

VIP Refresher: Đại số và vi tích phân

Afshine AMIDI và Shervine AMIDI

Ngày 17 tháng 5 năm 2020

Dịch bởi Hoàng Minh Tuấn và Phạm Hồng Vinh

Kí hiệu chung

□ **Vectơ** – Chúng ta kí hiệu $x \in \mathbb{R}^n$ là một vectơ với n phần tử, với $x_i \in \mathbb{R}$ là phần tử thứ i :

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

□ **Ma trận** – Kí hiệu $A \in \mathbb{R}^{m \times n}$ là một ma trận với m hàng và n cột, $A_{i,j} \in \mathbb{R}$ là phần tử nằm ở hàng thứ i , cột j :

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Ghi chú: vectơ x được xác định ở trên có thể coi như một ma trận $n \times 1$ và được gọi là vectơ cột.

□ **Ma trận đơn vị** – Ma trận đơn vị $I \in \mathbb{R}^{n \times n}$ là một ma trận vuông với các phần tử trên đường chéo chính bằng 1 và các phần tử còn lại bằng 0:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Ghi chú: với mọi ma trận vuông $A \in \mathbb{R}^{n \times n}$, ta có $A \times I = I \times A = A$.

□ **Ma trận đường chéo** – Ma trận đường chéo $D \in \mathbb{R}^{n \times n}$ là một ma trận vuông với các phần tử trên đường chéo chính khác 0 và các phần tử còn lại bằng 0:

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

Ghi chú: chúng ta kí hiệu D là $\text{diag}(d_1, \dots, d_n)$.

Các phép toán ma trận

□ **Vectơ/vectơ** – Có hai loại phép nhân vectơ/vectơ:

- phép nhân inner: với $x, y \in \mathbb{R}^n$, ta có:

$$x^T y = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$

- phép nhân outer: với $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, ta có:

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m \times n}$$

□ **Ma trận/vectơ** – Phép nhân giữa ma trận $A \in \mathbb{R}^{m \times n}$ và vectơ $x \in \mathbb{R}^n$ là một vectơ có kích thước \mathbb{R}^m :

$$Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^n a_{c,i} x_i \in \mathbb{R}^m$$

với $a_{r,i}^T$ là các vectơ hàng và $a_{c,j}$ là các vectơ cột của A , và x_i là các phần tử của x .

□ **Ma trận/ma trận** – Phép nhân giữa ma trận $A \in \mathbb{R}^{m \times n}$ và $B \in \mathbb{R}^{n \times p}$ là một ma trận kích thước $\mathbb{R}^{m \times p}$:

$$AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^n a_{c,i} b_{r,i}^T \in \mathbb{R}^{m \times p}$$

với $a_{r,i}^T, b_{r,i}^T$ là các vectơ hàng và $a_{c,j}, b_{c,j}$ lần lượt là các vectơ cột của A và B .

□ **Chuyển vị** – Chuyển vị của một ma trận $A \in \mathbb{R}^{m \times n}$, kí hiệu A^T , khi các phần tử hàng cột hoán đổi vị trí cho nhau:

$$\forall i, j, \quad A_{i,j}^T = A_{j,i}$$

Ghi chú: với ma trận A, B , ta có $(AB)^T = B^T A^T$

□ **Nghịch đảo** – Nghịch đảo của ma trận vuông khả đảo A được kí hiệu là A^{-1} và chỉ tồn tại duy nhất:

$$AA^{-1} = A^{-1}A = I$$

Ghi chú: không phải tất cả các ma trận vuông đều khả đảo. Ngoài ra, với ma trận A, B , ta có $(AB)^{-1} = B^{-1}A^{-1}$

□ **Truy vết** – Truy vết của ma trận vuông A , kí hiệu $\text{tr}(A)$, là tổng của các phần tử trên đường chéo chính của nó:

$$\text{tr}(A) = \sum_{i=1}^n A_{i,i}$$

Ghi chú: với ma trận A, B , chúng ta có $\text{tr}(A^T) = \text{tr}(A)$ và $\text{tr}(AB) = \text{tr}(BA)$

□ **Định thức** – Định thức của một ma trận vuông $A \in \mathbb{R}^{n \times n}$, kí hiệu $|A|$ hay $\det(A)$ được tính đệ quy với $A_{\setminus i, \setminus j}$, ma trận A xóa đi hàng thứ i và cột thứ j :

$$\det(A) = |A| = \sum_{j=1}^n (-1)^{i+j} A_{i,j} |A_{\setminus i, \setminus j}|$$

Ghi chú: A khả đảo nếu và chỉ nếu $|A| \neq 0$. Ngoài ra, $|AB| = |A||B|$ và $|A^T| = |A|$.

