计算机科学中的线性代数

2.1 基础

逆矩阵	正交矩阵	QR 分解

2.2 范数

2.3 向量范数

2.4 归纳矩阵范数

向量 p-范数	1-范数, 欧几里德（2）范数, 无穷（最大）范数

2.6 奇异值分解

2.9 Moore-Penrose 伪逆

2.10 张量

在这一节，我们将简要概述基本的线性代数属性和在这一章中将用到的数学符号。我们假定读者具备线性代数的基础（例如，向量的内积和叉积，基本矩阵运算如加法、标量乘法、转置、上/下三角矩阵，矩阵-向量乘法，矩阵乘法，矩阵的迹等）。

2.1 基础

我们将完全聚焦于线性空间中的矩阵和向量。我们使用符号 x ∈ R^n 表示 n 维向量，注意向量都是以粗体小写字母书写。这里**假定所有的向量都是列向量**，除非特别说明。所有元素为零的向量用 0 表示，所有元素为 1 的向量用 1 表示（类似 Broadcasting）；维度会隐含在上下文中或显式地用下标表示。

我们将使用粗体大写字母表示矩阵，例如 A ∈ R^mxn 表示一个 mxn 阶的矩阵；用 A_i* 表示 A 的第 i 行的行向量，用 A_*i 表示 A 的第 i 列的列向量。单位矩阵表示为 I_n，其中 n 是矩阵的行数和列数。最后，我们用 e_i 表示 I_n 的第 i 列，即第 i 个规范基。

逆矩阵：如果存在一个逆矩阵 A^-1 ∈ R^mxn 满足以下条件，那么矩阵 A ∈ R^mxn 被称为非奇异的或可逆的：

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_{n\times n} = \mathbf{A}^{-1}\mathbf{A}.$$

如果 A 的所有列向量（或行向量）线性无关，那么 A 是可逆的。换句话说，不存在一个非零向量 x ∈ R^n 使得 Ax=0。可逆矩阵的标准性质有：(A^−1)^T = (A^T)^−1 = A^−T（A 逆的转置等于 A 转置的逆）和 (AB)^−1 = B^−1* A^−1（A 左乘 B 的逆等于 B 逆左乘 A 逆。注：表达式展示不便，准确表达式请查看原材料）。

正交矩阵：如果矩阵 A ∈ R^n×n 满足 A^T=A^−1，则称 A 为正交矩阵。等价地说，对所有 i，j 属于 [1,n]，正交矩阵满足：

$$\mathbf{A}_{*i}^{\top}\mathbf{A}_{*j} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}.$$

对于 A 的行向量，上述性质同样满足。即 A 的所有列（或行）向量都是两两正交或互成法向量。

QR 分解：任意的矩阵 A ∈ R^n×n 都可以分解成一个正交矩阵和一个上三角矩阵的乘积：A=QR

其中 Q ∈ R^n×n 是正交矩阵，R ∈ R^n×n 是上三角矩阵。QR 分解在求解线性方程组的时候很有用，它的计算复杂度为 O(n^3)，并且是数值稳定的。为了用 QR 分解求解线性方程组 Ax=b，我们首先对等式两边同时左乘一个 Q^T，即 Q^TQRx = Rx = Q^Tb。然后，我们用反向代入求解 Rx = Q^Tb。

2.2 范数

范数（Norms）被用于度量矩阵的大小，或者相应地，度量向量的长度。范数是一个函数，它将 R^mxn（或 R^n）映射到 R。形式地说：

定义 1：任何函数满足 || · ||: R^m×n → R 和下列性质，则称为一个范数：

- 非负性：|| A ||≥0；|| A ||=0 当且仅当 A=0；

- 三角不等律：|| A+B ||≤|| A ||+|| B ||；
- 标量乘法律：|| αA ||=|α| || A ||，α∈R。

可以很容易地证明以下两个性质：

- || A ||=|| -A ||
- | || A ||-|| B || | ≤ || A-B ||

第二个性质被称为倒三角型不等式。

2.3 向量范数

若给定 n 维向量 x 和一个整数 p > 1，我们可以定义向量 p-范数为：

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

最常见的向量 p-范数为：

- 1-范数：
  $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|.$
- 欧几里德（2）范数：
  $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{\mathbf{x}^\top \mathbf{x}}.$
- 无穷（最大）范数：
  $\|\alpha \mathbf{A}\| = |\alpha|\|\mathbf{A}\|, \text{ for all } \alpha \in \mathbb{R}.$

若给定 n 维向量 x、y，我们可以使用 p-范数作为内积的上确界，即 Cauchy-Schwartz 不等式可以写为：

$$|x^\top y| \leqslant \|x\|_2\|y\|_2.$$

一般来说，该不等式给定了两个向量的欧几里德范数可以作为它们内积的上确界，Holder 不等式表明：

$$|x^\top y| \leqslant \|x\|_2\|y\|_2.$$

以下向量 p-范数的不等式性质可以轻易的证明：

$$\begin{aligned}\|x\|_\infty \leqslant \|x\|_1 \leqslant n\|x\|_\infty,\\ \|x\|_2 \leqslant \|x\|_1 \leqslant \sqrt{n}\|x\|_2,\\ \|x\|_\infty \leqslant \|x\|_2 \leqslant \sqrt{n}\|x\|_\infty.\end{aligned}$$

