàng, hơn nữa $\mathbf{v} - \mathbf{y} = \alpha(\mathbf{v} - \mathbf{z})$ thì khoảng cách giữa \mathbf{v} và \mathbf{y} sẽ gấp $\|\alpha\|$ lần khoảng cách giữa \mathbf{v} và \mathbf{z} .

Điều kiện thứ ba chính là bất đẳng thức tam giác nếu ta coi $\mathbf{x}_1 = \mathbf{w} - \mathbf{y}$, $\mathbf{x}_2 = \mathbf{z} - \mathbf{w}$ với \mathbf{w} là một điểm bất kỳ trong cùng không gian.

2.2 Một số chuẩn thường dùng

Giả sử các vectors $\mathbf{x} = [x_1; x_2; \dots; x_n]$, $\mathbf{y} = [y_1; y_2; \dots; y_n]$.

Nhận thấy rằng khoảng cách Euclid chính là một norm, norm này thường được gọi là **norm 2**:

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (1)$$

Với p là một số không nhỏ hơn 1 bất kỳ, hàm số sau đây:

$$\|\mathbf{x}\|_p = (\|x_1\|^p + \|x_2\|^p + \dots \|x_n\|^p)^{\frac{1}{p}} \quad (2)$$

được chứng minh thỏa mãn ba điều kiện bên trên, và được gọi là **norm p**.

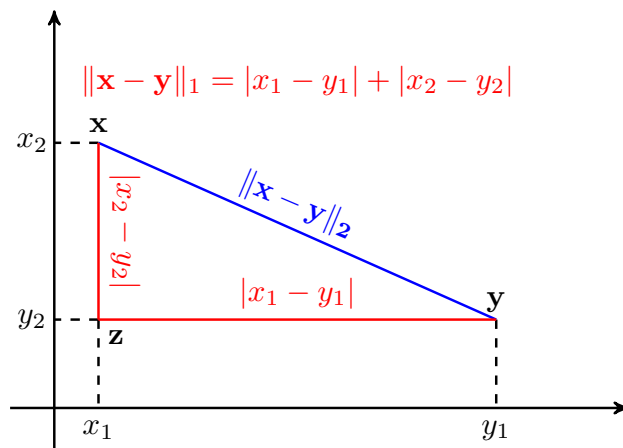


Figure 1: Minh họa norm 1 và norm 2

Nhận thấy rằng khi $p \rightarrow 0$ thì biểu thức bên trên trở thành *số các phần tử khác 0 của \mathbf{x}* . Hàm số (2) khi $p = 0$ được gọi là giả chuẩn (pseudo-norm) 0. Nó không phải là norm vì nó không thỏa mãn điều kiện 2 và 3 của norm. Giả-chuẩn này, thường được ký hiệu là $\|\mathbf{x}\|_0$, khá quan trọng trong Machine Learning vì trong nhiều bài toán, chúng ta cần có ràng buộc “sparse”, tức số lượng thành phần “active” của \mathbf{x} là nhỏ.

Có một vài giá trị của p thường được dùng:

1. Khi $p = 2$ chúng ta có norm 2 như ở trên.
2. Khi $p = 1$ chúng ta có:

$$\|\mathbf{x}\|_1 = \|x_1\| + \|x_2\| + \dots \|x_n\| \quad (3)$$

là tổng các trị tuyệt đối của từng phần tử của \mathbf{x} . Norm 1 thường được dùng như xấp xỉ của norm 0 trong các bài toán có ràng buộc "sparse". Dưới đây là một ví dụ so sánh norm 1 và norm 2 trong không gian hai chiều:

Norm 2 (màu xanh) chính là đường thẳng "chim bay" nối giữa hai vector \mathbf{x} và \mathbf{y} . Khoảng cách norm 1 giữa hai điểm này (màu đỏ) có thể diễn giải như là đường đi từ \mathbf{x} tới \mathbf{y} trong một thành phố mà đường phố tạo thành hình bàn cờ. Chúng ta chỉ có cách đi dọc theo cạnh của bàn cờ mà không được đi thẳng.