Những tính chất của ma trận

□ **Phân rã đối xứng** – Một ma trận A đã cho có thể được biểu diễn dưới dạng các phần đối xứng và phần đối xứng của nó như sau:

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{Đối xứng}} + \underbrace{\frac{A - A^T}{2}}_{\text{Phần đối xứng}}$$

□ **Chuẩn** – Một chuẩn (norm) là một hàm $N : V \rightarrow [0, +\infty[$ mà V là một không gian vectơ, và với mọi $x, y \in V$, ta có:

- $N(x + y) \leq N(x) + N(y)$
- $N(ax) = |a|N(x)$ với a là một số
- nếu $N(x) = 0$, thì $x = 0$

Với $x \in V$, các chuẩn thường dùng được tổng hợp ở bảng dưới đây:

Chuẩn	Kí hiệu	Định nghĩa	Trường hợp dùng
Manhattan, L^1	$\ x\ _1$	$\sum_{i=1}^n x_i $	LASSO chính quy hóa
Euclidean, L^2	$\ x\ _2$	$\sqrt{\sum_{i=1}^n x_i^2}$	Ridge chính quy hóa
p -norm, L^p	$\ x\ _p$	$\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Hölder bất đẳng thức
Infinity, L^∞	$\ x\ _\infty$	$\max_i x_i $	Uniform convergence

□ **Sự phụ thuộc tuyến tính** – Một tập hợp các vectơ được cho là phụ thuộc tuyến tính nếu một trong các vectơ trong tập hợp có thể được biểu diễn bởi một tổ hợp tuyến tính của các vectơ khác.

Ghi chú: nếu không có vectơ nào có thể được viết theo cách này, thì các vectơ được cho là độc lập tuyến tính

□ **Hạng ma trận (rank)** – Hạng của một ma trận A kí hiệu $\text{rank}(A)$ và là số chiều của không gian vectơ được tạo bởi các cột của nó. Điều này tương đương với số cột độc lập tuyến tính tối đa của A .

□ **Ma trận bán xác định dương** – Ma trận $A \in \mathbb{R}^{n \times n}$ là bán xác định dương (PSD) kí hiệu $A \succeq 0$ nếu chúng ta có:

$$A = A^T \quad \text{và} \quad \forall x \in \mathbb{R}^n, \quad x^T A x \geq 0$$

Ghi chú: tương tự, một ma trận A được cho là xác định dương và được kí hiệu $A \succ 0$, nếu đó là ma trận PSD thỏa mãn cho tất cả các vectơ khác không x , $x^T A x > 0$.

□ **Giá trị riêng, vectơ riêng** – Cho ma trận $A \in \mathbb{R}^{n \times n}$, λ được gọi là giá trị riêng của A nếu tồn tại một vectơ $z \in \mathbb{R}^n \setminus \{0\}$, được gọi là vectơ riêng, sao cho:

$$Az = \lambda z$$

□ **Định lý phổ** – Cho $A \in \mathbb{R}^{n \times n}$. Nếu A đối xứng, thì A có thể chéo hóa bởi một ma trận trực giao thực $U \in \mathbb{R}^{n \times n}$. Bằng cách kí hiệu $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, chúng ta có:

$$\exists \Lambda \text{ đường chéo}, \quad A = U \Lambda U^T$$

□ **Phân tích giá trị suy biến** – Đối với một ma trận A có kích thước $m \times n$, Phân tích giá trị suy biến (SVD) là một kỹ thuật phân tích nhân tố nhằm đảm bảo sự tồn tại của đơn vị U $m \times m$, đường chéo Σ $m \times n$ và đơn vị V $n \times n$ ma trận, sao cho:

$$A = U \Sigma V^T$$

Giải tích ma trận

□ **Gradien** – Cho $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ là một hàm và $A \in \mathbb{R}^{m \times n}$ là một ma trận. Gradien của f đối với A là ma trận $m \times n$, được kí hiệu là $\nabla_A f(A)$, sao cho:

$$\left(\nabla_A f(A) \right)_{i,j} = \frac{\partial f(A)}{\partial A_{i,j}}$$

Ghi chú: gradien của f chỉ được xác định khi f là hàm trả về một số.

□ **Hessian** – Cho $f : \mathbb{R}^n \rightarrow \mathbb{R}$ là một hàm và $x \in \mathbb{R}^n$ là một vectơ. Hessian của f đối với x là một ma trận đối xứng $n \times n$, ghi chú $\nabla_x^2 f(x)$, sao cho:

$$\left(\nabla_x^2 f(x) \right)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Ghi chú: hessian của f chỉ được xác định khi f là hàm trả về một số.

□ **Các phép toán của gradien** – Đối với ma trận A, B, C , các thuộc tính gradien sau cần để lưu ý:

$$\nabla_A \text{tr}(AB) = B^T \quad \nabla_A f(A) = (\nabla_A f(A))^T$$

$$\nabla_A \text{tr}(ABA^T C) = CAB + C^T AB^T$$

$$\nabla_A |A| = |A|(A^{-1})^T$$