I Các khái niệm

1 Cực trị có điều kiện

1.1 Bài toán tối ưu

Xét bài toán tìm cực trị của hàm $w(\mathbf{x}, \mathbf{y})$ với ràng buộc:

$$g(\mathbf{x}, \mathbf{y}) = b,$$

trong đó b là một hằng số. Ta sử dụng phương pháp Lagrange để giải bài toán này.

1.2 Hàm Lagrange

Hàm Lagrange được định nghĩa:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = w(\mathbf{x}, \mathbf{y}) + \lambda[b - g(\mathbf{x}, \mathbf{y})],$$

với λ là nhân tử Lagrange.

1.3 Các điều kiện cần

Để tìm cực trị, cần thỏa mãn:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{y}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0.$$

1.4 Ví dụ áp dụng

Xét bài toán tối ưu:

$$\mathbf{x} = \arg\max ||A\mathbf{x}||_2^2 \quad \text{subject to } ||\mathbf{x}||_2 = 1,$$

trong đó:

- $A \in \mathbb{R}^{m \times n}$ là ma trận,
- $||\mathbf{x}||_2 = 1$ là điều kiện chuẩn hóa vector.

1.5 Hàm Lagrange

Hàm Lagrange trong trường hợp này là:

$$\mathcal{L}(\mathbf{x}, \lambda) = ||A\mathbf{x}||_2^2 + \lambda(1 - ||\mathbf{x}||_2^2).$$

1.6 Giải bài toán

Từ hàm Lagrange, ta tính:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 2A^T A\mathbf{x} - 2\lambda\mathbf{x} = 0.$$

Suy ra:

$$(A^T A - \lambda I)\mathbf{x} = 0 \ (1)$$

với I là ma trận đơn vị. Đây là phương trình giá trị riêng:

$$\lambda \text{ là trị riêng của } A^T A, \quad \mathbf{x} \text{ là vector riêng tương ứng.}$$

Từ (1), ta nhân hai vế của (1) cho x^T, ta được:

$$x^T A^T A x = \lambda x^T x$$

$\iff$

$$||A\mathbf{x}||_2^2 = \lambda \cdot 1$$

=> $||A\mathbf{x}||_2^2$ đạt giá trị lớn nhất khi đạt giá trị lớn nhất

Kết quả

- Giá trị cực đại của $||A\mathbf{x}||_2^2$ đạt được khi $\mathbf{x}$ là vector riêng tương ứng với trị riêng lớn nhất của $A^T A$.
- Trong PCA, điều này tương ứng với việc chọn các vector riêng tương ứng với các trị riêng lớn nhất để tối đa hóa phương sai giữ lại.

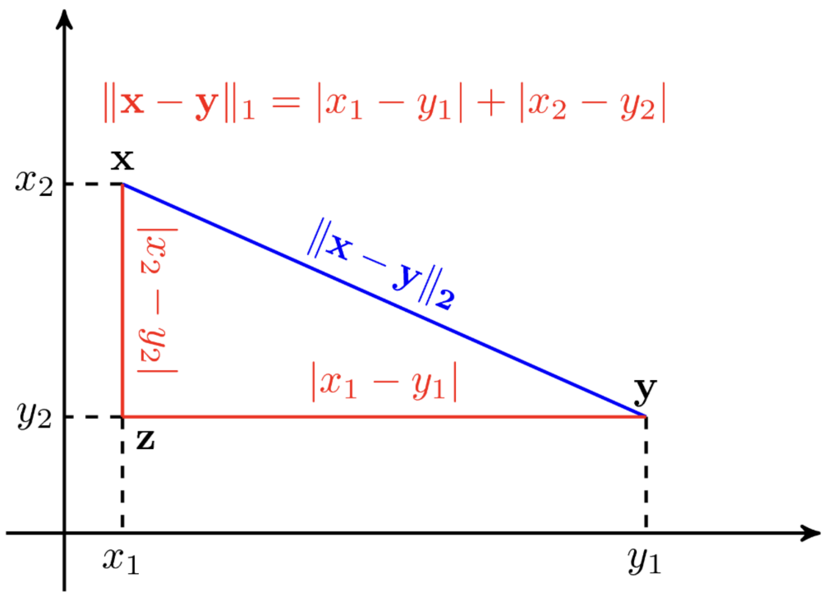
2 Chuẩn (Norm)

Một hàm số $f(x)$ ánh xạ một điểm x từ không gian n chiều sang tập số thực một chiều được gọi là Norm, nếu thoả:

- $f(x) \geq 0$, dấu " $=$ " xảy ra khi và chi khi $x = 0$ (khoảng cách giữa hai điểm x, y không thể là một số âm)
- $f(\alpha x) = |\alpha| f(x), \forall \alpha \in \mathbb{R}$ (nếu ba điểm y, v, z thẳng hàng, hơn nữa $v - y = \alpha(v - z)$ thì khoảng cách giữa v, y gấp $|\alpha|$ lần khoảng cách giữa v, z)
- $f(x_1) + f(x_2) \geq f(x_1 + x_2), \forall x_1, x_2 \in \mathbb{R}^n$ (bất đẳng thức tam giác)

Một số norm phổ biến:

- **Norm 2 (Euclid)**: $||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.
- **Norm 1**: $||\mathbf{x}||_1 = |x_1| + |x_2| + \cdots + |x_n|$.
- **Norm ∞**: $||\mathbf{x}||_\infty = \max(|x_i|)$ với $i = 1, 2, \ldots, n$.



Hình 1: Norm 1 và Norm 2 trong không gian 2 chiều

3 Chuẩn (Norm) của Ma Trận

Với một ma trận $A \in \mathbb{R}^{m \times n}$, chuẩn thường dùng là:

Chuẩn Frobenius:

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

Chuẩn (Norm) 2 của ma trận:

$$||A||_2 = ||A||_2^2 = \lambda_{max}$$

với λ là trị riêng lớn nhất của ma trận A

4 Trị riêng và Vector riêng

Trị riêng (*eigenvalue*) λ và vector riêng (*eigenvector*) x của một ma trận vuông $A \in \mathbb{R}^{n \times n}$ được định nghĩa như sau: Nếu tồn tại một số vô hướng λ và một vector $x \neq 0$ sao cho phương trình

$$Ax = \lambda x$$

thỏa mãn, thì λ gọi là **trị riêng** của ma trận A, và x gọi là **vector riêng** tương ứng với trị riêng đó.

Điều này có nghĩa là khi bạn nhân ma trận A với vector x, kết quả là một vector có cùng phương với x, chỉ khác ở yếu tố tỷ lệ λ. Nói cách khác, vector x không thay đổi hướng khi bị biến đổi bởi ma trận A, chỉ thay đổi độ dài (tỉ lệ λ).

5 Các tính chất của trị riêng và vector riêng:

- **Tính chất 1:** Nếu x là vector riêng của A ứng với trị riêng λ, thì bất kỳ vector kx (với $k \neq 0$) cũng sẽ là vector riêng ứng với trị riêng λ. Điều này có nghĩa là vector riêng không thay đổi khi nhân với một hằng số khác không.
- **Tính chất 2:** Mọi ma trận vuông bậc n đều có n trị riêng, tính cả các trị riêng lặp lại (multiplicities), và các trị riêng này có thể là số thực hoặc số phức.
- **Tính chất 3:** Nếu ma trận là đối xứng, tức là $A = A^T$, thì tất cả các trị riêng của nó đều là **số thực**.
- **Tính chất 4:** Nếu ma trận là xác định dương, tức là mọi vector $u \neq 0$ đều thỏa mãn $u^T Au > 0$, thì tất cả các trị riêng của ma trận này đều là **số thực dương**.
- **Tính chất 5:** Nếu ma trận là nửa xác định dương, tức là với mọi vector $u \neq 0$, ta có $u^T Au \geq 0$, thì tất cả các trị riêng của ma trận này đều là **số thực không âm** (tức là $\lambda \geq 0$).

Ma trận xác định dương: Một ma trận A được gọi là xác định dương nếu với mọi vector $u \neq 0$, ta có $u^T Au > 0$. Điều này có nghĩa là giá trị của $u^T Au$ luôn dương với mọi vector không bằng 0.

Giả sử u là một vector riêng ứng với trị riêng λ của A, tức là

$$Au = \lambda u.$$

Khi nhân cả hai vế của phương trình này với u^T, ta có

$$u^T Au = \lambda u^T u.$$

Vì $u^T u$ chính là chuẩn 2 của u, tức là $u^T u = \|u\|^2$, ta có:

$$u^T Au = \lambda \|u\|^2.$$

Tại sao $\lambda \geq 0$?

