

線性代數與微積分回顧

Afshine AMIDI 和 Shervine AMIDI

December 15, 2019

翻譯: kevingo. 審閱: Miyaya.

通用符號

□ **向量** – 我們定義 $x \in \mathbb{R}^n$ 是一個向量, 包含 n 維元素, $x_i \in \mathbb{R}$ 是第 i 維元素:

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

□ **矩陣** – 我們定義 $A \in \mathbb{R}^{m \times n}$ 是一個 m 列 n 行的矩陣, $A_{i,j} \in \mathbb{R}$ 代表位在第 i 列第 j 行的元素:

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

注意: 上述定義的向量 x 可以視為 $n \times 1$ 的矩陣, 或是更常被稱為行向量

□ **單位矩陣** – 單位矩陣 $I \in \mathbb{R}^{n \times n}$ 是一個方陣, 其主對角線皆為 1, 其餘皆為 0

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

注意: 對於所有矩陣 $A \in \mathbb{R}^{n \times n}$, 我們有 $A \times I = I \times A = A$

□ **對角矩陣** – 對角矩陣 $D \in \mathbb{R}^{n \times n}$ 是一個方陣, 其主對角線為非 0, 其餘皆為 0

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

注意: 我們令 D 為 $\text{diag}(d_1, \dots, d_n)$

矩陣運算

□ **向量-向量** – 有兩種類型的向量-向量相乘:

- 內積: 對於 $x, y \in \mathbb{R}^n$, 我們可以得到:

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

- 外積: 對於 $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, 我們可以得到:

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

□ **矩陣-向量** – 矩陣 $A \in \mathbb{R}^{m \times n}$ 和向量 $x \in \mathbb{R}^n$ 的乘積是一個大小為 \mathbb{R}^m 的向量, 使得:

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

其中 $a_{r,i}^T$ 是 A 的列向量、 $a_{c,j}$ 是 A 的行向量、 x_i 是 x 的元素

□ **矩陣-矩陣** – 矩陣 $A \in \mathbb{R}^{m \times n}$ 和 $B \in \mathbb{R}^{n \times p}$ 的乘積為一個大小 $\mathbb{R}^{m \times p}$ 的矩陣, 使得:

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

其中, $a_{r,i}^T, b_{r,i}^T$ 和 $a_{c,j}, b_{c,j}$ 分別是 A 和 B 的列向量與行向量

□ **轉置** – 一個矩陣的轉置矩陣 $A \in \mathbb{R}^{m \times n}$, 記作 A^T , 指的是其中元素的翻轉:

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

注意: 對於矩陣 A, B , 我們有 $(AB)^T = B^T A^T$

□ **可逆** – 一個可逆矩陣 A 記作 A^{-1} , 存在唯一的矩陣, 使得:

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

注意: 並非所有的方陣都是可逆的。同樣的, 對於矩陣 A, B 來說, 我們有 $(AB)^{-1} = B^{-1}A^{-1}$

□ **跡** – 一個方陣 A 的跡, 記作 $\text{tr}(A)$, 指的是主對角線元素之合:

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

注意: 對於矩陣 A, B 來說, 我們有 $\text{tr}(A^T) = \text{tr}(A)$ 及 $\text{tr}(AB) = \text{tr}(BA)$

□ **行列式** – 一個方陣 $A \in \mathbb{R}^{n \times n}$ 的行列式, 記作 $|A|$ 或 $\det(A)$, 可以透過 $A_{\setminus i, \setminus j}$ 來遞迴表示, 它是一個沒有第 i 列和第 j 行的矩陣 A :

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

注意: A 是一個可逆矩陣, 若且唯若 $|A| \neq 0$ 。同樣的, $|AB| = |A||B|$ 且 $|A^T| = |A|$

矩陣的性質

□ **對稱分解** – 給定一個矩陣 A ，它可以透過其對稱和反對稱的部分表示如下：

$$A = \underbrace{\frac{A + A^T}{2}}_{\text{對稱}} + \underbrace{\frac{A - A^T}{2}}_{\text{反對稱}}$$

