Mục lục

Chương 1

Ma trận

1.1 Định thức của ma trận

Trong các bài viết của mình thì vector sẽ được ký hiệu bởi chữ thường in đậm (ví dụ $\boldsymbol{v},\ \boldsymbol{x}, \cdots$); ma trận sẽ được ký hiệu bởi chữ hoa in đậm (ví dụ $\boldsymbol{A},\ \boldsymbol{B}, \cdots$); các đại lượng vô hướng (số) được ký hiệu bởi chữ thường không in đậm (ví dụ $x_1,\ N,\ t, \cdots$).

Ví dụ vector $\boldsymbol{a} = (x_1, x_2, \cdots, x_n)$ trong đó x_i là các tọa độ.

Ví dụ ma trận $\boldsymbol{A} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$.

Định nghĩa 1. Nghịch thế

Cho tập hợp $A = \{1, 2, \cdots, n\}$ và xét hoán vị σ trên A. Ta gọi hai phần tử i và j tạo thành **nghịch thế** (inversion) nếu $i < j$ và $\sigma(i) > \sigma(j)$.

Đặt $\sigma = \{k_1, k_2, \cdots, k_n\}$ là một hoán vị của A. Ta ký hiệu

$$P\{k_1, k_2, \cdots, k_n\}$$

là số lượng nghịch thế của σ và đặt

$$(-1)^{P\{k_1,k_2,\cdots,k_n\}} = \text{sign}\{k_1, k_2, \cdots, k_n\}.$$

Ví dụ với $n = 4$, $A = \{1, 2, 3, 4\}$. Xét hoán vị $\sigma = \{4, 2, 1, 3\}$.

Ta nhận thấy các cặp nghịch thế $(4, 2),\ (4, 1),\ (4, 3),\ (2, 1)$ gồm 4 cặp nghịch thế. Vậy $P\{4, 2, 1, 3\} = 4$ và $\text{sign}\{4, 2, 1, 3\} = (-1)^4 = 1$.

Định nghĩa 2. Định thức

Khi đó định thức của ma trận $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ được định nghĩa là:

$$\det(\boldsymbol{A}) = \sum_{(i_1,i_2,\cdots,i_n)} a_{1,i_1} \cdot a_{2,i_2} \cdot a_{n,i_n} \cdot \text{sign}\{i_1, i_2, \cdots, i_n\} \qquad (1.1)$$

với mọi hoán vị $(i_1, i_2, \cdots, i_n)$ của $(1, 2, \cdots, n)$. Như vậy có $n!$ phần tử cho tổng trên.

Ví dụ: tính định thức ma trận $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

Xét hoán vị $\sigma_1 = \{1, 2, 3\}$. Khi đó $P\{1, 2, 3\} = 0$, $a_{11} \cdot a_{22} \cdot a_{33} \cdot (-1)^0 = 1 \cdot 5 \cdot 9 \cdot 1 = 45$.

Xét hoán vị $\sigma_2 = \{1, 3, 2\}$. Khi đó $P\{1, 3, 2\} = 1$, $a_{11} \cdot a_{23} \cdot a_{32} \cdot (-1)^1 = 1 \cdot 6 \cdot 8 \cdot (-1) = -48$.

Xét hoán vị $\sigma_3 = \{2, 1, 3\}$. Khi đó $P\{2, 1, 3\} = 1$, $a_{12} \cdot a_{21} \cdot a_{33} \cdot (-1)^1 = 2 \cdot 4 \cdot 9 \cdot (-1) = -72$.

Xét hoán vị $\sigma_4 = \{2, 3, 1\}$. Khi đó $P\{2, 3, 1\} = 2$, $a_{12} \cdot a_{23} \cdot a_{31} \cdot (-1)^2 = 2 \cdot 6 \cdot 7 \cdot 1 = 84$.

Xét hoán vị $\sigma_5 = \{3, 1, 2\}$. Khi đó $P\{3, 1, 2\} = 2$, $a_{13} \cdot a_{21} \cdot a_{32} \cdot (-1)^2 = 3 \cdot 4 \cdot 8 \cdot 1 = 96$.

Xét hoán vị $\sigma_6 = \{3, 2, 1\}$. Khi đó $P\{3, 2, 1\} = 3$, $a_{13} \cdot a_{22} \cdot a_{31} \cdot (-1)^3 = 3 \cdot 5 \cdot 7 \cdot (-1) = -105$.

Như vậy $\det(A) = 45 - 48 - 72 + 84 + 96 - 105 = 0$.

Định thức của ma trận còn được định nghĩa như sau:

Với ma trận 1×1 là $\boldsymbol{A} = \begin{pmatrix} a_{11} \end{pmatrix}$ thì $\det(\boldsymbol{A}) = a_{11}$.

Với ma trận 2×2 là $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ thì $\det(\boldsymbol{A}) = a_{11}a_{22} - a_{21}a_{12}$.

Với ma trận $n \times n$, gọi $\boldsymbol{M}_{ij}$ là ma trận có được từ ma trận $\boldsymbol{A}$ khi bỏ đi hàng i và cột j của ma trận $\boldsymbol{A}$ và ký hiệu $A_{ij} = (-1)^{i+j} \det(\boldsymbol{M}_{ij})$. Khi đó:

Định lí 1. Định lý Laplace

Định lý Laplace cho phép ta khai triển định thức của ma trận cấp n thành tổng các ma trận cấp $n-1$.

Khai triển theo cột j:

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ij}A_{ij} = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj},\ \ j = \overline{1,n}.$$

Khai triển theo hàng i:

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} a_{ij}A_{ij} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in},\ \ i = \overline{1,n}.$$

1.2 Ma trận nghịch đảo

Ma trận $\boldsymbol{A}^{-1}$ được gọi là **ma trận nghịch đảo** của ma trận vuông $\boldsymbol{A}$ nếu $\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{I}$. Trong đó $\boldsymbol{I}$ là ma trận đơn vị cùng kích thước với $\boldsymbol{A}$.

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}[(A_{ij})_n]^T = \frac{1}{\det(\boldsymbol{A})}\begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix} \tag{1.2}$$

Như vậy, điều kiện cần và đủ để một ma trận có nghịch đảo là định thức khác 0.

1.3 Hạng của ma trận

Định nghĩa 3. Hạng của ma trận

Cho ma trận $\mathbf{A}_{m\times n}$. **Hạng** của ma trận là cấp của ma trận con lớn nhất có định thức khác 0. Nghĩa là một ma trận vuông mà là ma trận con (lấy 1 phần của ma trận gốc) kích thước $r \times r$ mà có định thức khác 0, thì hạng của ma trận khi đó là r. Dễ thấy do là ma trận con, và vuông, nên $r \leqslant \min(m,n)$.

Ví dụ, ma trận $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 4 \end{pmatrix}$ có định thức $\det(\boldsymbol{A}) = 0$. Nhưng ma trận

con của $\boldsymbol{A}$ là $\boldsymbol{B} = \begin{pmatrix} 2 & 3 \\ 2 & 4 \end{pmatrix}$ (lấy dòng 1 và 3, lấy cột 2 và 3) có định thức $\det(\boldsymbol{B}) = 2 \neq 0$, do đó $r = \text{rank}(\boldsymbol{A}) = 2$ ($\text{rank}(\boldsymbol{A})$ nghĩa là hạng của $\boldsymbol{A}$).

Chương 2

Discrete logarithm

2.1 Các thuật toán tính discrete logarithm

Thuật toán Baby-Step-Giant-Step (BSGS) giúp tính discrete logarithm trên nhóm cyclic với order là số nguyên tố 1.

Algorithm 1 Thuật toán Baby-Step-Giant-Step

Require: Nhóm cyclic G có order n, generator g và phần tử $h \in G$.
Ensure: Số x duy nhất thuộc $\{0, 1, \ldots, n-1\}$ thỏa $g^x = h$. $m \leftarrow \lfloor \sqrt{n} \rfloor$

1: **for** $j = 0 \rightarrow m - 1$ **do**
2: Tính g^j. Lưu (j, g^j) vào bảng.
3: **end for**
4: Tính g^{-m}.
5: $\gamma \leftarrow h$.
6: **for** $i = 0 \rightarrow m - 1$ **do**
7: a) Kiểm tra điều kiện $\gamma = g^j$ với $j = 0, 1, \ldots, m-1$.
8: b) Nếu điều kiện thỏa, trả về $im + j$.
9: c) Nếu không, đặt $\gamma \leftarrow \gamma \cdot g^{-m}$.
10: **end for**

Khi order của cyclic group là lũy thừa một số nguyên tố thì ta dùng thuật toán Pohlig-Hellman 2.

Algorithm 2 Thuật toán Pohlig-Hellman

Require: Nhóm cyclic G có order $n = p^e$, generator g và phần tử $h \in G$.
Ensure: Số x duy nhất thuộc $\{0, 1, \ldots, n-1\}$ thỏa $g^x = h$.

1: Khởi tạo $x_0 = 0$.
2: Tính $\gamma = g^{p^{e-1}}$. Theo định lý Lagrange, γ có order là p.
3: **for** $k = 0 \to e - 1$ **do**
4: a) Tính $h_k = (g^{-x_k} \cdot h)^{e-1-k}$.
5: b) Sử dụng thuật toán baby-step-giant-step, tìm $d_k \in \{0, 1, \ldots, p-1\}$ sao cho $\gamma^{d_k} = h_k$.
6: c) Tính $x_{k+1} = x_k + p^k d_k$.
7: **end for**
8: Trả về x_e là kết quả cần tìm.

Chương 3

Số học

3.1 Thặng dư chính phương

Định nghĩa 1. Số chính phương modulo p

Xét số dương nguyên tố lẻ p. Số a được gọi là **số chính phương modulo** p nếu $(a, m) = 1$ và tồn tại số x sao cho $x^2 = a \pmod{p}$.

Nói cách khác phương trình đồng dư $x^2 \equiv a \pmod{p}$ có nghiệm.

Chúng ta sử dụng kí hiệu Legendre (Legendre's symbol) để thể hiện một số a có phải là số chính phương modulo nguyên tố p không.

Định nghĩa 2. Legendre's symbol

Xét p là số nguyên tố, a là số nguyên không chia hết cho p. Khi đó kí hiệu Legendre được định nghĩa là

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{nếu } a \text{ là số chính phương modulo } p. \\ -1, & \text{nếu ngược lại.} \end{cases} \tag{3.1}$$

Chương 4

Lattice-based crypto

4.1 Thuật toán GGH

Phần này tham khảo trong [3]

Trong thuật toán GGH, ta chọn số nguyên tố q làm public parameter.

Sau đó chọn hai số f và g làm secret key. Hai số này phải thỏa mãn các điều kiện

$$f < \sqrt{q/2}, \quad \sqrt{q/4} < g < \sqrt{q/2}, \quad \gcd(f, qg) = 1$$

Tính $h = f^{-1}g \pmod{q}$. Khi đó public key là h.

Encryption. Để encrypt message m với số random r thỏa mãn

$$0 < m < \sqrt{q/4}, \quad 0 < r < \sqrt{q/2}$$

Ta tính $e = rh + m \pmod{q}$ là ciphertext với $0 < e < q$.

Decryption. Để decrypt ciphertext e ta tính

$$a = fe \pmod{q}, \quad b = f^{-1}a \pmod{g}$$

Lưu ý f^{-1} là nghịch đảo modulo g. Khi đó $b \equiv m$ là message ban đầu.

Chứng minh. Để chứng minh rằng số b sau khi tính toán bằng chính xác m ban đầu ta cần xem xét điều kiện của secret key và public key.

Đầu tiên ta có

$$a \equiv fe \equiv f(rh + m) = f(rf^{-1}g + m) = rg + fm \pmod{q}$$

Từ điều kiện của f, g, r và m ta có

$$rg + fm < \sqrt{\frac{q}{2}} \cdot \sqrt{\frac{q}{2}} + \sqrt{\frac{q}{2}} \cdot \sqrt{\frac{q}{4}} < q$$

Nói cách khác $rg + fm$ giữ nguyên giá trị trong phép modulo q, hay $a \equiv rg + fm$.

Suy ra $b = f^{-1}a = f^{-1}(rg + fm) = m \pmod{g}$ (giá trị a không thay đổi khi chuyển từ modulo q sang modulo g). Do $0 < m < \sqrt{q/4}$ và $\sqrt{q/4} < g < \sqrt{q/2}$ nên $m < g$. Nói cách khác b bằng đúng m ban đầu. □

Để tấn công hệ mật mã này ta xây dựng lattice. Để ý rằng $h = f^{-1}g \pmod{q}$, hay $fh + kq = g$ với $k \in \mathbb{Z}$.

Ta thấy rằng $f \cdot (h, 1) + k \cdot (q, 0) = (g, f)$. Như vậy lattice gồm hai vector $(h, 1)$ và $(q, 0)$. Thuật toán tối giản Gauss sẽ hoạt động trong trường hợp này (số chiều bằng 2).

Chương 5

Bài toán đếm Polya

5.1 Lớp tương đương

Xét nhóm G và tập hợp M. Khi đó hai phần tử m và n thuộc M được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $m = gn$.

Nhận xét 1

Quan hệ giữa các phần tử như trên là quan hệ tương đương.

Chứng minh. Để chứng minh một quan hệ là tương đương, ta cần chứng minh tính phản xạ, đối xứng và bắc cầu.

Đối với tính phản xạ, mọi vector đều có quan hệ với chính nó qua phần tử đơn vị $e \in G$.

Đối với tính đối xứng, nếu m có quan hệ với n thì tồn tại $g \in G$ sao cho $m = gn$. Theo tính chất nhóm thì tồn tại phần tử g^{-1} là nghịch đảo của g trong G. Do đó $g^{-1}m = n$. Nói cách khác n cũng có quan hệ với m. Như vậy ta có tính đối xứng.

Đối với tính bắc cầu, nếu m có quan hệ với n thì tồn tại $g_1 \in G$ sao cho $m = g_1 n$. Tiếp theo, nếu n có quan hệ với p thì tồn tại $g_2 \in G$ sao cho $n = g_2 p$. Suy ra $m = g_1 n = g_1(g_2 p) = (g_1 g_2)p$. Do $g_1, g_2 \in G$ nên $g_1 g_2 \in G$. Như vậy m cũng có quan hệ với p nên quan hệ có tính bắc cầu.

Vậy quan hệ được định nghĩa như trên là quan hệ tương đương. □

5.2 Tác động nhóm lên vector

Xét nhóm G và không gian vector $\mathbb{F}_2^n$, $n \in \mathbb{N}$. Khi đó hai vector $\boldsymbol{x}$ và $\boldsymbol{y}$ thuộc $\mathbb{F}_2^n$ được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $\boldsymbol{x} = g\boldsymbol{y}$.