Vì A là nửa xác định dương, ta biết rằng $u^T Au \geq 0$ với mọi $u \neq 0$. Mặt khác, $\|u\|^2 > 0$ đối với mọi vector $u \neq 0$. Do đó, từ phương trình

$$u^T Au = \lambda \|u\|^2$$

và $u^T Au \geq 0$, ta suy ra rằng

$$\lambda \geq 0.$$

Vậy kết luận là nếu ma trận A xác định dương hoặc nửa xác định dương, thì tất cả các trị riêng λ của nó đều là số thực không âm.

6 Trực giao và trực chuẩn

6.1 Trực giao (Orthogonality)

Trong không gian Euclid $\mathbb{R}^n$, hai vectơ u và v được gọi là *trực giao* nếu và chỉ nếu tích vô hướng của chúng bằng 0, tức là:

$$u \cdot v = u^T v = 0.$$

Điều này có nghĩa là góc giữa hai vectơ u và v là 90 độ. Các tính chất của vectơ trực giao bao gồm:

- Nếu u và v là trực giao, thì chúng không ảnh hưởng lẫn nhau khi chiếu lên nhau.
- Tập hợp các vectơ trực giao có thể tạo thành một cơ sở của không gian, và nếu chúng cũng có độ dài chuẩn 1, chúng sẽ tạo thành một cơ sở trực chuẩn.

Ví dụ: Hai vectơ trong không gian $\mathbb{R}^2$ là $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ và $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ là trực giao vì:

$$u \cdot v = 1 \times 0 + 0 \times 1 = 0.$$

6.2 Hệ trực giao (Orthogonal System)

Trong không gian Euclid $\mathbb{R}^n$, một tập hợp các vectơ $\{v_1, v_2, \ldots, v_k\}$ được gọi là *hệ trực giao* nếu mọi cặp vectơ trong tập hợp đó là trực giao với nhau. Cụ thể, các vectơ v_i và v_j trong hệ trực giao thỏa mãn điều kiện sau:

$$v_i \cdot v_j = 0 \quad \text{với mọi } i \neq j.$$

Điều này có nghĩa là tích vô hướng giữa mỗi cặp vectơ khác nhau trong hệ bằng 0, tức là các vectơ này vuông góc với nhau. Nếu tất cả các vectơ trong hệ là trực giao, thì chúng không ảnh hưởng lẫn nhau khi chiếu lên nhau.

Ví dụ: Trong không gian $\mathbb{R}^3$, tập hợp các vectơ $\{v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\}$ là một hệ trực giao vì:

$$v_1 \cdot v_2 = v_1 \cdot v_3 = v_2 \cdot v_3 = 0.$$

6.3 Trực chuẩn (Orthonormal System)

Một hệ trực giao $\{v_1, v_2, \ldots, v_k\}$ được gọi là *hệ trực chuẩn* nếu tất cả các vectơ trong hệ không chỉ trực giao mà còn có độ dài chuẩn bằng 1. Tức là, mỗi vectơ v_i trong hệ phải thỏa mãn:

$$\|v_i\| = 1 \quad \text{với mọi } i.$$

Nói cách khác, mỗi vectơ trong hệ trực chuẩn là một vectơ đơn vị và tất cả các vectơ đều vuông góc với nhau. Hệ trực chuẩn là một đặc biệt quan trọng trong nhiều lĩnh vực toán học và khoa học máy tính vì tính đơn giản và dễ sử dụng của nó trong các phép tính.

Ví dụ: Trong không gian $\mathbb{R}^3$, hệ vectơ $\{v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, v_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, v_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}\}$ là hệ trực chuẩn vì:

$$v_1 \cdot v_2 = v_1 \cdot v_3 = v_2 \cdot v_3 = 0, \quad \|v_1\| = \|v_2\| = \|v_3\| = 1.$$

Tính chất và ứng dụng của hệ trực chuẩn

- **Cơ sở trực chuẩn:** Một tập hợp các vectơ trực chuẩn tạo thành một cơ sở cho không gian nếu các vectơ trong hệ là độc lập tuyến tính và có thể dùng để biểu diễn bất kỳ vectơ nào trong không gian đó.
- **Tính đơn giản trong phép tính:** Khi làm việc với hệ trực chuẩn, các phép chiếu và phép tính tích vô hướng trở nên rất đơn giản vì các vectơ trực chuẩn có độ dài chuẩn 1 và là trực giao.
- **Giải quyết bài toán trực giao hóa:** Hệ trực chuẩn có thể được tạo ra từ hệ trực giao bằng cách chuẩn hóa các vectơ trong hệ. Phương pháp Gram-Schmidt là một thuật toán nổi tiếng để thực hiện điều này.