3. Khi $p \rightarrow \infty$, ta có norm p chính là trị tuyệt đối của phần tử lớn nhất của vector đó:

$$\|\mathbf{x}\|_\infty = \max_{i=1,2,\dots,n} \|x_i\| \quad (4)$$

2.3 Chuẩn của ma trận

Với một ma trận $\mathbf{A} \in \mathbb{R}^{m \times n}$, chuẩn thường được dùng nhất là chuẩn Frobenius, ký hiệu là $\|\mathbf{A}\|_F$ là căn bậc hai của tổng bình phương tất cả các phần tử của ma trận đó.

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$$

3 Đạo hàm của hàm nhiều biến

Trong mục này, chúng ta sẽ giả sử rằng các đạo hàm tồn tại. Chúng ta sẽ xét hai trường hợp: i) Hàm số nhận giá trị là ma trận (vector) và cho giá trị là một số thực vô hướng; và ii) Hàm số nhận giá trị là một số vô hướng hoặc vector và cho giá trị là một vector.

3.1 Hàm cho giá trị là một số vô hướng

Đạo hàm (gradient) của một hàm số $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ theo vector \mathbf{x} được định nghĩa như sau:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n \quad (5)$$

trong đó $\frac{\partial f(\mathbf{x})}{\partial x_i}$ là đạo hàm của hàm số theo thành phần thứ i của vector \mathbf{x} . Đạo hàm này được lấy khi giả sử tất cả các biến còn lại là hằng số.

Nếu không có thêm biến nào trong hàm số, $\nabla_{\mathbf{x}} f(\mathbf{x})$ thường được viết gọn là $\nabla f(\mathbf{x})$.

Điều quan trọng cần nhớ: **đạo hàm của hàm số này là một vector có cùng chiều với vector đang lấy đạo hàm**. Tức nếu vector viết ở dạng cột thì đạo hàm cũng phải viết ở dạng cột.

Đạo hàm bậc hai (second-order gradient) của hàm số trên còn được gọi là *Hessian* và được định nghĩa như sau:

$$\nabla^2 f(\mathbf{x}) \triangleq \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix} \in \mathbb{S}^n \quad (6)$$

với $\mathbb{S}^n \in \mathbb{R}^{n \times n}$ là tập các ma trận vuông đối xứng có số cột là n .

Đạo hàm của một hàm số $f(\mathbf{X}) : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ **theo ma trận \mathbf{X}** được định nghĩa là:

$$\nabla f(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \frac{\partial f(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1m}} \\ \frac{\partial f(\mathbf{X})}{\partial x_{21}} & \frac{\partial f(\mathbf{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{n1}} & \frac{\partial f(\mathbf{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{nm}} \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (7)$$

Một lần nữa, đạo hàm của một hàm số theo ma trận là một ma trận có chiều giống với ma trận đó.

Hiểu một cách đơn giản, đạo hàm của một hàm số (có đầu ra là 1 số vô hướng) theo một ma trận được tính như sau. Trước tiên, tính đạo hàm của hàm số đó theo từng thành phần của ma trận *khi toàn bộ các thành phần khác được giả sử là hằng số*. Tiếp theo, ta *ghép* các đạo hàm thành phần tính được thành một ma trận đúng theo thứ tự như trong ma trận đó. Chú ý rằng vector là một trường hợp của ma trận.

Ví dụ: Xét hàm số: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(\mathbf{x}) = x_1^2 + 2x_1x_2 + \sin(x_1) + 2$.

Đạo hàm bậc nhất theo \mathbf{x} của hàm số đó là:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 + \cos(x_1) \\ 2x_1 \end{bmatrix}$$

Đạo hàm bậc hai theo \mathbf{x} , hay *Hessian* là:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 x_2} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 - \sin(x_1) & 2 \\ 2 & 0 \end{bmatrix}$$

Chú ý rằng *Hessian* luôn là một ma trận đối xứng.

3.2 Hàm cho giá trị là một vector

Những hàm số cho giá trị là một vector được gọi là *vector-valued function* trong tiếng Anh.