Ví dụ, xét nhóm hoán vị $\mathcal{S}_3$. Giả sử các vector trong $\mathbb{F}_2^3$ có dạng

$$\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{F}_2^3.$$

Khi đó vector $(1,0,0)$ có quan hệ với $(0,0,1)$ với hoán vị $(1,3)(2)$. Cụ thể là $(x_1, x_2, x_3) \xrightarrow{(1,3)(2)} (x_3, x_2, x_1)$.

Tương tự, vector $(1,0,0)$ cũng có quan hệ với $(0,1,0)$ với hoán vị $(1,2)(3)$. Thêm nữa, vector $(1,0,0)$ có quan hệ với chính nó qua hoán vị đồng nhất $(1)(2)(3)$.

Trong môn toán rời rạc ta đã biết, nếu một tập có quan hệ tương đương thì ta có thể phân các phần tử của tập đó vào các lớp tương đương rời nhau. Nghĩa là nếu hai phần tử có quan hệ với nhau thì vào cùng một lớp tương đương. Từ phần trên ta đã biết rằng dưới tác động của nhóm, các phần tử trong tập hợp bất kì sẽ phân bổ thành các lớp tương đương.

Câu hỏi đặt ra là, có bao nhiêu lớp tương đương như vậy?

Để giải quyết vấn đề này ta sử dụng bổ đề Burnside.

Nhóm $\mathcal{S}_3$ có các hoán vị

$$\mathcal{S}_3 = \{(1)(2)(3), (1,2)(3), (1,3)(2), (2,3)(1), (1,3,2), (1,2,3)\}$$

Lần lượt xét từng hoán vị. Đầu tiên, với $(1)(2)(3)$ thì các phần tử trong vector đứng yên. Do đó dưới tác động của hoán vị này, x_1 biến thành x_1, x_2 biến thành x_2 và x_3 biến thành x_3. Số cách chọn cho mỗi x_i là 2 nên theo quy tắc nhân ta có $2^3 = 8$ cách.

Tiếp theo, với hoán vị $(1,2)(3)$ thì $x_1 \to x_2$, $x_2 \to x_1$ và $x_3 \to x_3$. Do đó x_1 và x_2 có cùng giá trị nên có 2 cách chọn, x_3 cũng có 2 cách chọn nên tổng số cách là $2 \cdot 2 = 4$. Hoán vị $(1,3)(2)$ và $(2,3)(1)$ tương tự.

Với hoán vị $(1,2,3)$ thì $x_1 \to x_2$, $x_2 \to x_3$ và $x_3 \to x_1$ nên $x_1 = x_2 = x_3$, có 2 cách chọn trong trường hợp này. Hoán vị $(1,3,2)$ tương tự.

Như vậy, theo bổ đề Burnside, số lớp tương đương các vector trong $\mathbb{F}_2^3$ là

$$t(\mathcal{S}_3) = \frac{1}{6}(1 \cdot 2^3 + 3 \cdot 2^2 + 2 \cdot 2) = 4$$

Thật vậy, ta có thể chia các vector thành 4 lớp tương đương là $\{000\}$, $\{001, 010, 011\}$, $\{011, 101, 110\}$, $\{111\}$.

Ngoài nhóm $\mathcal{S}_3$ ra còn các nhóm khác cũng tác động lên các vector. Một số nhóm hay được sử dụng là:

1. Nhóm general linear: gồm các ma trận khả nghịch $n \times n$ trên $\mathbb{F}_2$. Tác động nhóm lúc này là phép nhân ma trận $\boldsymbol{A} \in GL(n,2)$ với vector $\boldsymbol{x} \in \mathbb{F}_2^n$, hay $\boldsymbol{A} \cdot \boldsymbol{x}$.

2. Nhóm general affine: gồm các ma trận khả nghịch $n \times n$ trên $\mathbb{F}_2$ và vector bất kì trong $\mathbb{F}_2^n$. Tác động nhóm lúc này là biến đổi affine $\boldsymbol{A} \cdot \boldsymbol{x} + \boldsymbol{b}$ với $\boldsymbol{A} \in GL(n, 2)$ và $\boldsymbol{b} \in \mathbb{F}_2^n$.

Cần nhắc lại một chút, số lượng phần tử của nhóm $GL(n, 2)$ là

$$(2^n - 1) \cdot (2^n - 2) \cdots (2^n - 2^{n-1})$$

Khi $n = 3$ thì $|GL(3, 2)| = (2^3 - 1) \cdot (2^3 - 2) \cdot (2^3 - 4) = 168$ ma trận.

5.3 Tác động nhóm lên hàm boolean

Ta tiếp tục xét nhóm G và không gian vector $\mathbb{F}_2^n$, $n \in \mathbb{N}$. Khi đó hai hàm boolean n biến $f(x_1, \ldots, x_n)$ và $g(x_1, \ldots, x_n)$ được gọi là **quan hệ với nhau** nếu tồn tại $g \in G$ mà $g(\boldsymbol{x}) = f(g\boldsymbol{x})$ với mọi $\boldsymbol{x} \in \mathbb{F}_2^n$.

Ta cũng xét hoán vị $\mathcal{S}_3$. Ta cũng lần lượt xét các phần tử của nhóm.

Đặt $f_0, f_1, \ldots, f_7$ lần lượt là các giá trị hàm f với các vector $\boldsymbol{x} \in \mathbb{F}_2^3$.

Đầu tiên, với $(1)(2)(3)$, ta có bảng chuyển vector như sau

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

$\xrightarrow{(1)(2)(3)}$

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_1$, ..., $f_7 \to f_7$ nên có 8 chu trình. Vậy số lượng cách chọn là 2^8.

Tiếp theo, xét các hoán vị dạng $(1)(2, 3)$, ta có bảng chuyển vector như sau

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_2 \to f_1$, $f_3 \to f_3$, $f_4 \to f_4$, $f_5 \to f_6 \to f_5$, $f_7 \to f_7$. Ở đây có 6 chu trình nên số cách chọn là 2^6.

Tiếp theo ta xét các hoán vị dạng $(1, 2, 3)$.

Ta thấy rằng $f_0 \to f_0$, $f_1 \to f_2 \to f_4 \to f_1$, $f_3 \to f_6 \to f_5 \to f_3$, $f_7 \to f_7$ nên ở đây có 4 chu trình. Số cách chọn là 2^4.

$(1)(2,3)$

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

$\longrightarrow$

x_1	x_3	x_2	f
0	0	0	f_0
0	1	0	f_2
0	0	1	f_1
0	1	1	f_3
1	0	0	f_4
1	1	0	f_6
1	0	1	f_5
1	1	1	f_7

$(1,2,3)$

x_1	x_2	x_3	f
0	0	0	f_0
0	0	1	f_1
0	1	0	f_2
0	1	1	f_3
1	0	0	f_4
1	0	1	f_5
1	1	0	f_6
1	1	1	f_7

$\longrightarrow$

x_2	x_3	x_1	f
0	0	0	f_0
0	1	0	f_2
1	0	0	f_4
1	1	0	f_6
0	0	1	f_1
0	1	1	f_3
1	0	1	f_5
1	1	1	f_7

Như vậy theo bổ đề Burnside, số lớp hàm bool tương đương dưới tác động của nhóm $\mathcal{S}_3$ là

$$t(\mathcal{S}_3) = \frac{1}{6}(2^8 + 3 \cdot 2^6 + 2 \cdot 2^4) = 80.$$

Chương 6

Giải tích

Định nghĩa 1. Dãy Cauchy

Dãy (x_n) được gọi là dãy Cauchy nếu với mọi $\varepsilon > 0$, tồn tại $N_0 \in \mathbb{N}$ sao cho, với mọi $m, n > N_0$ thì $|x_m - x_n| < \varepsilon$.

Định lí 1. Tiêu chuẩn Cauchy

Dãy số (x_n) có giới hạn hữu hạn khi và chỉ khi nó là dãy Cauchy.

Định lí 2. Bổ đề Fermat

Cho f là một hàm số có đạo hàm trên (a, b). Nếu $x_0 \in (a, b)$ là một điểm cực trị của f thì ta có $f'(x_0) = 0$.

Chứng minh. Ta chứng minh trong trường hợp x_0 là điểm cực tiểu. Trường hợp điểm cực đại tương tự.

Hàm f có đạo hàm trên (a, b) nên tại điểm x_0 nó có đạo hàm bên trái và đạo hàm bên phải, và hai đạo hàm này bằng nhau.

Ta có $f'(x_0^+) = \lim\limits_{x \to x_0^+} \dfrac{f(x) - f(x_0)}{x - x_0}$. Vì $x \to x_0^+$ nghĩa là $x > x_0$ (x tiến tới x_0 từ bên phải), và do x_0 là cực tiểu $f(x) - f(x_0) \geqslant 0$ nên phân số dưới dấu giới hạn lớn hơn 0. Suy ra $f'(x_0^+) \geqslant 0$.

Hoàn toàn tương tự ta chứng minh được $f'(x_0^-) \leqslant 0$. Và do $f'(x_0^+) = f'(x_0^-) = f'(x_0)$ nên $f'(x_0) = 0$.

Ta có điều phải chứng minh. □

Định lí 3. Định lí Rolle

Xét hàm số f liên tục trên đoạn $[a, b]$, có đạo hàm trên khoảng (a, b) và $f(a) = f(b)$. Khi đó tồn tại c thuộc (a, b) sao cho $f'(c) = 0$.

Định lí 4. Định lí Lagrange

Xét hàm số f liên tục trên đoạn $[a, b]$, có đạo hàm trên khoảng (a, b). Khi đó tồn tại c thuộc (a, b) sao cho $f'(c)(b - a) = f(b) - f(a)$.

Định nghĩa 2. Hàm lõm

Hàm số f liên tục trên khoảng $\mathbb{I}$ nếu với mọi α, β mà $\alpha + \beta = 1$ ta đều có

$$f(\alpha x + \beta y) \leqslant \alpha f(x) + \beta f(y), \quad \forall x, y \in \mathbb{I} \tag{6.1}$$

Chương 7

Quantum computing

7.1 Introduction

Toán tử NOT được biểu diễn dưới dạng ma trận $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Khi đó $NOT(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle$.

Nếu cho $\alpha = \beta = 1$ và $x \in \{0, 1\}$ thì $NOT|x\rangle = |x \oplus 1\rangle$.

Toán tử Hadarmard được biểu diễn dưới dạng ma trận

$$\mathcal{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Khi đó

$$\mathcal{H}(\alpha|0\rangle + \beta|1\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha + \beta \\ \alpha - \beta \end{pmatrix} = \frac{\alpha + \beta}{\sqrt{2}}|0\rangle + \frac{\alpha - \beta}{\sqrt{2}}|1\rangle$$

Nếu $\alpha = 1, \beta = 0$ thì $\mathcal{H}(|0\rangle) = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$.

Nếu $\alpha = 0, \beta = 1$ thì $\mathcal{H}(|1\rangle) = \frac{1}{\sqrt{2}}|0\rangle + \frac{-1}{\sqrt{2}}|1\rangle$.

Tổng hợp lại, nếu $x \in \{0, 1\}$ thì $\mathcal{H}|x\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{(-1)^x}{\sqrt{2}}|1\rangle$.

Ta thấy rằng toán tử ngược của toán tử Hadamard là chính nó.

Đố vui: nếu ta có qubit là $\alpha|0\rangle + \beta|1\rangle$, hãy xây dựng mạch logic để biến đổi qubit trên thành $\alpha|000\rangle + \beta|111\rangle$.

Giải: sử dụng toán tử Hadamard và NOT. Ta có

$$NOT(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle$$

và

$$\mathcal{H}(\alpha|0\rangle + \beta|1\rangle) = \frac{1}{\sqrt{2}}(\alpha + \beta)|0\rangle + \frac{1}{\sqrt{2}}(\alpha - \beta)|1\rangle$$

Toán tử $CNOT$ được biểu diễn bởi ma trận

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Giả sử ta có hai qubit là $|\alpha\rangle = x|0\rangle + y|1\rangle$ và $|\beta\rangle = z|0\rangle + t|1\rangle$.
Khi đó $|\alpha\rangle \otimes |\beta\rangle = xz|00\rangle + xt|01\rangle + yz|10\rangle + yt|11\rangle$.
Qua toán tử $CNOT$ ta có

$$\begin{aligned} CNOT(|\alpha\rangle \otimes |\beta\rangle) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} xz \\ xt \\ yz \\ yt \end{pmatrix} = \begin{pmatrix} xz \\ xt \\ yt \\ yz \end{pmatrix} \\ &= xz|00\rangle + xt|01\rangle + yt|10\rangle + yz|11\rangle \end{aligned}$$

Trường hợp $|\alpha\rangle = |0\rangle$ thì $x = 1, y = 0$. Khi đó ta có tương đương

$$CNOT(|0\rangle \otimes |\beta\rangle) = z|00\rangle + t|01\rangle = |0\rangle \otimes (z|0\rangle + t|1\rangle) = |0\rangle \otimes |\beta\rangle.$$

Trường hợp $|\alpha\rangle = |1\rangle$ thì $x = 0, y = 1$. Khi đó ta có tương đương

$$CNOT(|1\rangle) \otimes |\beta\rangle) = t|10\rangle + z|11\rangle = |1\rangle \otimes (t|0\rangle + z|1\rangle) = |1\rangle \otimes NOT(|\beta\rangle).$$

Nói cách khác, nếu $x \in \{0, 1\}$ thì

$$CNOT(|x\rangle \otimes |\beta\rangle) = \begin{cases} |x\rangle \otimes |\beta\rangle, & \text{nếu } x = 0 \\ |x\rangle \otimes NOT(|\beta\rangle), & \text{nếu } x = 1 \end{cases}$$

Do đó các toán tử có ma trận $\begin{pmatrix} I_n & \mathcal{O} \\ \mathcal{O} & \mathcal{U} \end{pmatrix}$ được gọi là toán tử kiểm soát (controlled).