Ứng dụng: Các hệ trực chuẩn thường được sử dụng trong các bài toán phân tích số học tuyến tính, như trong phân tích Fourier, phân tích riêng, hay trong các thuật toán học máy (machine learning) và xử lý tín hiệu (signal processing).

7 Ma trận trực giao (Orthogonal Matrix)

Một ma trận vuông $U \in \mathbb{R}^{m\times m}$ được gọi là ma trận trực giao nếu các cột của nó là các vectơ trực chuẩn (orthonormal). Điều này có nghĩa là:

$$U^T U = U z U^T = I$$

trong đó I là ma trận đơn vị bậc m.

Tính chất quan trọng:

- $U^{-1} = U^T$. Tức là nghịch đảo của một ma trận trực giao chính là chuyển vị của nó.
- Định thức của ma trận trực giao luôn có giá trị bằng ± 1:

  $$\det(U) = \pm 1$$

  Điều này được suy ra từ tính chất:

  $$\det(U) = \det(U^T) \quad \text{và} \quad \det(U) \cdot \det(U^T) = \det(I) = 1.$$

7.1 Phép Xoay (Rotation) và Tích Vô Hướng

Ma trận trực giao có thể được sử dụng để thực hiện phép xoay các vectơ trong không gian Euclid $\mathbb{R}^m$. Cụ thể, nếu $x, y \in \mathbb{R}^m$ là hai vectơ và $U \in \mathbb{R}^{m\times m}$ là một ma trận trực giao, thì phép xoay x và y bởi U sẽ không thay đổi tích vô hướng giữa chúng:

$$(Ux)^T(Uy) = x^T U^T U y = x^T y$$

Điều này chứng tỏ phép xoay không thay đổi khoảng cách (norm) hay góc giữa các vectơ, tức là nó bảo toàn các tính chất hình học của các vectơ.

7.2 Hệ Trực Giao và Tính Chất Phân Rã (QR Decomposition)

Một hệ trực giao có thể được sử dụng trong quá trình phân rã QR. Khi phân rã một ma trận $A \in \mathbb{R}^{m\times n}$, ta có thể viết:

$$A = QR$$

trong đó $Q \in \mathbb{R}^{m\times m}$ là ma trận trực giao (hoặc trực chuẩn) và $R \in \mathbb{R}^{m\times n}$ là ma trận tam giác trên (upper triangular). Quá trình phân rã này rất quan trọng trong các bài toán giải hệ phương trình và tối ưu.

7.3 Hệ Trực Giao và Các Phép Biến Hình

Ma trận trực giao có thể mô tả các phép biến hình hình học như xoay, phản xạ hoặc kết hợp giữa chúng. Các phép biến hình này bảo toàn độ dài của các vectơ, do đó, chúng là các phép biến hình bảo toàn khoảng cách. Cụ thể, nếu $U \in \mathbb{R}^{m\times m}$ là ma trận trực giao và $x \in \mathbb{R}^m$ là một vectơ, thì phép biến hình U làm biến đổi x thành Ux và giữ nguyên độ dài của x:

$$\|Ux\| = \|x\|$$

Điều này có nghĩa là phép biến đổi này không làm thay đổi độ dài (norm) của vectơ x, chỉ thay đổi hướng của nó trong không gian.

7.4 Phép Biến Hình và Các Ma Trận Con

Giả sử $\hat{U} \in \mathbb{R}^{m\times r}$ là một ma trận con của ma trận trực giao $U \in \mathbb{R}^{m\times m}$, được tạo thành từ r cột của U. Khi đó, $\hat{U}^T\hat{U} = I_r$, tức là các cột của ma trận con $\hat{U}$ tạo thành một hệ trực chuẩn trong không gian con có chiều r.

Tính chất này rất hữu ích trong các bài toán phân rã ma trận hoặc trong các phép toán giải bài toán tối ưu với các không gian con.

8 Ma trận Hiệp phương sai

Ma trận hiệp phương sai cho một tập hợp các biến ngẫu nhiên $X = (X_1, X_2, \ldots, X_n)$ có dạng:

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \dots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \dots & \mathrm{Var}(X_n) \end{pmatrix}$$

Trong đó: - $\mathrm{Var}(X_i)$ là phương sai của biến ngẫu nhiên X_i, - $\mathrm{Cov}(X_i, X_j)$ là hiệp phương sai giữa các biến ngẫu nhiên X_i và X_j.