Giả sử một hàm số với **đầu vào là một số thực** $v(x) : \mathbb{R} \rightarrow \mathbb{R}^n$:

$$v(x) = \begin{bmatrix} v_1(x) \\ v_2(x) \\ \vdots \\ v_n(x) \end{bmatrix} \quad (8)$$

Đạo hàm của nó là một **vector hàng** như sau:

$$\nabla v(x) \triangleq \begin{bmatrix} \frac{\partial v_1(x)}{\partial x} & \frac{\partial v_2(x)}{\partial x} & \dots & \frac{\partial v_n(x)}{\partial x} \end{bmatrix} \quad (9)$$

Đạo hàm bậc hai của hàm số này có dạng:

$$\nabla^2 v(x) \triangleq \begin{bmatrix} \frac{\partial^2 v_1(x)}{\partial x^2} & \frac{\partial^2 v_2(x)}{\partial x^2} & \dots & \frac{\partial^2 v_n(x)}{\partial x^2} \end{bmatrix} \quad (10)$$

Ví dụ: Cho vector $\mathbf{a} \in \mathbb{R}^n$ và *vector-valued function* $v(x) = x\mathbf{a}$, thế thì:

$$\nabla v(x) = \mathbf{a}^T, \quad \nabla^2 v(x) = \mathbf{0} \in \mathbb{R}^{n \times n} \quad (11)$$

với $\mathbf{0}$ là ma trận với các thành phần đều là 0.

Xét một *vector-valued function* với **đầu vào là một vector** $h(\mathbf{x}) : \mathbb{R}^k \rightarrow \mathbb{R}^n$, đạo hàm bậc nhất của nó là:

$$\nabla h(\mathbf{x}) \triangleq \begin{bmatrix} \frac{\partial h_1(\mathbf{x})}{\partial x_1} & \frac{\partial h_2(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial h_n(\mathbf{x})}{\partial x_1} \\ \frac{\partial h_1(\mathbf{x})}{\partial x_2} & \frac{\partial h_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial h_n(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1(\mathbf{x})}{\partial x_k} & \frac{\partial h_2(\mathbf{x})}{\partial x_k} & \dots & \frac{\partial h_n(\mathbf{x})}{\partial x_k} \end{bmatrix} \quad (12)$$

$$= \begin{bmatrix} \nabla h_1(\mathbf{x}) & \nabla h_2(\mathbf{x}) & \dots & \nabla h_n(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{k \times n} \quad (13)$$

Một quy tắc dễ nhớ ở đây là nếu một hàm số $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ thì đạo hàm của nó là một ma trận thuộc $\mathbb{R}^{m \times n}$.

Đạo hàm bậc hai của hàm số trên là một *ma trận ba chiều*, tôi xin không đề cập ở đây.

<hr> Với các hàm số *matrix-valued* nhận giá trị đầu vào là ma trận, tôi cũng xin không đề cập ở đây. Tuy nhiên, ở phần dưới, khi tính toán đạo hàm cho các hàm cho giá trị là số

thực, chúng ta vẫn có thể sẽ sử dụng khái niệm này.

Trước khi đến phần tính đạo hàm của các hàm số thường gặp, chúng ta cần biết hai tính chất quan trọng khá giống với đạo hàm của hàm một biến được học trong chương trình cấp ba.

3.3 Hai tính chất quan trọng

3.3.1 Product rules

Để cho tổng quát, ta giả sử biến đầu vào là một ma trận (vector và số thực là các trường hợp đặt biệt của ma trận). Giả sử rằng các hàm số có chiều phù hợp để các phép nhân thực hiện được. Ta có:

$$\nabla (f(\mathbf{X})^T g(\mathbf{X})) = (\nabla f(\mathbf{X})) g(\mathbf{X}) + (\nabla g(\mathbf{X})) f(\mathbf{X}) \quad (14)$$

Biểu thức này giống như biểu thức chúng ta đã quá quen thuộc:

$$(f(x)g(x))' = f'(x)g(x) + g'(x)f(x)$$

Chú ý rằng với vector và ma trận, chúng ta không được sử dụng tính chất giao hoán.

3.3.2 Chain rules

Khi có các hàm hợp thì:

$$\nabla_{\mathbf{X}} g(f(\mathbf{X})) = \nabla_{\mathbf{X}} f^T \nabla_f g \quad (15)$$

Quy tắc này cũng giống với quy tắc trong hàm một biến:

$$(g(f(x)))' = f'(x)g'(f)$$

Nhắc lại rằng khi tính toán với ma trận, chúng ta cần chú ý tới chiều của các ma trận, và nhân ma trận không có tính chất giao hoán.