Một trường hợp riêng nữa là khi $\beta = 0$ hoặc $\beta = 1$. Khi đó, với toán tử NOT bên trên ta suy ra

$$CNOT(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |x \oplus y\rangle$$

Chương 8

Lý thuyết xác suất

8.1 Định nghĩa xác suất

Định nghĩa 1. Định nghĩa cổ điển của xác suất

Định nghĩa thống kê của xác suất nói rằng, giả sử trong một phép thử có n khả năng có thể xảy ra. Xét một biến cố A xảy ra khi thực hiện phép thử có k khả năng xảy ra. Khi đó xác suất của biến cố A ký hiệu là $P(A)$ và được tính

$$P(A) = \frac{k}{n}$$

Dễ thấy, do biến cố A là một trường hợp nhỏ trong tổng thể tất cả trường hợp khi thực hiện phép thử, do đó $0 \leq k \leq n$. Nghĩa là

$$0 \leq P(A) \leq 1$$

với mọi biến cố A bất kì.

Ví dụ 1. *Xét phép thử tung hai đồng xu. Gọi A là biến cố hai đồng xu cùng mặt.*

Ta ký hiệu S là đồng xu sấp, N là đồng xu ngửa. Khi đó các trường hợp có thể xảy ra của phép thử là $S-S$, $S-N$, $N-S$, $N-N$ (4 trường hợp).

Trong khi đó, các trường hợp có thể xảy ra của biến cố A là $S-S$, $N-N$ (2 trường hợp).

Kết luận: $P(A) = \frac{2}{4} = \frac{1}{2}$

*Chúng ta gọi tập hợp tất cả các trường hợp khi thực hiện phép thử là **không gian mẫu** và ký hiệu là Ω. Mỗi phần tử trong không gian mẫu được gọi là **biến cố sơ cấp**. Trong bài này, $\Omega = \{S-S, S-N, N-S, N-N\}$.*

*Tập hợp các trường hợp có thể xảy ra của biến cố gọi là **không gian biến cố** và ký hiệu là Ω_A. Trong bài này $\Omega_A = \{S-S, N-N\}$.*

Như vậy, $P(A) = \frac{|\Omega_A|}{|\Omega|}$

Ví dụ 2. *Tung hai con súc sắc cân đối và đồng chất. Tính xác suất tổng số nút của hai con súc sắc bằng 4.*

Việc tung mỗi con súc sắc có 6 trường hợp. Do đó $|\Omega| = 6^2 = 36$

Gọi A là biến cố tổng số nút của hai con súc sắc bằng 4. Ta có các trường hợp là $4 = 1 + 3 = 3 + 1 = 2 + 2$ *(3 trường hợp).*

Như vậy $|\Omega_A| = 3$ *và* $P(A) = \frac{3}{36} = \frac{1}{12}$

Định nghĩa 2. Biến cố xung khắc

Hai biến cố được gọi là **xung khắc** nếu biến cố này xảy ra thì biến cố kia chắc chắn không xảy ra. Nói cách khác giao của chúng bằng rỗng.

Khi đó, nếu A và B là hai biến cố xung khắc,

$$P(A + B) = P(A) + P(B)$$

Ta còn có thể ký hiệu $P(A + B)$ là $P(A \cup B)$ (hợp hai biến cố).

Định nghĩa 3. Biến cố độc lập

Hai biến cố được gọi là **độc lập** nếu việc xảy ra của biến cố này không ảnh hưởng đến việc xảy ra của biến cố kia.

Khi đó, nếu A và B là hai biến cố độc lập thì

$$P(AB) = P(A)P(B)$$

8.2 Xác suất có điều kiện

Xét hai tập hợp A và B. Số phần tử của phép hợp hai tập hợp trong trường hợp tổng quát được tính như sau:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Tương tự, xác suất của phép cộng xác suất đối với hai biến cố có giao khác rỗng là:

$$P(A + B) = P(A) + P(B) - P(A \cap B)$$

Xét các tập hợp A_1, A_2, ..., A_n. Khi đó, số phần tử khi hợp các tập hợp này là:

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_n| = |A_1| + |A_2| + \cdots + |A_n| - \sum_{i,j} |A_i \cap A_j| \\ + \sum_{i,j,k} |A_i \cap A_j \cap A_k| + \cdots \\ = \sum_{i=1}^{n} (-1)^{i+1} \sum_{j_1, j_2, \cdots, j_i} |A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i}|\end{aligned}$$

Tương tự, ta có phép cộng xác suất:

Định lí 1. Phép cộng xác suất mở rộng

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} (-1)^{i+1} \sum_{j_1, j_2, \cdots, j_i} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i})$$

Định lí 2. Xác suất có điều kiện

Xét hai biến cố A và B. Khi đó xác suất xảy ra của biến cố B với điều kiện biến cố A xảy ra là:

$$P(B|A) = \frac{P(AB)}{P(A)} \tag{8.1}$$

Lúc này, A và B không độc lập.

Tổng quát, nếu n biến cố A_i, $i = 1, \ldots, n$ không độc lập thì:

$$\begin{aligned}P(A_1 A_2 \cdots A_n) =& P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 A_1) \cdots \\ & P(A_n|A_1 A_2 \cdots A_{n-1})\end{aligned}$$

Ví dụ 3. *Xét hai câu hỏi trắc nghiệm có 4 lựa chọn. Tính xác suất học sinh trả lời đúng câu thứ hai với điều kiện câu đầu trả lời sai.*

Giải. *Gọi A là biến cố câu đầu tiên học sinh trả lời sai. $P(A) = \frac{3}{4}$*

Gọi B là biến cố câu thứ hai học sinh trả lời đúng. $P(B) = \frac{1}{4}$.

Do A và B là hai biến cố độc lập nên $P(AB) = P(A)P(B) = \frac{3}{16}$

Như vậy, xác suất học sinh trả lời đúng câu thứ hai với điều kiện câu đầu trả lời đúng là: $P(B|A) = \frac{P(AB)}{P(A)} = \frac{3/16}{3/4} = \frac{1}{4}$

8.3 Công thức xác suất đầy đủ

Định nghĩa 4. Hệ biến cố đầy đủ

Xét phép thử có không gian mẫu là Ω. Một hệ các biến cố A_1, A_2, ..., A_n được gọi là **đầy đủ** nếu chúng thỏa các điều kiện:

- $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$
- $A_i \cap A_j = \emptyset$ với mọi $i \neq j$

Định lí 3. Công thức xác suất đầy đủ

Gọi $A_1, A_2, \ldots, A_n$ là một hệ biến cố đầy đủ. Khi đó, với biến cố B bất kì trong phép thử:

$$P(B) = P(A_1) \cdot P(B|A_1) + \cdots P(A_n) \cdot P(B|A_n) \tag{8.2}$$

Định lí 4. Công thức Bayes

Xét hệ có n biến cố đầy đủ $\{A_1, A_2, \ldots, A_n\}$.

Với biến cố B bất kì thì:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

với $1 \leq i \leq n$.

Chương 9

Biến ngẫu nhiên

9.1 Biến ngẫu nhiên

Xét phép thử với không gian mẫu Ω. Với mỗi biến cố sơ cấp $\omega \in \Omega$ ta liên kết với một số thực $\xi(\omega) \in \mathbb{R}$ thì ξ được gọi là **biến ngẫu nhiên** (BNN).

Định nghĩa 1. Biến ngẫu nhiên

Biến ngẫu nhiên ξ của một phép thử với không gian mẫu Ω là ánh xạ:

$$\xi = \xi(\omega), \quad \omega \in \Omega$$

Giá trị $\xi(\omega)$ được gọi là một giá trị của biến ngẫu nhiên ξ.

- Nếu $\xi(\Omega)$ là một tập hữu hạn $\{\xi_1, \xi_2, \ldots, \xi_n\}$ hay tập vô hạn đếm được thì ξ được gọi là **biến ngẫu nhiên rời rạc**.
- Nếu $\xi(\Omega)$ là một khoảng của $\mathbb{R}$ hay toàn bộ $\mathbb{R}$ thì ξ được gọi là **biến ngẫu nhiên liên tục**.

Định nghĩa 2. Hàm phân phối

Hàm phân phối của biến ngẫu nhiên ξ là hàm số $F(x)$, xác định bởi:

$$F(x) = P(\xi \leq x), \quad x \in \mathbb{R} \tag{9.1}$$

Ở đây ta viết gọn $P(\xi \leq x)$ từ $P(\{\omega : \xi(\omega) \leq x\})$. Tập hợp $\{\omega : \xi(\omega) \leq x\}$ có thể không thuộc một biến cố nào, do đó có thể là tập rỗng (ứng với xác suất là 0).

9.2 Tính chất của hàm phân phối

Tính chất 1. Hàm phân phối $F(x)$ không giảm trên mọi đoạn thẳng.

Chứng minh. Đặt $x_2 > x_1$. Ta thấy rằng

$$\{\xi \le x_2\} = \{\xi \le x_1\} + \{x_1 < \xi \le x_2\},$$

Do đó nếu ta lấy xác suất thì cũng có

$$P(\xi \le x_2) = P(\xi \le x_1) + P(x_1 < \xi \le x_2)$$

Xác suất luôn không âm, hay $P(x_1 < \xi \le x_2) \ge 0$, suy ra $P(\xi \le x_2) \ge P(\xi \le x_1)$, hay $F(x_2) \ge F(x_1)$. □

Tính chất 2. $\lim\limits_{x\to-\infty} F(x) = 0$.

Tính chất 3. $\lim\limits_{x\to+\infty} F(x) = 1$.

Tính chất 4. Hàm phân phối $F(x)$ liên tục phải trên toàn trục số.

Để chứng minh các tính chất 2, 3, 4 chúng ta cần các tiên đề của sự liên tục (continunity axioms) và sẽ không đề cập ở đây.

9.3 Biến ngẫu nhiên rời rạc

Cho BNN rời rạc $\xi = \xi(\omega)$, $\xi = \{a_1, a_2, \ldots, a_n, \ldots\}$. Giả sử $a_1 < a_2 < \ldots < a_n < \ldots$ với xác suất tương ứng là $P(\xi = a_i) = p_i$, $i = 1, 2, \ldots$

Ta có thể biểu diễn biến ngẫu nhiên và xác suất tương ứng của nó bằng bảng phân phối xác suất của ξ.

ξ	a_1	a_2	$\cdots$	a_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

Rõ ràng rằng $p_n \ge 0$ với mọi n. Hơn nữa

$$\sum_{n=1}^{\infty} p_n = 1$$

Không gian mẫu lúc này là hợp của các tập biến ngẫu nhiên rời rạc:

$$\Omega = \{\xi = a_1\} \cup \{\xi = a_2\} \cup \ldots$$

Các biến ngẫu nhiên xung khắc nhau (vì ξ không thể nhận hai giá trị khác nhau cùng lúc), do đó xác suất cả không gian mẫu là

$$1 = P(\Omega) = P(\xi = a_1) + P(\xi = a_2) + \ldots = p_1 + p_2 + \ldots$$

Định nghĩa 3. Phân phối nhị thức

Biến ngẫu nhiên ξ được gọi là có **phân phối nhị thức** với tham số p, n, với $p \in (0,1)$ và n là số tự nhiên, nếu ξ nhận các giá trị $0, 1, \ldots, n$ và

$$P(\xi = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n \tag{9.2}$$

Ở đây $q = 1 - p$.

Ví dụ 4. *Một bài kiểm tra có 100 câu hỏi trắc nghiệm bốn đáp án. Xác suất chọn ngẫu nhiên đúng đáp án của mỗi câu hỏi thì bằng nhau và bằng* $\frac{1}{4}$.

Ở đây xác suất chọn ngẫu nhiên đúng đáp án của một câu hỏi bất kì là $p = \frac{1}{4}$, *và số lượng câu hỏi là* $n = 100$.

Gọi ξ *là biến ngẫu nhiên số câu hỏi trả lời đúng. Khi đó* ξ *nhận các giá trị* $0, 1, \ldots, 100$.

Do đó bài toán này có phân phối nhị nhức và

$$P(\xi = k) = C_{100}^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{100-k}$$

Định nghĩa 4. Phân phối Poisson

Biến ngẫu nhiên ξ được gọi là có **phân phối Poisson** với tham số λ, nếu ξ nhận các giá trị $0, 1, \ldots, n$ và

$$P(\xi = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots, n \tag{9.3}$$

Tham số λ thể hiện số lần trung bình mà một sự kiện xảy ra trong một khoảng thời gian nhất định. Khi đó, nếu một biến ngẫu nhiên có số lần xuất hiện trung bình của một sự kiện trong thời gian t thì nó có phân phối Poisson với tham số λt, với λ là số lần trung bình trong một đơn vị thời gian.

9.4 Biến ngẫu nhiên liên tục

Định nghĩa 5. Biến ngẫu nhiên liên tục

Biến ngẫu nhiên ξ được gọi là **liên tục**, nếu nó nhận giá trị tại mọi điểm thuộc một đoạn liên tục nào đó trên trục số, và tồn tại một hàm số không âm $p(x)$ sao cho với mọi đoạn $[a, b]$ (hữu hạn hoặc vô hạn) ta có

$$P(a \leq \xi \leq b) = \int_a^b p(x)\,dx \tag{9.4}$$

Hàm $p(x)$ được gọi là **hàm mật độ** của biến ngẫu nhiên ξ.

Tương tự biến ngẫu nhiên rời rạc, $p(x) \geq 0$ với mọi $x \in \mathbb{R}$ và khi hai cận là vô cực thì biến ngẫu nhiên bao quát toàn bộ không gian mẫu. Nghĩa là

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

Từ định nghĩa của hàm phân phối $F(x) = P(\xi \leq x)$ ta có hai tính chất của hàm mật độ:

1. $F(x) = \int_{-\infty}^{x} p(x)\,dx$
2. $p(x) = F'(x)$

Tính chất thứ nhất là từ định nghĩa hàm phân phối. Tính chất thứ hai suy ra từ việc cận trên của tích phân là hữu hạn.

Hàm mật độ của X là

$$f(x) = \begin{cases} p_i & \text{khi } x = x_i, \\ 0 & \text{khi } x \neq x_i, \forall i \end{cases}$$

Nhận xét 1

Ta có các lưu ý sau:

- $p_i \geq 0$, $\sum p_i = 1$, $i = 1, 2, \ldots$
- $P(a < X \leq b) = \sum_{a < x_i \leq b} p_i$

9.5 Hàm mật độ của biến ngẫu nhiên liên tục

Định nghĩa 6

Hàm số $f : \mathbb{R} \mapsto \mathbb{R}$ được gọi là **hàm mật độ** của biến ngẫu nhiên liên tục X nếu:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \forall a, b \in \mathbb{R}$$

Nhận xét 2

Với mọi $x \in \mathbb{R}$, $f(x) \geq 0$ và $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

Ý nghĩa hình học. Xác suất của biến ngẫu nhiên X nhận giá trị trong $[a, b]$ bằng diện tích hình thang cong giới hạn bởi $x = a$, $x = b$, $y = f(x)$ và Ox.

