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 2025-06-16

1. 向量 L_2 范数与二阶矩

设 $\boldsymbol{x} \in \mathbb{R}^n$，有：

$$||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}, \tag{1}$$

$$||\boldsymbol{x}||_{\text{RMS}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \frac{||\boldsymbol{x}||_2}{\sqrt{n}}, \tag{2}$$

$$||\boldsymbol{x}||_1 = \frac{1}{n}\sum_{i=1}^{n} |x_i|, \tag{3}$$

$$||\boldsymbol{x}||_2 \leq ||\boldsymbol{x}||_1 \leq \sqrt{n}||\boldsymbol{x}||_2 \tag{4}$$

其中，$||\boldsymbol{x}||_2$ 表示向量的 L_2 范数，$||\boldsymbol{x}||_{\text{RMS}}$ 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 $\boldsymbol{x}$ 中仅有一个非零分量时，式(4)左边等号成立。当且仅当向量 $\boldsymbol{x}$ 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：

$$\lim_{n\to\infty} ||\boldsymbol{x}||_{\text{RMS}}^2 = \lim_{n\to\infty} \frac{||\boldsymbol{x}||_2^2}{n} = \mathbb{E}[x_i^2] \tag{5}$$

2. 矩阵的谱范数、Frobenius 范数与 RMS 范数

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

$$||\boldsymbol{A}\boldsymbol{x}||_2 \leq ||\boldsymbol{A}||_2||\boldsymbol{x}||_2, \tag{6}$$

$$||\boldsymbol{A}||_{\text{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}^2} \tag{7}$$

$$||\boldsymbol{A}||_{\text{RMS}} = \sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}^2} = \frac{||\boldsymbol{A}||_{\text{F}}}{\sqrt{mn}} \tag{8}$$

其中，$||\boldsymbol{A}||_2$ 表示矩阵的谱范数，$||\boldsymbol{A}||_{\mathrm{F}}$ 表示矩阵的 Frobenius 范数，$||\boldsymbol{A}||_{\mathrm{RMS}}$ 主要用来描述矩阵元素的平均尺度指标。当且仅当 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的右奇异向量，且对应于最大奇异值 $\sigma_{\max}$，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：

$$\frac{||\boldsymbol{A}||_{\mathrm{F}}}{\sqrt{\min(m,n)}} \le ||\boldsymbol{A}||_2 \le ||\boldsymbol{A}||_{\mathrm{F}} \le \operatorname{rank}(\boldsymbol{A})||\boldsymbol{A}||_2 \tag{9}$$

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 $\boldsymbol{Q}$，有：

$$\operatorname{rank}(\boldsymbol{Q}) = n, \tag{10}$$

$$\sigma_i(\boldsymbol{Q}) = |\lambda_i| = 1, \tag{11}$$

$$||\boldsymbol{Q}||_2 = 1, \tag{12}$$

$$||\boldsymbol{Q}||_{\mathrm{F}} = \sqrt{n} \tag{13}$$

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

$$\operatorname{rank}(\boldsymbol{\Lambda}) = \sum_{i=1}^{n} |\operatorname{sign}(\lambda_i)|, \tag{14}$$

$$||\boldsymbol{\Lambda}||_2 = \max_i(|\lambda_i|), \tag{15}$$

$$||\boldsymbol{\Lambda}||_{\mathrm{F}} = \sum_{i=1}^{n} \lambda_i^2 \tag{16}$$

对于外积矩阵，有：

$$\operatorname{rank}(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}) = 1 \Leftrightarrow ||\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}||_2 = ||\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}||_{\mathrm{F}} = ||\boldsymbol{u}||_2||\boldsymbol{v}||_2 \tag{17}$$

对于外积矩阵元素的符号化，有：

$$\begin{aligned} &\operatorname{sign}(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})_{i,j} = \operatorname{sign}(u_i v_j) = \operatorname{sign}(u_i)\operatorname{sign}(v_j) \\ \implies &\operatorname{sign}(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}) = \operatorname{sign}(\boldsymbol{u})\operatorname{sign}(\boldsymbol{v})^{\mathrm{T}} \end{aligned} \tag{18}$$

4. 矩阵分解为近似（半）正交矩阵与标量的数乘

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

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

那么有：

$$\operatorname{rank}(\boldsymbol{A}) = \min(m, n), \tag{20}$$

$$||\boldsymbol{A}||_2 = ||\kappa \boldsymbol{Q}||_2 = \kappa ||\boldsymbol{Q}||_2 \approx \kappa \times 1 = \kappa, \tag{21}$$

$$||\boldsymbol{A}||_{\mathrm{F}} = ||\kappa \boldsymbol{Q}||_{\mathrm{F}} = \kappa ||\boldsymbol{Q}||_{\mathrm{F}} \approx \kappa \times \sqrt{\min(m, n)} \tag{22}$$

$$||\boldsymbol{A}\boldsymbol{x}||_2 = ||\kappa \boldsymbol{Q}\boldsymbol{x}||_2 \approx \kappa ||\boldsymbol{x}||_2 \tag{23}$$

5. 矩阵是否为线性

在数学（特别是线性代数）中，一个运算 $f(.)$ 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 $f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$。

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

设有如下运算：

$$\boldsymbol{y} = \boldsymbol{A}(t)\boldsymbol{x} \tag{24}$$

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