2.4 归纳矩阵范数

给定一个 m×n 阶矩阵 A，和一个 p > 1 整数，我们定义矩阵的 p-范数为：

$$\|\mathbf{A}\|_p = \max_{x\neq 0}\frac{\|\mathbf{A}x\|_p}{\|x\|_p} = \max_{\|x\|_p=1}\|\mathbf{A}x\|_p.$$

一般我们最常用的矩阵 p-范数为：

- 1-范数，取矩阵列加和绝对值的最大值：

$$\|\mathbf{A}\|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|\mathbf{A}_{ij}| = \max_{1\leqslant j\leqslant n}\|\mathbf{A}e_j\|_1.$$

- 无穷范数，取矩阵行加和绝对值的最大值：

$$\|\mathbf{A}\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |\mathbf{A}_{ij}| = \max_{1\leqslant j\leqslant m} \|\mathbf{A}^\top \mathbf{e}_i\|_1.$$

- 2-范数，

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2=1} \sqrt{\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x}}.$$

这一系列的范数被称为「归纳（induced）」，因为它们是通过不取决于 A 和 p 的非零向量 x 而实现的。因此，一般存在一个单位范数向量（p-范数中的单位范数）x 令||A||p = ||Ax||p。归纳矩阵 p-范数遵循以下 submultiplicativity 法则：

$$\|\mathbf{A}\mathbf{x}\|_p \leqslant \|\mathbf{A}\|_p \|\mathbf{x}\|_p \quad \text{and} \quad \|\mathbf{A}\mathbf{B}\|_p \leqslant \|\mathbf{A}\|_p \|\mathbf{B}\|_p.$$

此外，矩阵 p-范数对于矩阵的初等变换是不变的，即||PAQ||p = ||A||p，其中 P 和 Q 为对应维度的初等变换矩阵。同样，如果我们考虑矩阵分割：

$$\mathbf{PAQ} = \begin{pmatrix} \mathbf{B} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

那么子矩阵的范数就和全部矩阵的范数相关：即||B||p <= ||A||p。矩阵 p-范数间的以下关系可以相对简单地证明。若给定一个 m×n 阶矩阵，

$$\frac{1}{\sqrt{n}}\|\mathbf{A}\|_\infty \leqslant \|\mathbf{A}\|_2 \leqslant \sqrt{m}\|\mathbf{A}\|_\infty,$$
$$\frac{1}{\sqrt{m}}\|\mathbf{A}\|_1 \leqslant \|\mathbf{A}\|_2 \leqslant \sqrt{n}\|\mathbf{A}\|_1.$$

此外，||A^T||1 = ||A||∞，||A^T||∞ = ||A||1。其中转置影响了矩阵的无穷范数和 1-范数，而不影

响 2-范数，即||A^T||2 = ||A||2。同样，矩阵 2-范数并不会受到矩阵 pre(post)- multiplication 操作的影响，其中它的列（或行）为正交向量：||UAV^T||2 = ||A||2，其中 U 和 V 为对应维度的正交矩阵（U^T*U = I and V^T*V = I）。

2.6 奇异值分解

我们知道方阵可以分解为特征值与特征向量，但非方阵的矩阵并没不能实现特征值分解。因此奇异值分解（SVD）是每个矩阵中最重要的矩阵分解方式，因为不是所有的矩阵都能进行特征分解，但是所有的矩阵都能进行奇异值分解。

定义 6. 给定一个矩阵 A ∈ R^m×n，我们定义全 SVD 为：

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \sum_{i=1}^{\min\{m,n\}} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top,$$

其中 U ∈ R^m×m 和 V ∈ R^n×n 分别是包含 A 的左、右奇异向量的正交矩阵，Σ ∈ R^m×n 是对角矩阵，其中 A 的奇异值在主对角线上递减。

我们经常使用 u_i（或 v_j），i=1,..., m（或 j=1,..., n）来表示矩阵 U（或 V）的列。同样，我们将使用σ_i，i = 1,..., min{m, n} 来表示奇异值：

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{\min\{m,n\}} \geqslant 0.$$

A 的奇异值是非负的，其数目等于 min{m, n}。A 的非零奇异值个数等于 A 的秩。由于正交不变性，我们得到：

$$\boldsymbol{\Sigma}_{\mathbf{PAQ}^\top} = \boldsymbol{\Sigma}_{\mathbf{A}},$$

其中 P 和 Q 是对应维度上的正交矩阵（P^TP = I 且 Q^TQ = I）。或者说，PAQ 的奇异值与 A 的奇异值相同。

涉及矩阵 A 和 B 的奇异值的以下不等式是非常重要的。首先，如果 A 和 B 都在 R^m×n 上，对于所有 i = 1, ... , min{m, n},

$$|\sigma_i(\mathbf{A}) - \sigma_i(\mathbf{B})| \leqslant \|\mathbf{A} - \mathbf{B}\|_2.$$