Chương 10

Ôn thi

10.1 Ôn thi ngày 20/11/2023

10.1.1 Toán tử tuyến tính

Toán tử tuyến tính là một ánh xạ

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

Nếu A là một ma trận cỡ $m \times n$ thì đây là một ánh xạ tuyến tính với phép nhân ma trận với vector $A \cdot \boldsymbol{x} = \boldsymbol{y}$.

Ở đây $\boldsymbol{x} \in \mathbb{R}^n$ và $\boldsymbol{y} \in \mathbb{R}^m$.

Định nghĩa 1. Hạt nhân

Hạt nhân của ánh xạ tuyến tính A là tập hợp nghiệm của hệ thuần nhất và được ký hiệu là $\ker(A)$. Nói cách khác

$$\ker(A) = \{\boldsymbol{x} \in \mathbb{R}^n : \ A \cdot \boldsymbol{x} = \boldsymbol{0}\} \tag{10.1}$$

Định nghĩa 2. Ảnh

Ảnh của ánh xạ tuyến tính A là tập hợp tất cả giá trị có thể của phép nhân ma trận và được ký hiệu là $\mathrm{im}(A)$. Nói cách khác

$$\mathrm{im}(A) = \{A \cdot \boldsymbol{x} : \ \boldsymbol{x} \in \mathbb{R}^n\} \tag{10.2}$$

Tính chất đối với ánh xạ $A : \mathbb{R}^n \to \mathbb{R}^m$ là $\dim(\ker A) + \dim(\mathrm{im}\, A) = n$.

10.1.2 Trị riêng và vector riêng

Định nghĩa 3. Trị riêng, vector riêng

Xét hệ phương trình tuyến tính thuần nhất biểu diễn bởi phép nhân

$$A \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x}$$

Giá trị λ khiến phương trình có nghiệm không tầm thường được gọi là **trị riêng** (eigenvalue) của ánh xạ tuyến tính.

Vector $\boldsymbol{x}$ là cơ sở của không gian vector nghiệm khi đó được gọi là **vector riêng** (eigenvector) ứng với trị riêng λ.

Lưu ý rằng có thể có nhiều vector riêng tương ứng với một trị riêng.

Để tìm trị riêng ta giải phương trình đặc trưng $\det(A - \lambda I) = 0$ và tìm tất cả nghiệm thực λ của phương trình.

Sau đó ta thế từng λ vào hệ $A\boldsymbol{x} = \lambda\boldsymbol{x}$ và tìm cơ sở của không gian nghiệm. Các vector trong cơ sở là vector riêng tương ứng với λ đó.

Một số tính chất của trị riêng và vector riêng (giả sử rằng đối với ma trận A cỡ $n \times n$ thì phương trình đặc trưng có đầy đủ n nghiệm thực).

1. $\operatorname{tr} A = \lambda_1 + \lambda_2 + \ldots + \lambda_n$
2. $\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$

Tính chất liên quan đến rank và trace:

1. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$
2. $\operatorname{rank}(AB) \leqslant \min(\operatorname{rank}(A), \operatorname{rank}(B))$

Bài 1. Cho vector cột $\boldsymbol{v} \in \mathbb{R}^n$. Đặt $A = \boldsymbol{v} \cdot \boldsymbol{v}^T$. Tìm spa A.

Các cột của A có dạng $\boldsymbol{v} \cdot v_1$, $\boldsymbol{v} \cdot v_2$, ..., $\boldsymbol{v} \cdot v_n$. Như vậy các cột đều tỉ lệ với cột đầu nên $\operatorname{rank} A = 1$.

Suy ra $\dim \ker A = n - 1$ và do đó $\lambda = 0$ là nghiệm bậc $n - 1$ trong phương trình đặc trưng.

Như vậy phương trình đặc trưng còn một nghiệm $\lambda \neq 0$.

Do $(\boldsymbol{v} \cdot \boldsymbol{v}^T)\boldsymbol{x} = \lambda\boldsymbol{x} \Leftrightarrow \boldsymbol{v}(\boldsymbol{v}^T \cdot \boldsymbol{x}) = \lambda\boldsymbol{x}$.

Đặt $\boldsymbol{v}^T \cdot \boldsymbol{x} = \alpha$ thì $\alpha\boldsymbol{v} = \lambda\boldsymbol{x}$. Suy ra $\boldsymbol{x} = \boldsymbol{v}$ và do đó $\alpha = \lambda = \|\boldsymbol{v}\|^2$.

Vậy spa $A = \{\|\boldsymbol{v}\|^2, 0, 0, \ldots, 0\}$.

Bài 3. Cho ma trận $A_{3\times3}$. Biết rằng $\operatorname{tr} A = \operatorname{tr} A^{-1} = 0$ và $\det A = 1$. Chứng minh rằng $A^3 = I$.

Phương trình đặc trưng có dạng $P_3(\lambda) = -\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$.

Theo tính chất trên thì $a_2 = \sum \lambda = \operatorname{tr} A = 0$.

Do λ là trị riêng nên $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Do A khả nghịch nên $\dfrac{1}{\lambda}\boldsymbol{x} = A^{-1}\boldsymbol{x}$.

Nghĩa là $\dfrac{1}{\lambda}$ là trị riêng của ma trận A^{-1}. Suy ra $\dfrac{1}{\lambda_1} + \dfrac{1}{\lambda_2} + \dfrac{1}{\lambda_3} = \operatorname{tr} A^{-1} = 0$.

Từ đó suy ra $\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 = 0$.

Cuối cùng $\det A = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 = 1$.

Vậy phương trình đặc trưng là $P_3(\lambda) = -\lambda^3 + 1$. Theo định lý Cayley-Hamilton thì $P_3(A) = -A^3 + I = 0$, hay $A^3 = I$.

Bài 4. Cho ma trận $A_{n\times n}$, $A_{ij} \geqslant 0$. Giả sử ma trận có đủ n trị riêng thực. Chứng minh rằng $\lambda_1^k + \lambda_2^k + \ldots + \lambda_n^k \geqslant 0$ với mọi $k \in \mathbb{N}$.

Ta thấy rằng với $k = 1$ thì $\lambda_1 + \ldots + \lambda_n = \operatorname{tr}(A) \geqslant 0$.

Vì λ_i là thỏa phương trình $A\boldsymbol{x} = \lambda_i\boldsymbol{x}$ nên nhân hai vế cho A ta có $A \cdot A\boldsymbol{x} = A \cdot \lambda_i\boldsymbol{x}$. Tương đương với $A^2\boldsymbol{x} = \lambda_i(A\boldsymbol{x}) = \lambda_i^2\boldsymbol{x}$.

Nói cách khác, λ_i^2 là trị riêng của ma trận A^2. Thực hiện tương tự ta có λ_i^k là trị riêng của ma trận A^k.

Do đó $\lambda_1^k + \ldots + \lambda_n^k = \operatorname{tr}(A^k) \geqslant 0$.

Bài 5. Cho ma trận A khả nghịch. X là ma trận sao cho $AX + XA = 0$. Chứng minh rằng $\operatorname{tr} X = 0$.

Nhân bên trái hai vế cho A^{-1} ta có $X + A^{-1}XA = 0$. Ta biết rằng $A^{-1}XA$ là ma trận tương đương ma trận X nên $\operatorname{tr}(A^{-1}XA) = \operatorname{tr} X$.

Suy ra $\operatorname{tr} X + \operatorname{tr} X = \operatorname{tr} 0 = 0$. Từ đây có $\operatorname{tr} X = 0$.

Chương 11

RUDN Olympiad 2023

Lần đầu tiên mình được tham dự thi toán đồng đội theo hình thức MathBoy (trận chiến toán).

Trong cách thi này, mỗi đội có 3 vị trí: người thuyết trình (докладчик), người phản biện (оппонент) và người giám sát (наблюдатель).

Ở mỗi vòng sẽ có 3 đội thi với nhau. Mỗi đội sẽ có 1 vị trí tương ứng với 3 vị trí trên. Sau đây là ví dụ

	Đội 1	Đội 2	Đội 3
Vòng 1	О	Д	Н
Vòng 2	Н	О	Д
Vòng 3	Д	Н	О

Ở mỗi vòng, đội đóng vai trò người thuyết trình lên bảng ghi bài giải trong thời gian cho phép và thuyết trình về bài giải của đội mình. Đội phản biện có nhiệm vụ phản biện bài thuyết trình đó. Đội giám sát, dựa trên bài thuyết trình cũng như phản biện mà ghi chép lại các lỗi, chỗ khó hiểu, ... và trình lên cho giám khảo.

Ngoài ra, đội thuyết trình trước đó phải trình bài giải viết tay cho giám khảo chấm trước khi lên thuyết trình.

Ở đây có rất nhiều câu chuyện hack não đã xảy ra. Lúc mình thi vòng 1, câu hỏi quá khó nên đội thuyết trình chỉ viết được một ít. Đồng nghĩa việc đội phản biện cũng như đội giám sát ... thất nghiệp, không có gì để nói.

Đối với vòng 2, trận chiến cân bằng hơn, đội mình làm việc giám sát. Dựa trên bài giải của đội thuyết trình, chúng mình thấy những trường hợp chưa được xét tới và có thể bị sai, do đó cả ba đội đều có điểm (đội thuyết trình có nhiều điểm nhất vì các bạn giải hơn 1 nửa rồi).

Đối với vòng 3, đội mình thuyết trình. Đội mình clear bài đó nên giành điểm tuyệt đối cho phần thuyết trình. Tuy nhiên các bạn phản biện cũng không vừa, vẫn cố gắng bắt một số lỗi do trình bày quá cô đọng. Kết quả là đội mình

(thuyết trình) full điểm cho vòng 3, đội phản biện được 3 điểm.

Chương 12

Hình học giải tích

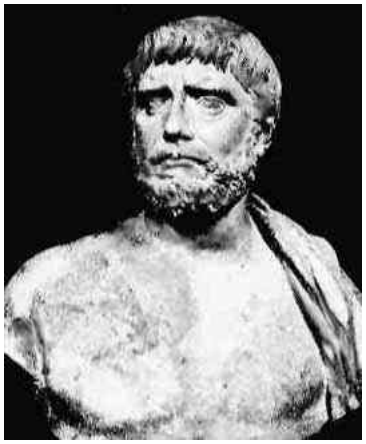
12.1 Theo dòng lịch sử

Hình học xuất hiện từ thời xa xưa, xuất phát từ những nhu cầu thực tế nhất của con người là đo đạc để phân chia đất đai, xây dựng, canh tác, ... Từ đó con người đã có nhận thức rất sớm về quan hệ song song và vuông góc giữa hai đường thẳng.

Một cách hình ảnh (mà thật ra hình học là môn học về hình ảnh) thì hai đường thẳng song song không cắt nhau dù có kéo dài chúng ra vô tận. Các đường thẳng song song luôn có nhiều điều thú vị, cả ở mặt phẳng Euclid lẫn trong không gian. Đầu tiên phải kể đến định lý mang tên triết gia vĩ đại của Hy Lạp: Thales.

12.1.1 Thales của Miletus

Thales của Miletus được cho rằng sinh vào khoảng năm 624 Trước Công nguyên (TCN) và mất năm 547 TCN tại Miletus (Thổ Nhĩ Kì ngày nay) ([6]).



Hình 12.1: Thales của Miletus

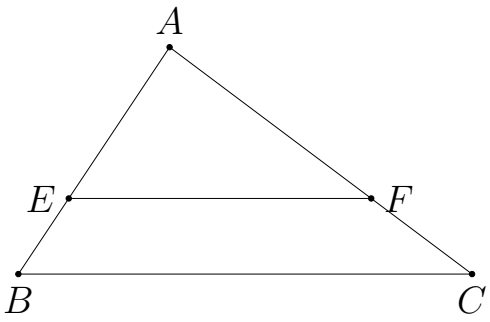
Ông được xem là nhà triết học đầu tiên khi không cố gắng giải thích tự nhiên

bằng thần thoại hay các thế lực siêu nhiên như trước. Trường phái triết học do ông sáng lập, trường phái Milet, cho rằng mọi vật có nguồn gốc từ nước. Nhà triết học nổi tiếng Aristotle đánh giá rằng Thales là người sáng lập ra *triết học duy vật sơ khai*.

Trong toán học, Thales được biết tới với định lý mang tên ông về các đường song song. Định lý Thales được phát biểu như sau:

Định lí 1. Định lý Thales

Trong một tam giác, đường thẳng song song với một cạnh chắn trên hai cạnh còn lại các đoạn thẳng tương ứng tỉ lệ.



Hình 12.2: Định lý Thales trên mặt phẳng

Theo định lý Thales, nếu EF song song với BC thì ta có $\frac{AE}{AB} = \frac{AF}{AC} = \frac{EF}{BC}$ (hình 12.2).

Không dừng lại ở mặt phẳng, khi mở rộng lên không gian định lý Thales cũng cho chúng ta một kết quả quan trọng khi nói tới các mặt phẳng song song nhau.

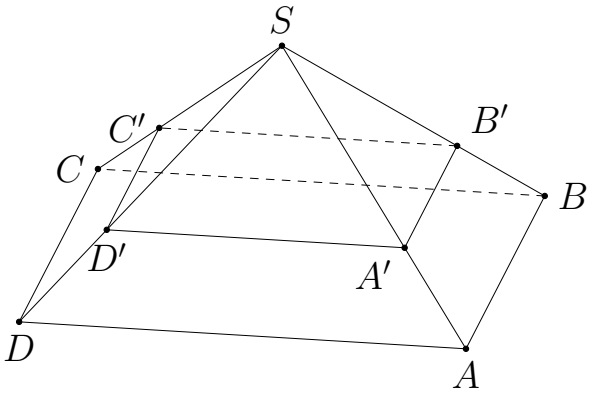
Định lí 2. Định lý Thales trong không gian

Trong khối chóp, mặt phẳng song song mặt đáy chắn các cạnh nối từ đỉnh hình chóp tới các đỉnh của mặt phẳng đáy các đoạn thẳng tương ứng tỉ lệ.

Theo định lý Thales, nếu mặt phẳng $(ABCD)$ song song với mặt phẳng $(A'B'C'D')$ thì $\frac{SA}{SA'} = \frac{SB}{SB'} = \frac{SC}{SC'} = \frac{SD}{SD'}$ (hình 12.3).

