

深度学习分析中必备的矩阵知识

朱梦

初稿于 2025-05-18, 修改于 2025-06-20

1. 向量 L_2 范数与二阶矩

设 $\mathbf{x} \in \mathbb{R}^n$, 有:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}, \quad (1)$$

$$\|\mathbf{x}\|_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} = \frac{\|\mathbf{x}\|_2}{\sqrt{n}}, \quad (2)$$

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|, \quad (3)$$

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2 \quad (4)$$

其中, $\|\mathbf{x}\|_2$ 表示向量的 L_2 范数, $\|\mathbf{x}\|_{\text{RMS}}$ 主要用来描述向量分量的平均尺度指标 (即平均数量级)。当且仅当向量 \mathbf{x} 中仅有一个非零分量时, 式(4)左边等号成立。当且仅当向量 \mathbf{x} 中所有非零分量的绝对值相等时, 式(4)右边等号成立。

假设 1 (i.i.d) $\forall x_i$ 独立同分布。

由大数定律有:

$$\lim_{n \rightarrow \infty} \|\mathbf{x}\|_{\text{RMS}}^2 = \lim_{n \rightarrow \infty} \frac{\|\mathbf{x}\|_2^2}{n} = \mathbb{E}[x_i^2] \quad (5)$$

2. 矩阵的谱范数、Frobenius 范数与 RMS 范数

设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 有:

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2, \quad (6)$$

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n w_{i,j}^2} \quad (7)$$

$$\|\mathbf{A}\|_{\text{RMS}} = \sqrt{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n w_{i,j}^2} = \frac{\|\mathbf{A}\|_F}{\sqrt{mn}} \quad (8)$$

其中, $\|\mathbf{A}\|_2$ 表示矩阵的谱范数, $\|\mathbf{A}\|_F$ 表示矩阵的 Frobenius 范数, $\|\mathbf{A}\|_{\text{RMS}}$ 主要用来描述矩阵元素的平均尺度指标。当且仅当 \mathbf{x} 是 \mathbf{A} 的右奇异向量, 且对应于最大奇异值 σ_{\max} , 式(6)中的等号成立。从奇异值角度分析, 谱范数等于矩阵最大的奇异值, Frobenius 范数等于矩阵全体奇异值的平方和的平方根, 所以总有:

$$\frac{\|\mathbf{A}\|_F}{\sqrt{\min(m, n)}} \leq \|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \text{rank}(\mathbf{A})\|\mathbf{A}\|_2 \quad (9)$$

这个不等式描述了谱范数与 Frobenius 范数的等价性, 即遇到由于谱范数的复杂性难以分析下去的问题, 可以考虑将谱范数换成 Frobenius 范数来得到一个近似结果。

3. 正交矩阵、对角矩阵、外积矩阵

对于 $n \times n$ 正交矩阵 \mathbf{Q} , 有:

$$\text{rank}(\mathbf{Q}) = n, \quad (10)$$

$$\sigma_i(\mathbf{Q}) = |\lambda_i| = 1, \quad (11)$$

$$\|\mathbf{Q}\|_2 = 1, \quad (12)$$

$$\|\mathbf{Q}\|_F = \sqrt{n} \quad (13)$$

对于对角矩阵 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 有:

$$\text{rank}(\mathbf{\Lambda}) = \sum_{i=1}^n |\text{sign}(\lambda_i)|, \quad (14)$$

$$\|\mathbf{\Lambda}\|_2 = \max_i (|\lambda_i|), \quad (15)$$

$$\|\mathbf{\Lambda}\|_F = \sum_{i=1}^n \lambda_i^2 \quad (16)$$

对于外积矩阵, 有:

$$\text{rank}(\mathbf{uv}^T) = 1 \Leftrightarrow \|\mathbf{uv}^T\|_2 = \|\mathbf{uv}^T\|_F = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \quad (17)$$

对于外积矩阵元素的符号化, 有:

$$\begin{aligned} \text{sign}(\mathbf{uv}^T)_{i,j} &= \text{sign}(u_i v_j) = \text{sign}(u_i) \text{sign}(v_j) \\ \Rightarrow \text{sign}(\mathbf{uv}^T) &= \text{sign}(\mathbf{u}) \text{sign}(\mathbf{v})^T \end{aligned} \quad (18)$$

4. 矩阵分解为近似（半）正交矩阵与标量的数乘

设 $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m \geq n$), 可以做到将其分解为近似列（半）正交矩阵 \mathbf{Q} 与标量 κ 的数乘:

$$\mathbf{A} = \kappa \mathbf{Q}, \text{ s.t. } \kappa = \arg \min_{\nu > 0} \left\| \frac{\mathbf{A}^T \mathbf{A}}{\nu^2} - \mathbf{I} \right\| \quad (19)$$

那么有:

$$\text{rank}(\mathbf{A}) = \min(m, n), \quad (20)$$

$$\|\mathbf{A}\|_2 = \|\kappa \mathbf{Q}\|_2 = \kappa \|\mathbf{Q}\|_2 \approx \kappa \times 1 = \kappa, \quad (21)$$

$$\|\mathbf{A}\|_F = \|\kappa \mathbf{Q}\|_F = \kappa \|\mathbf{Q}\|_F \approx \kappa \times \sqrt{\min(m, n)} \quad (22)$$

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\kappa \mathbf{Q}\mathbf{x}\|_2 \approx \kappa \|\mathbf{x}\|_2 \quad (23)$$

5. 矩阵是否为线性

在数学（特别是线性代数）中，一个运算 $f(\cdot)$ 被称为线性运算，必须同时满足以下两个条件:

(1) **可加性:** 对于任意输入向量，都有 $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$ 。

(2) **齐次性:** 对于任意标量，都有 $f(c\mathbf{x}) = cf(\mathbf{x})$ 。

设有如下运算:

$$\mathbf{y} = \mathbf{A}(t)\mathbf{x} \quad (24)$$

其中， $\mathbf{A}(t)$ 表示定义在 t 上的矩阵。如果任意时刻， $\mathbf{A}(t)$ 为固定不变的，那么式(24)为线性运算。如果 $\mathbf{A}(t)$ 随着 t 变化，那么式(24)为非线性运算。