□ **範數** – 範數指的是一個函式 $N: V \rightarrow [0, +\infty]$ ，其中 V 是一個向量空間，且對於所有 $x, y \in V$ ，我們有：

- $N(x + y) \leq N(x) + N(y)$
- 對一個純量來說，我們有 $N(ax) = |a|N(x)$
- 若 $N(x) = 0$ 時，則 $x = 0$

對於 $x \in V$ ，最常用的範數總結如下表：

範數	表示法	定義	使用情境
Manhattan, L^1	$\ x\ _1$	$\sum_{i=1}^n x_i $	LASSO regularization
Euclidean, L^2	$\ x\ _2$	$\sqrt{\sum_{i=1}^n x_i^2}$	Ridge regularization
p -norm, L^p	$\ x\ _p$	$\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Hölder inequality
Infinity, L^∞	$\ x\ _\infty$	$\max_i x_i $	Uniform convergence

□ **線性相關** – 當集合中的一個向量可以用被定義為集合中其他向量的線性組合時，則則稱此集合的向量為線性相關

注意：如果沒有向量可以如上表示時，則稱此集合的向量彼此為線性獨立

□ **矩陣的秩** – 一個矩陣 A 的秩記作 $\text{rank}(A)$ ，指的是其列向量空間所產生的維度，等價於 A 的線性獨立的最大最大行向量

□ **半正定矩陣** – 當以下成立時，一個矩陣 $A \in \mathbb{R}^{n \times n}$ 是半正定矩陣(PSD)，且記作 $A \succeq 0$ ：

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

注意：同樣的，一個矩陣 A 是一個半正定矩陣(PSD)，且滿足所有非零向量 x ， $x^T A x > 0$ 時，稱之為正定矩陣，記作 $A \succ 0$

□ **特徵值、特徵向量** – 給定一個矩陣 $A \in \mathbb{R}^{n \times n}$ ，當存在一個向量 $z \in \mathbb{R}^n \setminus \{0\}$ 時，此向量被稱為特徵向量， λ 稱之為 A 的特徵值，且滿足：

$$Az = \lambda z$$

□ **譜分解** – 令 $A \in \mathbb{R}^{n \times n}$ ，如果 A 是對稱的，則 A 可以被一個實數正交矩陣 $U \in \mathbb{R}^{n \times n}$ 給對角化。令 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ，我們得到：

$$\exists \Lambda \text{ 對角線}, \quad A = U \Lambda U^T$$

□ **奇異值分解** – 對於給定維度為 $m \times n$ 的矩陣 A ，其奇異值分解指的是一種因子分解技巧，保證存在 $m \times m$ 的單式矩陣 U 、對角線矩陣 $\Sigma \in \mathbb{R}^{m \times n}$ 和 $n \times n$ 的單式矩陣 V ，滿足：

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

矩陣導數

□ **梯度** – 令 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ 是一個函式，且 $A \in \mathbb{R}^{m \times n}$ 是一個矩陣。 f 相對於 A 的梯度是一個 $m \times n$ 的矩陣，記作 $\nabla_A f(A)$ ，滿足：

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

注意： f 的梯度僅在 f 為一個函數且該函數回傳一個純量時有效

□ **海森** – 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一個函式，且 $x \in \mathbb{R}^n$ 是一個向量，則一個 f 的海森對於向量 x 是一個 $n \times n$ 的對稱矩陣，記作 $\nabla_x^2 f(x)$ ，滿足：

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

注意： f 的海森僅在 f 為一個函數且該函數回傳一個純量時有效

□ **梯度運算** – 對於矩陣 A 、 B 、 C ，下列的梯度性質值得牢牢記住：

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

$$\nabla_{A^T} f(A) = (\nabla_A f(A))^T$$

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

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