3.4 Đạo hàm của các hàm số thường gặp

3.4.1 $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$

Giả sử $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$, ta viết lại:

$$f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

Có thể nhận thấy rằng:

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = a_i, \quad \forall i = 1, 2, \dots, n$$

Vậy nên:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}$$

Thêm nữa, vì $\mathbf{a}^T \mathbf{x} = \mathbf{x}^T \mathbf{a}$ nên:

$$\nabla(\mathbf{x}^T \mathbf{a}) = \mathbf{a}$$

3.4.2 $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$

Đây là một *vector-valued function* $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ với $\mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m \times n}$. Giả sử rằng \mathbf{a}_i là hàng thứ i của ma trận \mathbf{A} . Ta có:

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \mathbf{x} \\ \mathbf{a}_2 \mathbf{x} \\ \vdots \\ \mathbf{a}_m \mathbf{x} \end{bmatrix}$$

Theo định nghĩa (13.2), và công thức (17), ta có thể suy ra:

$$\nabla_{\mathbf{x}}(\mathbf{A}\mathbf{x}) = [\mathbf{a}_1^T \quad \mathbf{a}_2^T \quad \dots \quad \mathbf{a}_m^T] = \mathbf{A}^T \quad (16)$$

Từ đây ta có thể suy ra đạo hàm của hàm số $f(\mathbf{x}) = \mathbf{x} = \mathbf{I}\mathbf{x}$, với \mathbf{I} là ma trận đơn vị với chiều phù hợp, là:

$$\nabla \mathbf{x} = \mathbf{I}$$

3.4.3 $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$

với $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n \times n}$. Áp dụng Product rules (14) ta có:

$$\begin{aligned} \nabla f(\mathbf{x}) &= \nabla ((\mathbf{x}^T) (\mathbf{A} \mathbf{x})) \\ &= (\nabla(\mathbf{x})) \mathbf{A} \mathbf{x} + (\nabla(\mathbf{A} \mathbf{x})) \mathbf{x} \\ &= \mathbf{I} \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x} \\ &= (\mathbf{A} + \mathbf{A}^T) \mathbf{x} \end{aligned} \tag{17}$$

Từ (17) và (16), ta có thể suy ra: $\nabla^2 \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T + \mathbf{A}$

Nếu \mathbf{A} là một ma trận đối xứng, ta sẽ có:

$$\nabla \mathbf{x}^T \mathbf{A} \mathbf{x} = 2\mathbf{A} \mathbf{x}, \quad \nabla^2 \mathbf{x}^T \mathbf{A} \mathbf{x} = 2\mathbf{A} \tag{18}$$

Nếu \mathbf{A} là ma trận đơn vị, tức $f(\mathbf{x}) = \mathbf{x}^T \mathbf{I} \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2^2$, ta có:

$$\nabla \|\mathbf{x}\|_2^2 = 2\mathbf{x}, \quad \nabla^2 \|\mathbf{x}\|_2^2 = 2\mathbf{I} \tag{19}$$

3.4.4 $f(\mathbf{x}) = \|\mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2$

Có hai cách tính đạo hàm của hàm số này:

Cách 1: Trước hết, biến đổi:

$$\begin{aligned} f(\mathbf{x}) &= \|\mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2 = (\mathbf{A} \mathbf{x} - \mathbf{b})^T (\mathbf{A} \mathbf{x} - \mathbf{b}) = (\mathbf{x}^T \mathbf{A}^T - \mathbf{b}^T) (\mathbf{A} \mathbf{x} - \mathbf{b}) \\ &= \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{b}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{b} \end{aligned}$$

Lấy đạo hàm cho từng số hạng rồi cộng lại ta có:

$$\nabla \|\mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2 = 2\mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{A}^T \mathbf{b} = 2\mathbf{A}^T (\mathbf{A} \mathbf{x} - \mathbf{b})$$

Cách 2: Dùng Chain rule. Sử dụng $\nabla(\mathbf{A} \mathbf{x} - \mathbf{b}) = \mathbf{A}^T$ và $\nabla \|\mathbf{x}\|_2^2 = 2\mathbf{x}$ và công thức chain rules (15), ta sẽ thu được kết quả tương tự.