12.1.2 Pythagoras của Samos

Khi nhắc tới vuông góc, chúng ta thường nhớ tới định lý ngày nào được học ở thời học sinh: định lý Pythagoras. Định lý này nói về quan hệ giữa độ dài



Hình 12.3: Định lý Thales trong không gian

các cạnh trong một tam giác vuông. Định lý tuy đơn giản nhưng có ý nghĩa rất quan trọng trong đời sống và khoa học của con người suốt chiều dài lịch sử. Đây cũng là tiền đề cho định lý mang tính lịch sử của nhân loại: định lý cuối cùng của Fermat.

Hình 12.4: Pythagoras của Samos

Pythagoras của Samos cũng là nhà triết học Hy Lạp cổ, được cho rằng sinh vào khoảng năm 570 TCN và mất năm 490 TCN ([5]). Ông được học tập từ nhà triết học Thales và cũng có nhiều đóng góp cho sự phát triển của toán học, thiên văn học và âm nhạc. Tuy nhiên khác với thầy mình, trường phái triết học của ông cho rằng những con số là nguồn gốc của vạn vật và sử dụng những con số để giải thích những hiện tượng khoa học. Từ đây, các lý thuyết về âm nhạc được ra đời, cụ thể là các mối liên hệ về tần số với sự rung của dây nhạc cụ.

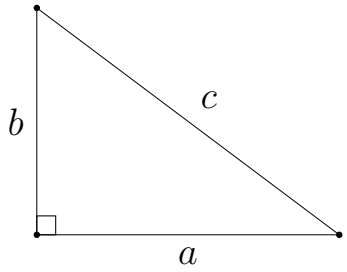
Ông là một trong những người hiếm hoi cho phép cả phụ nữ đi học ở lớp của mình vào thời ấy. Điều đó giúp phổ biến toán học nói riêng và kiến thức nói chung tới nhiều tầng lớp nhân dân. Tuy nhiên ông cũng có một hội kín rất thú vị. Như đã nói ở trên, trường phái triết học Pythagoras cố gắng giải thích nguồn gốc vạn vật bằng những con số. Điều này đã dẫn họ tới những khám phá động trời vào thời ấy.

Một trong những khám phá đó là về sự tồn tại của số vô tỉ dựa vào định lý mang tên ông. Lịch sử đã chỉ ra rằng trước Pythagoras, người Babylon và Ai Cập đã tìm ra rất nhiều bộ số nguyên (a, b, c) thỏa mãn $a^2 + b^2 = c^2$ là độ dài ba cạnh tam giác vuông. Định lý Pythagoras mà ngày nay chúng ta biết được phát biểu rằng:

Định lí 3. Định lý Pythagoras

Trong một tam giác vuông, bình phương độ dài cạnh huyền bằng tổng bình phương độ dài hai cạnh góc vuông.

Như vậy nếu gọi độ dài cạnh huyền là c, độ dài hai cạnh góc vuông lần lượt là a và b thì $a^2 + b^2 = c^2$ (hình 12.5).



Hình 12.5: Định lý Pythagoras

Nếu $a = b = 1$ thì sao? Khi đó bình phương độ dài cạnh huyền $c^2 = 2$. Tuy nhiên không thể tìm ra một số hữu tỉ nào để bình phương lên là 2 cả. Phát hiện này là một chấn động đối với thời Pythagoras và ông yêu cầu tất cả thành viên trong hội phải giữ kín bí mật về sự phát hiện này. Tuy nhiên thông tin vẫn lọt ra ngoài và truyền thuyết kể rằng ông đã xử tội chết cho thành viên của hội không tuân thủ.

Pythagoras đã đưa một khái niệm cực kì quan trọng trong toán học, gọi là *chứng minh* (proof). Để chứng minh một mệnh đề là đúng, chúng ta cần các mệnh đề (thường đơn giản hơn) đúng trước đó. Bằng các phép suy luận thích hợp dựa trên các mệnh đề đúng trước đó, chúng ta có thể kết luận rằng mệnh đề cần chứng minh là đúng. Phép chứng minh có thể gọi là *xương sống* của toán học, vì nếu không có một phép chứng minh đúng đắn thì một mệnh đề không thể được xác định được là có đúng hay không. Trong trường hợp của Fermat, khi ông đưa ra định lý Fermat nhưng không kèm chứng minh (vì lề sách quá chật nên không viết lời giải được) thì chúng ta không thể biết định lý Fermat có đúng hay không (?).

Nếu việc suy luận dựa trên các mệnh đề, hoặc định lý, đã đúng trước đó, thì phải có một lúc nào đó việc này dừng lại. Chúng ta không thể suy ngược tới vô hạn lần được. Do đó chúng ta cần những mệnh đề luôn đúng nhưng tính đúng đắn của nó được kiểm nghiệm trong thực tiễn. Chúng được gọi là *tiên đề*

(axiom). Nhân vật tiếp theo được đề cập tới sẽ dẫn chúng ta tới hệ thống tiên đề làm nền tảng cho hình học.

12.1.3 Euclid của Alexandria

Đúng vậy, Euclid là người đặt nền móng cho hình học với bộ sách nổi tiếng *Elements* của mình. Trong bộ sách này đề cập tới những tiên đề, định lý làm nền tảng cho bộ môn hình học và vẫn còn ý nghĩa cho tới tận ngày nay. Những gì viết trong đó không quá xa lạ với những gì được giảng dạy trong nhà trường.

Hình 12.6: Euclid của Alexandria

Euclid của Alexandria sinh vào khoảng năm 325 TCN và mất vào khoảng năm 265 TCN ([2]). Thông tin về ông không có nhiều. Nhưng chỉ mỗi bộ sách *Elements* cũng đủ để người đời sau cho rằng ông là người có ảnh hưởng nhất trong 2000 năm lịch sử phát triển của toán học.

Năm tiên đề cơ bản của hình học được ông phát biểu trong bộ *Elements* được phát biểu như sau:

1. Qua hai điểm bất kì luôn vẽ được một đường thẳng
2. Đường thẳng có thể kéo dài vô hạn về cả hai phía
3. Ta có thể xác định một đường tròn bằng tâm và bán kính của nó
4. Mọi góc vuông đều bằng nhau
5. Nếu một đường thẳng cắt hai đường thẳng khiến tổng hai góc trong cùng phía nhỏ hơn hai vuông thì hai đường thẳng đó chắc chắn sẽ cắt nhau tại một điểm nào đó

Tiên đề số 5 là rắc rối và phức tạp nhất. Nó không thực sự tự nhiên và có nhiều sự vướng mắc. Đây chính là tiên đề cho sự ra đời của hình học phi-Euclid hơn 1500 năm sau.

Bộ *Elements* của Euclid bao gồm 13 quyển. Trong đó đề cập tới rất nhiều vấn đề của hình học, từ những phần tử đơn giản nhất cấu tạo nên hình học là điểm, đoạn thẳng, đường thẳng, tới những hình học lớn hơn như hình chữ nhật, hình tròn, đa giác, mặt phẳng. Thậm chí ông cũng đã có những dấu chân ở hình

học không gian như hình chóp, hình cầu, hình nón ([1], [4]).

12.2 Danh mục thuật ngữ và ký hiệu

Đầu tiên chúng ta thống nhất các thuật ngữ cũng như ký hiệu được sử dụng kể từ đây của bài viết.

Điểm là đơn vị cơ bản của hình học. Bất kì đối tượng hình học nào cũng là một *tập hợp điểm*. Điểm được ký hiệu bởi chữ in hoa, ví dụ như A, B_1, B_2.

Đường thẳng đi qua hai điểm phân biệt cho trước. Đường thẳng có thể kéo dài vô hạn về hai phía. Đường thẳng được ký hiệu bởi chữ in thường hoặc chữ Hy Lạp trong ngoặc đơn, ví dụ như (d), (Δ).

Đoạn thẳng chỉ phần đường thẳng nằm giữa hai điểm.

Nửa đường thẳng chỉ phần đường thẳng nằm một phía của một điểm trên đường thẳng và chỉ kéo dài vô hạn về phía đó.

Vector là đoạn thẳng có hướng. Với điểm đầu là A và điểm cuối là B thì vector từ A tới B được ký hiệu là $\overrightarrow{AB}$. Để chỉ một vector không cần biết điểm đầu và điểm cuối ta dùng chữ thường in đậm, ví dụ như $\boldsymbol{a}$.

Góc giữa hai vector $\overrightarrow{OA}$ và $\overrightarrow{OB}$ là góc $\angle AOB$ và ký hiệu là $(\overrightarrow{OA}, \overrightarrow{OB})$.

Tương tự đối với vector $\boldsymbol{a}$ và $\boldsymbol{b}$ thì góc giữa chúng ký hiệu là $(\boldsymbol{a}, \boldsymbol{b})$.

12.3 Phương pháp tọa độ trong mặt phẳng

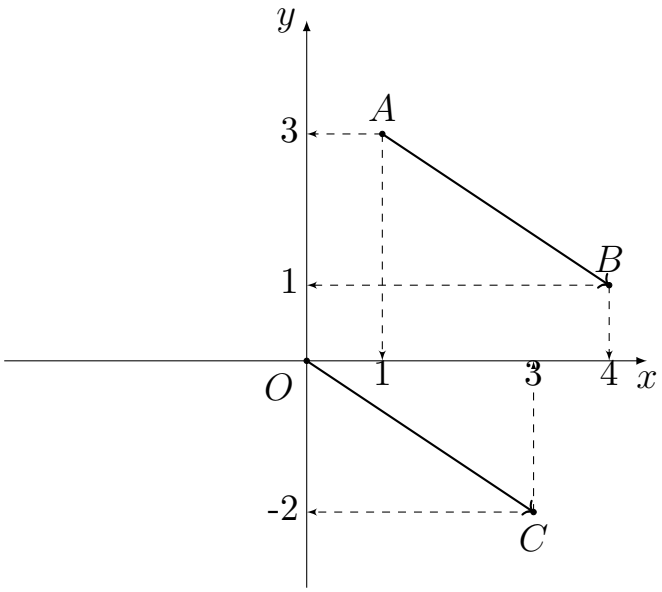
Cuộc cách mạng trong hình học xảy ra khi nhà toán học lãng tử René Descartes phát minh ra hệ tọa độ và từ đó mọi đối tượng hình học có thể được biểu diễn bởi các phương pháp đại số như phương trình, đẳng thức.

12.3.1 Vector trong mặt phẳng

Trong hệ tọa độ Oxy với tâm O và hai trục Ox (trục hoành) và Oy (trục tung) vuông góc nhau, đặt $O = (0, 0)$ là tọa độ của tâm O.

Tiếp theo, mọi điểm trong mặt phẳng Euclid đi liền với cặp số (x, y) chỉ tọa độ của điểm đó. Ví dụ $A = (1, 3)$, $B = (4, 1)$.

Tọa độ của điểm cũng là tọa độ của vector từ O tới điểm đó. Với hình 12.7 thì $\overrightarrow{OA} = (1, 3)$ và $\overrightarrow{OB} = (4, 1)$. Tọa độ của vector $\overrightarrow{AB}$ khi đó sẽ là $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = (4, 1) - (1, 3) = (3, -2)$. Cũng theo hình 12.7 thì ta thấy $\overrightarrow{AB} = \overrightarrow{OC} = (3, -2)$.



Hình 12.7: Tọa độ của điểm trong mặt phẳng

Như vậy, nếu ta có hai điểm $A = (x_A, y_A)$ và $B = (x_B, y_B)$ thì vector $\overrightarrow{AB}$ là

$$\overrightarrow{AB} = (x_B - x_A, y_B - y_A) \tag{12.1}$$

Tích vô hướng của hai vector $\boldsymbol{a} = (x_1, y_1)$ và $\boldsymbol{b} = (x_2, y_2)$ được định nghĩa là

$$\boldsymbol{a} \cdot \boldsymbol{b} = x_1 x_2 + y_1 y_2 \tag{12.2}$$

Ta ký hiệu $|\boldsymbol{a}|$ là độ dài (chuẩn Euclid) của vector $\boldsymbol{a}$. Trong hệ tọa độ Descartes vuông góc, theo định lý Pythagoras, độ dài của vector là độ dài cạnh huyền tam giác vuông (hình 12.7). Như vậy, độ dài đoạn thẳng AB với $A = (x_A, y_A)$ và $B = (x_B, y_B)$ là

$$AB = |\overrightarrow{AB}| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \tag{12.3}$$

Khi đó cosin góc giữa hai vector $\boldsymbol{a}$ và $\boldsymbol{b}$ là

$$\cos((\boldsymbol{a}, \boldsymbol{b})) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|} = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \cdot \sqrt{x_2^2 + y_2^2}} \tag{12.4}$$

Nếu góc giữa hai vector bằng 90 độ thì hai vector được gọi là vuông góc nhau. Khi đó tích vô hướng $\boldsymbol{a} \cdot \boldsymbol{b} = 0$.

12.3.2 Phương trình đường thẳng trong mặt phẳng

Theo tiên đề Euclid, một đường thẳng được xác định khi biết hai điểm phân biệt thuộc đường thẳng đó. Trong hệ tọa độ, chúng ta có hai cách tìm phương trình đường thẳng.

Bằng vector pháp tuyến. Vector pháp tuyến của đường thẳng là vector vuông góc với mọi vector có phương là đường thẳng đó. Giả sử $\boldsymbol{v} = (a, b)$ là vector pháp tuyến của đường thẳng đi qua điểm $M_0 = (x_0, y_0)$. Khi đó đường thẳng đi qua qua M_0 nhận $\boldsymbol{v}$ làm vector pháp tuyến là *tập hợp điểm* $M = (x, y)$ trên mặt phẳng sao cho $\boldsymbol{v} \cdot \overrightarrow{M_0M} = 0$. Điều này tương đương với

$$\boldsymbol{v} \cdot \overrightarrow{M_0M} = a \cdot (x - x_0) + b \cdot (y - y_0) = 0 \tag{12.5}$$

Bằng vector chỉ phương. Vector chỉ phương của đường thẳng là vector có phương song song với đường thẳng đó. Giả sử $\boldsymbol{v}' = (a', b')$ là vector chỉ phương của đường thẳng đi qua điểm $M_0 = (x_0, y_0)$. Khi đó đường thẳng đi qua M_0 nhận $\boldsymbol{v}'$ làm vector chỉ phương là *tập hợp điểm* $M = (x, y)$ trên mặt phẳng sao cho $\boldsymbol{v}' \parallel \overrightarrow{M_0M}$. Điều này tương đương với

$$\boldsymbol{v}' \parallel \overrightarrow{M_0M} \Leftrightarrow \frac{x - x_0}{a'} = \frac{y - y_0}{b'} \tag{12.6}$$

1. Cả hai cách biểu diễn khi khai triển ra đều có dạng $ax + by + c = 0$ với c là hằng số. Đây được gọi là dạng tổng quát của phương trình đường thẳng.