8.1 Các tính chất của ma trận hiệp phương sai:

- Ma trận hiệp phương sai luôn là ma trận đối xứng, tức là $\Sigma = \Sigma^T$, vì $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$.
- Phương sai của mỗi biến trong ma trận là các phần tử trên đường chéo chính, tức là $\mathrm{Var}(X_1), \mathrm{Var}(X_2), \dots, \mathrm{Var}(X_n)$.
- Hiệp phương sai giữa các cặp biến ngẫu nhiên không nhất thiết phải là dương, có thể âm hoặc bằng không.
- Ma trận hiệp phương sai có thể được sử dụng để tính toán các tính chất thống kê như độ lệch chuẩn, độ tương quan, v.v.

8.2 Cách tính ma trận hiệp phương sai

Giả sử bạn có một tập dữ liệu với các biến ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$, ma trận hiệp phương sai được tính như sau:

$$\Sigma = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \dots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \dots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}$$

Trong đó: - $\mathrm{Var}(X_i) = \mathrm{Cov}(X_i, X_i)$ là phương sai của biến X_i, - $\mathrm{Cov}(X_i, X_j)$ là hiệp phương sai giữa hai biến X_i và X_j.

Công thức tính hiệp phương sai giữa hai biến X_i và X_j là:

$$\mathrm{Cov}(X_i, X_j) = \frac{1}{m} \sum_{k=1}^{m} (x_{ik} - \overline{X_i})(x_{jk} - \overline{X_j})$$

Trong đó: - m là số lượng mẫu, - x_{ik} và x_{jk} là giá trị của biến X_i và X_j tại mẫu thứ k, - $\overline{X_i}$ và $\overline{X_j}$ là giá trị trung bình của các biến X_i và X_j.

Phương sai của biến X_i được tính bằng:

$$\mathrm{Var}(X_i) = \frac{1}{m} \sum_{k=1}^{m} (x_{ik} - \overline{X_i})^2$$

8.3 Áp dụng:

Cho một ma trận K kích thước m $\times N$, trong đó:

- K: Ma trận dữ liệu, với mỗi cột là một biến và mỗi hàng là một quan sát. - m: Số chiều (số hàng) của dữ liệu. - N: Số mẫu (số cột) của dữ liệu.

1. **Tính ma trận trung bình:**

$$M_{ij} = \text{mean}(K_i), \quad \text{với } M \text{ là ma trận kích thước } m \times N, \text{ trong đó:}$$

- M_{ij}: Giá trị trung bình của biến thứ i trên tất cả các quan sát. - Mỗi hàng trong M lặp lại giá trị trung bình của biến tương ứng.

2. **Tính ma trận đã chuẩn hóa:**

$$D = K - M, \quad \text{trong đó:}$$

- D: Ma trận dữ liệu đã được chuẩn hóa, tức là đã trừ đi giá trị trung bình của từng biến. - Việc chuẩn hóa đảm bảo mỗi biến có giá trị trung bình bằng 0.

3. **Tính ma trận hiệp phương sai:**

$$C = \frac{1}{N} DD^T, \quad \text{trong đó:}$$

- C: Ma trận hiệp phương sai, kích thước $m \times m$. - DD^T: Tích ma trận D với chuyển vị của chính nó, đại diện cho sự tương quan giữa các chiều. - $\frac{1}{N}$: Hệ số chuẩn hóa theo số mẫu N, để đảm bảo giá trị trong C phản ánh đúng mức độ hiệp phương sai.

Bảng các đạo hàm thường gặp

1. Cho vector

Hàm số	**Đạo hàm**
$a^T x$	a
$x^T A x$	$(A + A^T)x$
$x^T x = \|x\|^2$	$2x$
$\|Ax - b\|_2^2$	$2A^T(Ax - b)$
$a^T x^T x b$	$2a^T b$
$a^T x x^T b$	$(ab^T + ba^T)x$

2. Cho ma trận

Hàm số	**Đạo hàm**
$a^T X^T X b$	$X(ab^T + ba^T)$
$a^T X X^T b$	$(ab^T + ba^T)X$
$a^T Y X^T b$	$ba^T Y$
$a^T X^T Y b$	Yba^T
$a^T X Y^T b$	$(ab^T + ba^T)XY^T$
Yab^T	$a^T XY^T b$