3.4.5 $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} \mathbf{x}^T \mathbf{b}$

Bằng cách viết lại $f(\mathbf{x}) = (\mathbf{a}^T \mathbf{x})(\mathbf{x}^T \mathbf{b})$, ta có thể dùng Product rules (14) và ra kết quả:

$$\nabla(\mathbf{a}^T \mathbf{x} \mathbf{x}^T \mathbf{b}) = \mathbf{a} \mathbf{x}^T \mathbf{b} + \mathbf{b} \mathbf{a}^T \mathbf{x} = \mathbf{a} \mathbf{b}^T \mathbf{x} + \mathbf{b} \mathbf{a}^T \mathbf{x} = (\mathbf{a} \mathbf{b}^T + \mathbf{b} \mathbf{a}^T) \mathbf{x}$$

trong đây tôi đã sử dụng tính chất $\mathbf{y}^T \mathbf{z} = \mathbf{z}^T \mathbf{y}$ và tích của một số thực với một vector cũng bằng tích của vector và số thực đó.

3.5 Bảng các đạo hàm thường gặp

3.5.1 Cho vector

Table 1: Đạo hàm theo vector

$f(\mathbf{x})$	$\nabla f(\mathbf{x})$
$\mathbf{a}^T \mathbf{x}$	\mathbf{a}
$\mathbf{x}^T \mathbf{A} \mathbf{x}$	$\mathbf{A} + \mathbf{A}^T) \mathbf{x}$
$\mathbf{x}^T \mathbf{x} = \ \mathbf{x}\ _2^2$	$2\mathbf{x}$
$\ \mathbf{A} \mathbf{x} - \mathbf{b}\ _2^2$	$2\mathbf{A}^T (\mathbf{A} \mathbf{x} - \mathbf{b})$
$\mathbf{a}^T \mathbf{x}^T \mathbf{x} \mathbf{b}$	$2\mathbf{a}^T \mathbf{b} \mathbf{x}$
$\mathbf{a}^T \mathbf{x} \mathbf{x}^T \mathbf{b}$	$(\mathbf{a} \mathbf{b}^T + \mathbf{b} \mathbf{a}^T) \mathbf{x}$

3.5.2 Cho ma trận

Table 2: Đạo hàm theo ma trận

$f(\mathbf{X})$	$\nabla f(\mathbf{X})$
$\ \mathbf{X}\ _F^2$	$2\mathbf{X}$
$\mathbf{A} \mathbf{X}$	\mathbf{A}^T
$\ \mathbf{A} \mathbf{X} - \mathbf{B}\ _F^2$	$2\mathbf{A}^T (\mathbf{A} \mathbf{X} - \mathbf{B})$
$\ \mathbf{X} \mathbf{A} - \mathbf{B}\ _F^2$	$2(\mathbf{X} \mathbf{A} - \mathbf{B}) \mathbf{A}^T$
$\mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{b}$	$\mathbf{X}(\mathbf{a} \mathbf{b}^T + \mathbf{b} \mathbf{a}^T)$
$\mathbf{a}^T \mathbf{X} \mathbf{X}^T \mathbf{b}$	$(\mathbf{a} \mathbf{b}^T + \mathbf{b} \mathbf{a}^T) \mathbf{X}$,
$\mathbf{a}^T \mathbf{Y} \mathbf{X}^T \mathbf{b}$	$\mathbf{b} \mathbf{a}^T \mathbf{Y}$
$\mathbf{a}^T \mathbf{Y}^T \mathbf{X} \mathbf{b}$	$\mathbf{Y} \mathbf{a} \mathbf{b}^T$
$\mathbf{a}^T \mathbf{X} \mathbf{Y}^T \mathbf{b}$	$\mathbf{a} \mathbf{b}^T \mathbf{Y}$
$\mathbf{a}^T \mathbf{X}^T \mathbf{Y} \mathbf{b}$	$\mathbf{Y} \mathbf{b} \mathbf{a}^T$

3.6 Tài liệu tham khảo

[1] Matrix calculus