2. Cách viết $\dfrac{x - x_0}{a'} = \dfrac{y - y_0}{b'}$ được gọi là dạng chính tắc của phương trình đường thẳng.

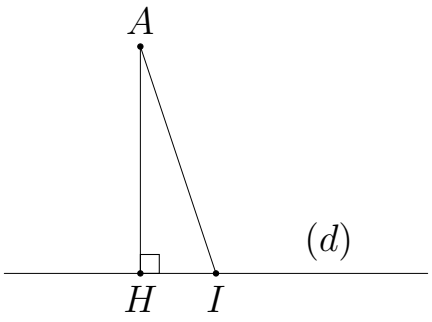
3. Dạng chính tắc của phương trình đường thẳng còn có một tác dụng đặc biệt khác
$$\frac{x - x_0}{a'} = \frac{y - y_0}{b'} = t$$
với $t \in \mathbb{R}$. Khi đó tọa độ $M = (x, y)$ có thể được biểu diễn dưới dạng
$$\begin{cases} x = x_0 + a't \\ y = y_0 + b't \end{cases}, \quad t \in \mathbb{R} \tag{12.7}$$
Đây được gọi là phương trình dạng tham số.

Chúng ta chú ý rằng nếu đường thẳng song song với một trong hai trục tọa độ thì vector chỉ phương của nó sẽ cùng phương với vector đơn vị $(1, 0)$ hoặc $(0, 1)$. Do đó không thể viết dưới dạng chính tắc được (không thể chia cho 0) nhưng có thể viết dưới dạng tổng quát hoặc dạng tham số.

12.3.3 Khoảng cách giữa điểm và đường thẳng

Nhắc lại một chút kiến thức cơ sở. **Khoảng cách** từ một điểm A nằm ngoài đường thẳng (d) là độ dài đoạn thẳng AH với $H \in (d)$ sao cho AH nhỏ nhất (hình 12.8).

Khi đó H được gọi là **hình chiếu** của A lên đường thẳng (d) và AH là **khoảng cách** từ A tới (d). Do AH là đoạn thẳng có độ dài ngắn nhất, điều này xảy ra khi $AH \perp (d)$.



Hình 12.8: Hình chiếu và khoảng cách tới đường thẳng

Như vậy, để tìm hình chiếu của điểm A lên đường thẳng (d), ta dựng đường thẳng đi qua điểm A và vuông góc với (d).

Giả sử phương trình đường thẳng (d) với vector pháp tuyến $\boldsymbol{v} = (a, b)$ là $(d) : ax + by + c = 0$.

Gọi (d') là đường thẳng đi qua $A = (x_0, y_0)$ và vuông góc với d. Do $\boldsymbol{v}$ là vector pháp tuyến của (d) nên $\boldsymbol{v}$ là vector chỉ phương của (d'). Khi đó phương trình dạng tham số của (d') là

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \end{cases}, t \in \mathbb{R}$$

Gọi H là hình chiếu của A lên (d). Khi đó H là giao điểm của (d) và (d'). Vì $H \in (d')$ nên tọa độ của H có dạng $(x_0 + at, y_0 + bt)$ với t nào đó thuộc $\mathbb{R}$. Chúng ta sẽ đi tìm t này.

Vì $H \in (d)$ nên ta thay tọa độ của H vừa tìm được vào phương trình của (d) thu được

$$a(x_0 + at) + b(y_0 + bt) + c = 0 \Leftrightarrow t = -\frac{ax_0 + by_0 + c}{a^2 + b^2}$$

Như vậy là ta đã tìm được t từ đó xác định được tọa độ của H.

Từ đây ta tính được khoảng cách từ A tới (d) hay nói cách khác là độ dài đường AH. Ta có $A = (x_0, y_0)$ và $H = (x_0 + at, y_0 + bt)$ nên $\overrightarrow{AH} = (at, bt)$. Suy ra

$$\begin{aligned} AH =& |\overrightarrow{AH}| = \sqrt{(at)^2 + (bt)^2} = |t|\sqrt{a^2 + b^2} \\ =& \left| -\frac{ax_0 + by_0 + c}{a^2 + b^2} \right| \cdot \sqrt{a^2 + b^2} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \end{aligned}$$

12.4 Đạo hàm

Phép tính vi tích phân đã được con người nghiên cứu từ lâu. Câu chuyện về ai là người phát minh ra phép tính vi tích phân: Newton hay Leibniz, được coi là một trong những vụ tranh cãi đáng xấu hổ nhất lịch sử toán học. Nhưng họ cũng đã để lại một mảnh đất màu mỡ cho toán học về sau.

12.4.1 Cơ học và sự ra đời của đạo hàm

Trường phái Newton sử dụng đạo hàm như công cụ khảo sát vận tốc từ quãng đường. Ở bậc trung học chúng ta biết rằng *vận tốc trung bình* bằng quãng đường chia thời gian. Tuy nhiên điều đó chỉ đúng cho *chuyển động thẳng đều*. Nếu quãng đường là một hàm số phụ thuộc thời gian (quãng đường là $s(t)$ với t là thời gian) thì điều đó không đúng nữa.

Do quãng đường phụ thuộc thời gian nên có thể là vận tốc cũng phụ thuộc thời gian? Hợp lí đấy. Nhưng với mỗi một giá trị thời gian t cho ta một vị trí $s(t)$ trên trục số, còn vận tốc thì không thể phụ thuộc một giá trị thời gian được. Rõ ràng vật phải di chuyển một quãng đường từ thời gian t_0 tới t_1 thì mới có vận tốc trên quãng đường đó chứ?

Cách tiếp cận ở đây là, chúng ta cho sự thay đổi thời gian, tức hiệu $\Delta t = t_1 - t_0$, rất nhỏ. Khi đó vật đi từ $s(t_0)$ tới $s(t_1)$, vậy là chúng ta có thể tính vận tốc với công thức $v = \dfrac{s(t_1) - s(t_0)}{t_1 - t_0}$. Do Δt rất nhỏ, hay *tiến về 0*, thì vận tốc gần như xảy ra vào đúng một thời điểm. Do đó vận tốc lúc này được gọi là *vận tốc tức thời*. Đó cũng chính là ý nghĩa cơ học và sự ra đời của đạo hàm theo trường phái Newton.

12.4.2 Định nghĩa đạo hàm

Xét hàm số $f(x)$ liên tục trên khoảng (a, b) có chứa điểm x_0. Đạo hàm của $f(x)$ tại x_0 được định nghĩa là giới hạn

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{12.8}$$

Lưu ý rằng nếu giới hạn trên không phải là giới hạn hữu hạn (không tồn tại hoặc tiến tới vô cực) thì hàm số không có đạo hàm tại điểm x_0.

Ví dụ, để tính đạo hàm của hàm số $f(x) = x^3 + 2x^2 - 4$ tại $x_0 = 4$, ta khai

triển

$$\begin{aligned}\frac{f(x)-f(x_0)}{x-x_0} =& \frac{f(x)-f(4)}{x-4}\\ =& \frac{x^3+2x^2-4-(4^3+2\cdot 4^2-4)}{x-4}\\ =& \frac{(x^3-4^3)+2(x^2-4^2)}{x-4}\\ =& \frac{(x-4)(x^2+4x+16)+2(x-4)(x+4)}{x-4}\\ =& x^2+4x+16+2(x+4)\end{aligned}$$

Cho x tiến tới 4 thì ta có đạo hàm tại $x=4$

$$\begin{aligned}f'(4) =& \lim_{x\to 4}\frac{f(x)-f(4)}{x-4}\\ =& \lim_{x\to 4}(x^2+4x+16+2(x+4))\\ =& 4^2+4\cdot 4+16+2\cdot(4+4)=64\end{aligned}$$

Trong định nghĩa ở 12.8, nếu ta đặt $\Delta x = x - x_0$ và $\Delta y = y - y_0 = f(x) - f(x_0)$, ta gọi Δx là *số gia* của biến x, tương tự Δy là *số gia* của biến y.

Trong định nghĩa, x tiến tới x_0 tương đương với Δx tiến tới 0. Chuyển vế x_0 ta có $x = x_0 + \Delta x$ và từ đó $f(x) = f(x_0 + \Delta x)$. Định nghĩa đạo hàm ở trên có thể được viết lại

$$f'(x_0) = \lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \tag{12.9}$$

Nếu hàm số có đạo hàm tại mọi điểm trên khoảng (a, b) thì ta nói hàm số khả vi trên khoảng đó.

Ví dụ đối với hàm số $f(x) = x^3 + 2x^2 - 4$ như trên. Với mọi $x_0 \in \mathbb{R}$ ta có

$$\begin{aligned}f'(x_0) =& \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}\\ =& \lim_{x\to x_0}\frac{x^3+2x^2-4-(x_0^3+2x_0^2-4)}{x-x_0}\\ =& \lim_{x\to x_0}\frac{(x^3-x_0^3)+2(x^2-x_0^2)}{x-x_0}\\ =& \lim_{x\to x_0}(x^2+xx_0+x_0^2)+2(x+x_0)\\ =& x_0^2+x_0\cdot x_0+x_0^2+2(x_0+x_0)=3x_0^2+4x_0\end{aligned}$$

Ta thấy rằng giới hạn trên luôn tồn tại với mọi $x_0 \in \mathbb{R}$ nên thay x_0 thành x ta có đạo hàm $f'(x) = 3x^2 + 4x$ của $f(x)$ trên $\mathbb{R}$.

12.4.3 Vi phân

Trong cách ký hiệu

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

ta thay Δy thành dy và Δx thành dx thì vi phân được định nghĩa là

$$f'(x) = \frac{dy}{dx} \Leftrightarrow dy = f'(x)\,dx \tag{12.10}$$

Cách ký hiệu vi phân có ý nghĩa là vế trái là vi phân theo biến y và vế phải là vi phân theo biến x. Do $y = f(x)$ nên khi vi phân hai vế sẽ cho ra $dy = f'(x)\,dx$ (vế trái là đa thức bậc 1 biến y).

Ví dụ phương trình $y^2 = x^3 + 4x - 7$ thì khi vi phân hai vế ta có

$$(y^2)'\,dy = (x^3 + 4x - 7)\,dx \Leftrightarrow 2y\,dy = (3x^2 + 4)\,dx$$

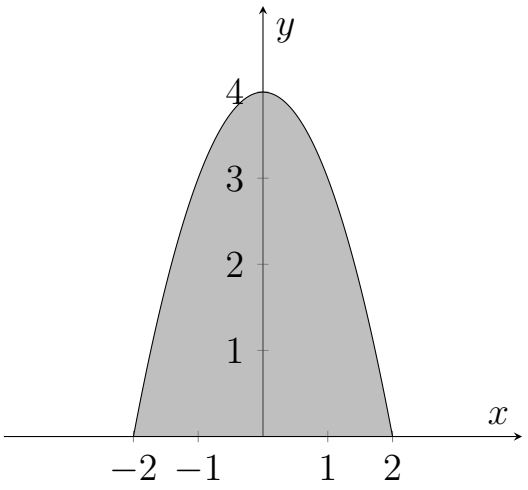
12.5 Tích phân

Tích phân là khái niệm quan trọng trong giải tích. Sau đây sẽ trình bày cách tính tích phân từ định nghĩa.

Xét phương trình của một đường cong $y = f(x) > 0$ trên đoạn $[a, b]$.

Theo định nghĩa, tích phân từ a tới b là diện tích phần hình phẳng giới hạn bởi đường cong $y = f(x)$, trục hoành Ox và hai trục đứng $x = a$, $x = b$.

Ở hình 12.9, diện tích phần tô màu xám là tích phân từ -2 tới 2 của hàm số $f(x) = -x^2 + 4$.



Hình 12.9: Tích phân từ -2 tới 2 của $f(x) = -x^2 + 4$

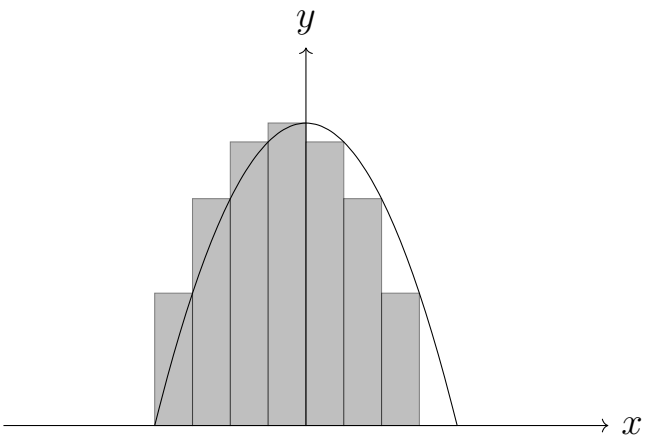
Chúng ta có thể tính diện tích hình chữ nhật, hình thang, hình vuông. Vậy có cách nào để tính diện tích một hình giới hạn bởi các đường cong bất kì không? Có đấy. Chúng ta sẽ tính xấp xỉ bằng tổng diện tích các hình chữ nhật.

Ví dụ với hàm số $f(x) = -x^2 + 4$ ở trên, ta chia đoạn $[a, b]$ thành n phần bằng nhau

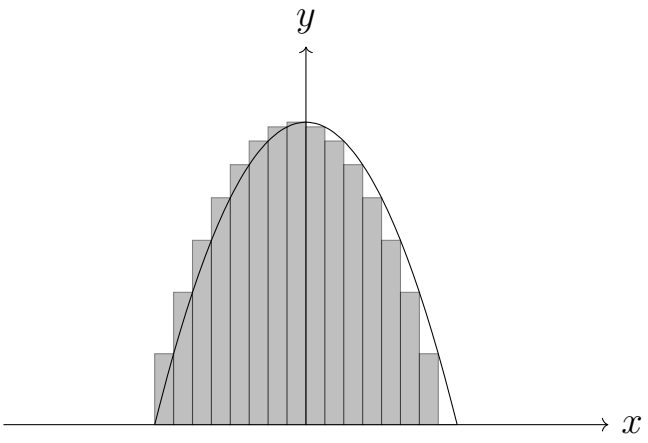
$$a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$$

Trong đó $x_{i+1} - x_i$ cố định và bằng $\dfrac{b-a}{n}$.

Đối với hình 12.10 ta xấp xỉ bằng 7 hình chữ nhật. Đối với hình 12.11 ta xấp xỉ bằng 15 hình chữ nhật. Đối với hình 12.12 ta xấp xỉ bằng 31 hình chữ nhật.



Hình 12.10: Xấp xỉ diện tích bởi 7 hình chữ nhật

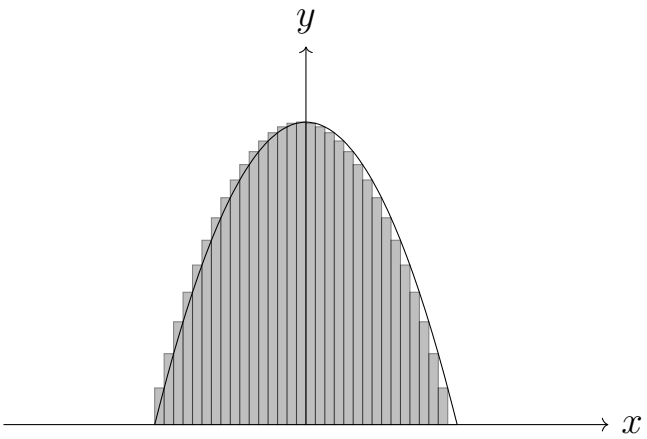


Hình 12.11: Xấp xỉ diện tích bởi 15 hình chữ nhật

Càng dùng nhiều hình chữ nhật, tổng diện tích của chúng càng gần với diện tích cần tìm, hay là tích phân cần tìm.

Ở ba hình trên, mỗi hình chữ nhật trong đó có chiều rộng bằng nhau là $\dfrac{b-a}{n}$ với n là số đoạn. Chiều dài là $f(x_i)$ với $x_i = a + \dfrac{b-a}{n}i$, $i = 1, 2, \ldots, n$ (biến sau).

Cụ thể hơn, hình chữ nhật từ x_{i-1} tới x_i sẽ có chiều dài là $f(x_i)$ và chiều rộng là $\dfrac{b-a}{n}$.



Hình 12.12: Xấp xỉ diện tích bởi 31 hình chữ nhật

Khi đó, tổng diện tích của các hình chữ nhật là

$$\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i) = \sum_{i=1}^{n}\frac{b-a}{n}f(x_i) \tag{12.11}$$

Khi số lượng hình chữ nhật tăng lên tới vô hạn thì tổng diện tích sẽ tiến tới diện tích chính xác của hình cần tìm, hay nói cách khác là tích phân. Do đó kết quả sẽ là

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n}\frac{b-a}{n}f(x_i), \quad x_i = a + \frac{b-a}{n}i \tag{12.12}$$

Ví dụ, tính tích phân từ -2 tới 2 của hàm số $f(x) = -x^2 + 4$ ở trên. Ta có $b = 2$ và $a = -2$ nên

$$\begin{aligned}\frac{b-a}{n}f(x_i) =& \frac{4}{n}\Big(-\Big(-2+\frac{4}{n}i\Big)^2 + 4\Big)\\ =& \frac{4}{n}\Big(-4 + \frac{16}{n}i - \frac{16}{n^2}i^2 + 4\Big)\\ =& \frac{64}{n}\Big(\frac{i}{n} - \frac{i^2}{n^2}\Big)\end{aligned}$$

Tính tổng i từ 1 tới n ta có $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Tính tổng i^2 từ 1 tới n ta có $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

Suy ra

$$\sum_{i=1}^{n} \frac{64}{n}\left(\frac{i}{n} - \frac{i^2}{n^2}\right) = \frac{64}{n^2}\sum_{i=1}^{n} i - \frac{64}{n^3}\sum_{i=1}^{n} i^2$$
$$= -\frac{64}{n^2} \cdot \frac{n(n+1)}{2} - \frac{64}{n^3} \cdot \frac{n(n+1)(2n+1)}{6}$$

Khi n tiến tới vô cực thì biểu thức trên tiến tới $\frac{64}{2} - \frac{64 \cdot 2}{6} = \frac{32}{3}$. Đây chính là giá trị của tích phân $\int_{-2}^{2} -x^2 + 4\, dx$.

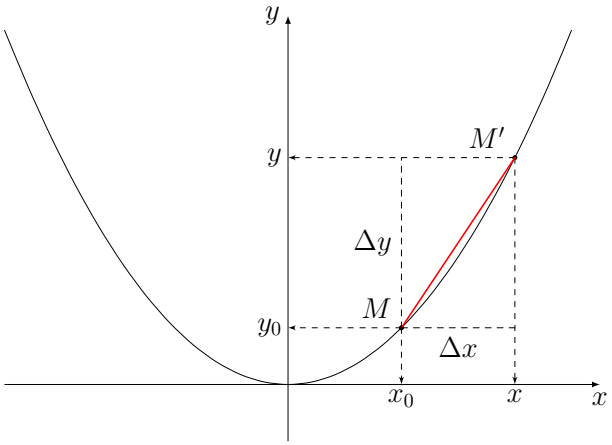
12.6 Tiếp tuyến và Tích phân đường

Xét hàm số $y = f(x)$ liên tục trên khoảng (a, b) chứa điểm x_0.

Gọi $M' = (x, y)$ là một điểm thuộc hàm số $y = f(x)$. Khi đó đạo hàm của $f(x)$ tại x_0 là giới hạn

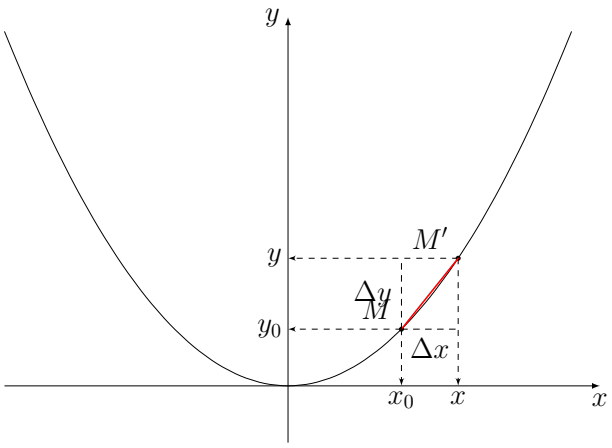
$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Xét hình 12.13, tỉ số $\Delta y / \Delta x$ là tangent của góc hợp bởi trục hoành Ox và đường thẳng MM'.



Hình 12.13: Hệ số góc 1

Tiếp theo, xét hình 12.14, ta thấy đường thẳng MM' ngày càng tiến sát lại với đường cong. Như vậy, khi Δx tiến tới 0 thì đường thẳng MM' cắt đường cong tại hai điểm càng sát nhau. Đến khi hai điểm đó trùng nhau, đường thẳng MM' chỉ đi qua đúng một điểm thuộc đường cong và khi đó MM' trở thành tiếp tuyến của đường cong tại điểm $M = (x_0, y_0)$.



Hình 12.14: Hệ số góc 2

Khi đó $f'(x_0)$ là tangent của góc hợp bởi MM' và trục hoành Ox, hay nói cách khác là *hệ số góc* của đường tiếp tuyến. Thêm nữa $f'(x_0) = \dfrac{\Delta y}{\Delta x} = \dfrac{y - y_0}{x - x_0}$ nên phương trình đường tiếp tuyến đi qua $M = (x_0, y_0)$ là

$$y = f'(x_0)(x - x_0) + y_0 \tag{12.13}$$

Dựa trên cùng ý tưởng với tích phân (chia diện tích dưới đường cong thành nhiều hình chữ nhật) và việc khảo sát hình học của tiếp tuyến ở trên, ta có thể tính *độ dài đường cong* bằng việc chia đường cong thành nhiều đường nhỏ và tính tổng của chúng.

Chương 13

Machine Learning

13.1 Linear Regression

Giả sử ta có N điểm dữ liệu đầu vào $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ với $\boldsymbol{x}_i \in \mathbb{R}^d$. Ứng với từng điểm dữ liệu đầu vào $\boldsymbol{x}_i$ ta có một đầu ra y_i. Nghĩa là ta có N cặp dữ liệu $(\boldsymbol{x}_i, y_i)$.

Mục tiêu là xây dựng hàm số $\hat{y} = f(x_1, x_2, \ldots, x_d)$ sao cho tổng sai số của y_i và $\hat{y}_i$ là nhỏ nhất, tức là

$$\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2 \to \min$$

Để hàm số đạt giá trị nhỏ nhất (hoặc lớn nhất) ta tìm cực trị của hàm số và khảo sát. Tuy nhiên không phải hàm số nào cũng đạo hàm được. Một cách tiếp cận đơn giản là sử dụng hàm tuyến tính, dễ xây dựng và luôn khả vi. Ta đặt

$$\hat{y} = f(x_1, x_2, \ldots, x_d) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d$$

Lúc này, hàm mất mát ở trên có dạng

$$\mathcal{L} = \sum_{i=1}^{N} \|y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}\|^2$$

Bình phương chuẩn Euclid chính là bình phương của vector. Do đó dưới dấu tổng là các hàm số bình phương. Khi đạo hàm riêng theo w_j ta có

$$\frac{\partial \mathcal{L}}{\partial w_j} = \sum_{i=1}^{N} 2x_{ij}(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}))$$

với $1 \leq j \leq d$. Với $j = 0$ có chút khác biệt, $\dfrac{\partial \mathcal{L}}{\partial w_0} = \sum_{i=1}^{N} 2(y_i - (w_0 + w_1 x_{i1} + \ldots + w_d x_{id}))$.

Ta cho các đạo hàm riêng $\dfrac{\partial \mathcal{L}}{\partial w_j}$ bằng 0 thì được

$$\begin{aligned}
&\sum_{i=1}^{N} x_{ij}(w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d x_{id}) = \sum_{i=1}^{N} x_{ij} y_i \\
\Leftrightarrow\ & w_0 \sum_{i=1}^{N} x_{ij} + w_1 \sum_{i=1}^{N} x_{ij} x_{i1} + w_2 \sum_{i=1}^{N} x_{ij} x_{i2} \\
&\qquad + \cdots + w_d \sum_{i=1}^{N} x_{ij} x_{id} = \sum_{i=1}^{N} x_{ij} y_i
\end{aligned}$$

Bây giờ chúng ta cần biểu diễn các dấu tổng lại thành dạng đại số (ma trận, vector) vì chúng sẽ được sử dụng để nhân với vector $\boldsymbol{w} = (w_0, w_1, \ldots, w_d)$.

Ta có $\sum_{i=1}^{N} x_{ij} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{pmatrix}$.

Ta cũng có $\sum_{i=1}^{N} x_{ij} x_{i1} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} \end{pmatrix} \cdot \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{pmatrix}$.

Cứ tương tự như vậy, ta xếp các dấu sigma thành dạng cột thì tương đương với

$$\begin{pmatrix} * & \sum_{i=1}^{N} x_{ij} & * \\ * & \sum_{i=1}^{N} x_{ij} x_{i1} & * \\ \vdots & \vdots & \vdots \\ * & \sum_{i=1}^{N} x_{ij} x_{id} & * \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ \cdots & \cdots & \ddots & \cdots \\ x_{1d} & x_{2d} & \cdots & x_{Nd} \end{pmatrix} \cdot \begin{pmatrix} * & x_{1j} & * \\ * & x_{2j} & * \\ \vdots & \vdots & \vdots \\ * & x_{Nj} & * \end{pmatrix}$$

Ghép các cột theo thứ tự j từ 0 tới d ta có

$$\begin{aligned}\begin{pmatrix} w_0 & w_1 & \cdots & w_d \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ \cdots & \cdots & \ddots & \cdots \\ x_{1d} & x_{2d} & \cdots & x_{Nd} \end{pmatrix} \\ \times \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & x_{N1} & \cdots & x_{Nd} \end{pmatrix} \\ = \begin{pmatrix} y_1 & y_2 & \cdots & y_N \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \ddots & \cdots \\ 1 & x_{N1} & \cdots & x_{Nd} \end{pmatrix}\end{aligned}$$

Hay nói cách khác, nếu ta đặt $\boldsymbol{w} = (w_0, w_1, \ldots, w_d)$ là ma trận hàng, $\boldsymbol{X}$ là ma trận có các hàng là các input, thì phương trình trên được viết lại là $\boldsymbol{wX}^T\boldsymbol{X} = \boldsymbol{yX}$.

Nếu đặt $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$ và $\boldsymbol{b} = \boldsymbol{yX}$ thì đây là hệ phương trình theo các ẩn $w_0, w_1, \ldots, w_d$. Tuy nhiên không phải lúc nào $\boldsymbol{A}$ cũng khả nghịch nên chúng ta sẽ sử dụng một khái niệm gọi là *giả nghịch đảo* để tìm nghiệm cho hệ phương trình.

Ký hiệu $\boldsymbol{A}^\dagger$ là giả nghịch đảo của ma trận $\boldsymbol{A}$. Khi đó nghiệm của hệ phương trình là $\boldsymbol{w} = \boldsymbol{bA}^\dagger$.

13.2 K-Means clustering

Một công việc thường được quan tâm là phân loại một nhóm các đối tượng thành những nhóm nhỏ hơn theo những tiêu chí nhất định.

Tương tự như phần trước, chúng ta có N điểm dữ liệu $\boldsymbol{x}_i$ thuộc $\mathbb{R}^d$. Ta muốn phân cụm các vector này vào những cluster (cụm) sao cho chúng gần nhau nhất (về mặt khoảng cách Euclid).

Giả sử ta muốn phân N điểm dữ liệu trên vào $K < N$ cluster. Ta cần tìm các điểm $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_K$ là tâm của các cụm, sao cho tổng khoảng cách từ các điểm $\boldsymbol{x}_i$ tới tâm cluster mà nó được phân vào là nhỏ nhất. Nghĩa là ứng với center $\boldsymbol{m}_1$ ta cần tìm các điểm $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_t}$ sao cho $\sum_{j=1}^{t} \|\boldsymbol{x}_{i_j} - \boldsymbol{m}_1\|^2$ nhỏ nhất. Tương tự cho các tâm khác.

Nhưng câu chuyện phức tạp ở đây là, tâm nằm ở đâu để có thể bao quát các điểm? Tâm được chọn phải có tính tổng quát, và việc phân các điểm vào cluster tương ứng với tâm thực hiện như thế nào?

Một kỹ thuật thường được sử dụng là *one-hot*. Với mỗi điểm dữ liệu $\boldsymbol{x}_i$ ta thêm một label $\boldsymbol{y}_i = (y_{i1}, \cdots y_{iK})$. Điểm $\boldsymbol{x}_i$ sẽ thuộc cluster j khi $y_{ij} = 1$, không thuộc thì bằng 0. Như vậy chỉ có đúng một phần tử của $\boldsymbol{y}_i$ bằng 1, còn lại bằng 0. Như vậy ràng buộc của $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iK})$ là $y_{ij} \in \{0, 1\}$ và $\sum_{j=1}^{K} y_{ij} = 1$.

Khi đó, ta mong muốn phân các điểm $\boldsymbol{x}_i$ vào cluster $\boldsymbol{m}_k$ để khoảng cách tới tâm $\boldsymbol{m}_k$ là ngắn nhất, hay $\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 \to \min$. Thêm nữa, với cách ký hiệu y_{ij} như trên, biểu thức tương đương với

$$\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 = y_{ik}\|\boldsymbol{x}_i - \boldsymbol{m}_k\|^2 = \sum_{j=1}^{K} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

vì điểm $\boldsymbol{x}_i$ sẽ thuộc cluster $\boldsymbol{m}_k$ nào đó với $1 \leq k \leq K$.

Sai số cho toàn bộ dữ liệu lúc này sẽ là

$$\mathcal{L}(\boldsymbol{Y}, \boldsymbol{M}) = \sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

Ta cần tối ưu $\boldsymbol{Y}$ và $\boldsymbol{M}$. Việc tối ưu hai ma trận cùng lúc là rất khó thậm chí bất khả thi. Do đó chúng ta có một cách tiếp cận khác là luân phiên cố định một bên và tối ưu bên còn lại. Từ đó công việc được chia làm hai bước.

Bước 1. Cố định $\boldsymbol{M}$, tìm $\boldsymbol{Y}$.

Giả sử ta đã biết các center $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_K$. Lúc này ta cần phân các điểm $\boldsymbol{x}_i$ vào cluster gần nó nhất. Dễ thấy rằng center gần nó nhất sẽ có khoảng cách Euclid ngắn nhất. Do đó ta tìm j sao cho $\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$ đạt nhỏ nhất. Không cần thiết phải tính căn bậc hai để giảm độ phức tạp.

Bước 2. Cố định $\boldsymbol{Y}$, tìm $\boldsymbol{M}$.

Khi đã biết $\boldsymbol{Y}$ tức là ta đã biết điểm nào được phân vào cluster nào. Khi đó ta cần tìm tâm cho từng cluster. Gọi $l(\boldsymbol{m}_j)$ là hàm tổng bình phương khoảng cách các điểm trong cluster tới tâm $\boldsymbol{m}_j$. Nghĩa là

$$l(\boldsymbol{m}_j) = \sum_{i=1}^{N} y_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|^2$$

Mục tiêu của chúng ta là tối ưu tâm $\boldsymbol{m}_j$. Do đó ta đạo hàm theo vector $\boldsymbol{m}_j$

thu được $\frac{\partial l(\boldsymbol{m}_j}{\partial \boldsymbol{m}_j} = \sum_{i=1}^{N} 2y_{ij}(\boldsymbol{x}_i - \boldsymbol{m}_j)$. Cho đạo hàm bằng 0 và biến đổi ta có

$$\begin{aligned} & 2\sum_{i=1}^{N} y_{ij}(\boldsymbol{x}_i - \boldsymbol{m}_j) = 0 \\ \Leftrightarrow & \boldsymbol{m}_j \sum_{i=1}^{N} y_{ij} = \sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i \\ \Leftrightarrow & \boldsymbol{m}_j = \frac{\sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i}{\sum_{i=1}^{N} y_{ij}} \end{aligned}$$

Để ý rằng, $\sum_{i=1}^{N} y_{ij}$ là số lượng điểm trong cluster, và $\sum_{i=1}^{N} y_{ij}\boldsymbol{x}_i$ là tổng các điểm trong cluster. Như vậy $\boldsymbol{m}_j$ là trung bình cộng các điểm trong cluster j.

Algorithm 3 Thuật toán K-Means clustering

Require: Dữ liệu $\boldsymbol{X}$ (có N điểm dữ liệu) và số cluster K
Ensure: Các center $\boldsymbol{M}$ và label $\boldsymbol{y}$ cho mỗi điểm dữ liệu

1. Chọn K điểm bất kì làm các cluster ban đầu.
2. Phân mỗi điểm dữ liệu vào cluster gần nó nhất (cố định M, tìm Y).
3. Nếu việc phân dữ liệu vào các cluster ở bước 2 không thay đổi so với trước đó thì dừng thuật toán.
4. Cập nhật center mới cho mỗi cluster bằng cách lấy trung bình cộng các điểm trong cluster (cố định Y, tìm M).
5. Quay lại bước 2.

13.3 Gradient Descent

Trong nhiều trường hợp chúng ta thường không thể tìm nghiệm của phương trình đạo hàm để từ đó tìm các cực trị địa phương. Một phương pháp hiệu quả là gradient descent.

13.3.1 Hàm một biến

Giả sử x^* là local extremum (cực trị địa phương) của hàm số $f(x)$. Khi đó chúng ta xây dựng dãy số $\{x_n\}$ hội tụ về x^*. Ý tưởng thực hiện là dựa trên nhận xét, nếu x_n nằm bên phải x^* thì x_{n+1} nằm giữa x^* và x_n. Ta đã biết nếu x^* là một điểm cực trị thì $f'(x) > 0$ với $x > x^*$ mà x_n đi từ bên phải sang bên trái (ngược chiều Ox nên mang dấu âm). Từ đó chúng ta có công thức chung sau

$$x_{n+1} = x_n - \eta f'(x_n)$$

Trong đó η là một số dương nhỏ, gọi là *learning rate* (tốc độ học).

Ta chọn x_0 là một điểm bất kì. Tuy nhiên việc chọn x_0 cũng có thể ảnh hưởng đến tốc độ hội tụ.

Ví dụ với hàm số $f(x) = x^2 + 5\sin x$. Ta có đạo hàm là $f'(x) = 2x + 5\cos x$. Việc giải phương trình đạo hàm bằng 0 là điều không dễ dàng. Do đó gradient descent tỏ ra hiệu quả trong trường hợp này.

Chọn $\eta = 0.1$ và $x_0 = 5$. Sau đó chọn $\eta = 0.1$ và $x_0 = -5$. Ta thấy trường hợp sau tốn ít vòng lặp hơn do $x_0 = -5$ gần điểm cực trị hơn (≈ -1.11).

13.3.2 Hàm nhiều biến

Lúc này đầu vào của hàm số là một vector $\boldsymbol{x}$. Đặt $\nabla f(\boldsymbol{x})$ là đạo hàm của hàm f theo vector $\boldsymbol{x}$. Tương tự, ta xây dựng dãy vector $\{\boldsymbol{x}_n\}$ hội tụ về cực trị $\boldsymbol{x}^*$. Công thức lúc này là

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \eta \cdot \nabla f(\boldsymbol{x}_n)$$

Ta đã biết đạo hàm của hàm số theo vector cũng là vector cùng cỡ. Do đó giả sử $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$ thì đạo hàm của nó là

$$\nabla f(\boldsymbol{x}) = \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\Big)$$

Với ví dụ là bài toán Linear Regression, lúc này hàm mất mát là

$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N} \|y_i - \boldsymbol{x}_i \boldsymbol{w}^T\|^2 = \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}^T\|^2$$

Đạo hàm của hàm mất mát là

$$\nabla \mathcal{L} = \frac{1}{N}(\boldsymbol{w}\boldsymbol{X}^T - \boldsymbol{y})\boldsymbol{X}$$

Lúc này, với vector khởi đầu $\boldsymbol{w}_0$ chúng ta xây dựng dãy $\{\boldsymbol{w}_n\}$ tới khi nhận được $\boldsymbol{w}_n/d < \varepsilon$, với d là độ dài vector $\boldsymbol{w}$.

13.4 Perception Learning Algorithm

Một trong những nhiệm vụ quan trọng nhất của ML là phân loại (tiếng Anh - classification).

Perception là thuật toán phân loại cho trường hợp đơn giản nhất khi có hai lớp. Nếu ta có các điểm dữ liệu cho trước trong không gian d chiều, ta muốn

tìm một siêu phẳng ($(d-1)$-phẳng) chia các điểm dữ liệu đó thành hai phần. Sau đó khi có một điểm dữ liệu mới ta chỉ cần bỏ nó vào bên này hoặc bên kia của siêu phẳng.

Trong dạng này, mỗi điểm dữ liệu được biểu diễn ở dạng cột của ma trận. Giả sử các điểm dữ liệu là $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, với $\boldsymbol{x}_i \in \mathbb{R}^d$, thì ma trận dữ liệu là $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T & \boldsymbol{x}_2^T & \cdots & \boldsymbol{x}_N^T \end{pmatrix}$. Ta gọi nhãn tương ứng với N điểm dữ liệu trên là vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ với $y_i = 1$ nếu $\boldsymbol{x}_i$ thuộc class xanh, và $y_i = -1$ nếu $\boldsymbol{x}_i$ thuộc class đỏ.

Một siêu phẳng có phương trình là

$$f_{\boldsymbol{w}}(\boldsymbol{x}) = w_0 + w_1 x_1 + \ldots + w_d x_d = \boldsymbol{w} \cdot \boldsymbol{x}^T$$

Một điểm thuộc nửa không gian (tạm gọi là *bên này*) đối với siêu phẳng thì $f_{\boldsymbol{w}}(\boldsymbol{x}) < 0$, nếu thuộc nửa bên kia thì $f_{\boldsymbol{w}}(\boldsymbol{x}) > 0$, nếu nằm trên phẳng thì bằng 0.

Gọi label($\boldsymbol{x}$) là nhãn của điểm $\boldsymbol{x}$. Khi đó điểm $\boldsymbol{x}$ thuộc một trong hai bên của phẳng nên label($\boldsymbol{x}$) = sgn($\boldsymbol{w} \cdot \boldsymbol{x}^T$) với sgn là hàm dấu. Ta quy ước sgn(0) = 1.

Khi một điểm bị phân loại sai class thì ta nói điểm đó bị **misclassified**. Ý tưởng của thuật toán là làm giảm thiểu số lượng điểm bị misclassified qua nhiều lần lặp. Đặt

$$J_1(\boldsymbol{w}) = \sum_{\boldsymbol{x}_i \in \mathcal{M}} (-y_i \cdot \text{sgn}(\boldsymbol{w} \cdot \boldsymbol{w}_i^T))$$

trong đó $\mathcal{M}$ là tập các điểm bị misclassified (tập này sẽ thay đổi theo $\boldsymbol{w}$.

Nếu $\boldsymbol{x}_i$ bị misclassified thì y_i và sgn($\boldsymbol{w} \cdot \boldsymbol{x}_i^T$) ngược dấu nhau. Nói cách khác, $-y_i \cdot \text{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}_i^T) = 1$. Từ đó $J_1(\boldsymbol{w})$ là hàm đếm số lượng điểm bị misclassified. Ta thấy rằng $J_1(\boldsymbol{w}) \geq 0$ nên ta cần tối ưu để hàm này đạt giá trị nhỏ nhất bằng 0. Khi đó không điểm nào bị misclassified.

Tuy nhiên có một vấn đề. Hàm $J_1(\boldsymbol{w})$ là hàm rời rạc (hàm sgn) nên rất khó tối ưu vì không thể tính đạo hàm. Do đó chúng ta cần một cách tiếp cận khác, một hàm mất mát khác tốt hơn.

Nếu ta bỏ đi hàm sgn thì có hàm

$$J(\boldsymbol{w}) = \sum_{\boldsymbol{x}_i \in \mathcal{M}} (-y_i \cdot \boldsymbol{w} \cdot \boldsymbol{x}^T)$$

Nhận xét. Một điểm bị misclassified nằm càng xa biên giới (siêu phẳng) thì giá trị $\boldsymbol{w} \cdot \boldsymbol{x}_i^T$ càng lớn, tức là hàm J đi ra xa so với giá trị nhỏ nhất. Hàm J cũng đạt min ở 0 nên ta cũng có thể dùng hàm này để loại bỏ các điểm bị misclassified.

Lúc này hàm $J(\boldsymbol{x})$ khả vi nên ta có thể dùng GD hoặc SGD để tìm nghiệm cho bài toán.

Nếu xét tại một điểm thì

$$J(\boldsymbol{w}, \boldsymbol{x}_i, y_i) = -y_i \cdot \boldsymbol{w} \cdot \boldsymbol{x}_i^T \Rightarrow \frac{\partial J}{\partial \boldsymbol{w}} = -y_i \boldsymbol{x}_i$$

Khi đó quy tắc để cập nhật là $\boldsymbol{w} = \boldsymbol{w} + \eta \cdot y_i \cdot \boldsymbol{x}_i$ với η là learning rate (thường chọn bằng 1). Nói cách khác ta đang xây dựng dãy $\{\boldsymbol{w}_n\}$ hội tụ lại nghiệm bài toán với công thức $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \eta \cdot y_i \cdot \boldsymbol{x}_i$.

Thuật toán PLA có thể được mô tả như sau:

1. Chọn ngẫu nhiên vector $\boldsymbol{w}$ với w_i xấp xỉ 0.
2. Duyệt ngẫu nhiên qua các $\boldsymbol{x}_i$:
   - Nếu $\boldsymbol{x}_i$ được phân lớp đúng, tức $\text{sgn}(\boldsymbol{w} \cdot \boldsymbol{x}_i^T) = y_i$ thì ta không cần làm gì.
   - Nếu $\boldsymbol{x}_i$ bị misclassified, ta cập nhật $\boldsymbol{w}$ theo công thức $\boldsymbol{w} = \boldsymbol{w} + \eta \cdot y_i \cdot \boldsymbol{x}$.
3. Kiểm tra xem có bao nhiêu điểm bị misclassified. Nếu không còn điểm nào thì ta dừng thuật toán, ngược lại thì quay lại bước 2.

Tài liệu tham khảo

[1] Euclid. *Euclid's Elements of Geometry*. Revised and corrected. Richard Fitzpatrick. ISBN: 978-0-6151-7984-1.

[2] *Euclid of Alexandria*. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Euclid/.

[3] Jeffrey Hoffstein, Jill Pipher и Joseph H. Silverman. *An Introduction to Mathematical Cryptography*. Англ. 2014. DOI: 10.1007/978-1-4939-1711-2.

[4] John Casey, Euclid. *The First Six Books of the Elements of Euclid*. 2007.

[5] *Pythagoras of Samos*. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Pythagoras/.

[6] *Thales of Miletus*. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Thales/.