第二，如果 A ∈ R^p×m 和 B ∈ R^m×n，对于所有 i = 1, ... , min{m, n},

$$\sigma_i(\mathbf{AB}) \leqslant \sigma_1(\mathbf{A})\sigma_i(\mathbf{B}),$$

其中σ_1(A) = ||A||_2。我们经常对于仅保持非零奇异值和相应的（矩阵 A 的）左、右奇异向量感兴趣。给定矩阵 A ∈ R^m×n 和 rank(A)=ρ，我们可以定义它的稀疏 SVD。

定义 9. 给定矩阵 A ∈ R^m×n，秩为ρ ≤ min{m, n}，我们定义稀疏 SVD 为：

$$\mathbf{A} = \underbrace{\mathbf{U}}_{m\times\rho}\underbrace{\mathbf{\Sigma}}_{\rho\times\rho}\underbrace{\mathbf{V}^\top}_{\rho\times n} = \sum_{i=1}^{\rho}\sigma_i \mathbf{u}_i \mathbf{v}_i^\top,$$

其中 U ∈ R^m×ρ和 V ∈ R^n×ρ是包含对应于非零奇异值的左、右奇异向量的两两正交列（即 U^TU = I 且 V^TV = I）的矩阵；Σ ∈ R^ρ×ρ是 A 的非零奇异值在对角线上递减的对角矩阵。

如果 A 是非奇异矩阵，我们可以使用 SVD 计算它的逆：

$$\mathbf{A}^{-1} = (\mathbf{U\Sigma V}^\top)^{-1} = \mathbf{V\Sigma}^{-1}\mathbf{U}^\top.$$

（如果 A 是非奇异的，那么它是方形和满秩的，在这种情况下，稀疏 SVD 和全 SVD 是一样的）众所周知，SVD 非常重要，任何矩阵的最佳 k 秩近似都可以通过 SVD 来计算。

定理 10. 让 A = UΣV^⊤ ∈ R^m×n 作为 A 的稀疏 SVD；设 k < rank(A) = ρ为整数，让

$$\sigma_{k+1} = \min_{\mathbf{B}\in\mathbb{R}^{m\times n},\ \mathrm{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_2 = \|\mathbf{A}-\mathbf{A}_k\|_2$$

随后，

$$\sigma_{k+1} = \min_{\mathbf{B}\in\mathbb{R}^{m\times n},\ \mathrm{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_2 = \|\mathbf{A}-\mathbf{A}_k\|_2$$

and

$$\sum_{j=k+1}^{\rho} \sigma_j^2 = \min_{\mathbf{B}\in\mathbb{R}^{m\times n},\ \mathrm{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_F^2 = \|\mathbf{A}-\mathbf{A}_k\|_F^2.$$

换句话说，上述定理指出，如果我们寻找一个矩阵 A 的 k 秩近似，使得「误差」矩阵的 2-范数或 Frobenius 范数最小化（即 A 和它的近似之间的差异最小化），随后我们需要保留 A 的最前 k 个奇异值和相应的左、右奇异向量。

我们会经常使用这些符号：让 U_k ∈ R^m×k（或 V_k ∈ R^n×k）表示矩阵 A 的最前 k 个左（或右）奇异向量的矩阵；让 Σ_k ∈ R^k×k 表示包含 A 的最前 k 个奇异值的对角矩阵。同样的，让 U_k,⊥ ∈ R^m×(ρ−k)（或 V_k,⊥ ∈ R^n×(ρ−k)）表示 A 的底部ρ-k 个非零左（或右）奇异向量的矩阵；然后令Σ_k,⊥ ∈ R^(ρ−k)×(ρ−k) 表示包含 A 的底部ρ-k 个奇异值的对角矩阵。然后，

$$\mathbf{A}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T \quad \text{and} \quad \mathbf{A}_{k,\perp} = \mathbf{A}-\mathbf{A}_k = \mathbf{U}_{k,\perp}\mathbf{\Sigma}_{k,\perp}\mathbf{V}_{k,\perp}^T.$$

2.9 Moore-Penrose 伪逆

对于非方矩阵而言，其逆矩阵是没有定义的。而一种非常出名的推广型矩阵求逆方法 Moore-Penrose 伪逆在这类问题上取得了一定的进展。形式上来说，若给定 m×n 阶矩阵 A，那么如果矩阵 A† 满足以下属性，它就是矩阵 A 的 Moore-Penrose 伪逆：

1. $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.
2. $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$.
3. $(\mathbf{A}\mathbf{A}^\dagger)^\top = \mathbf{A}\mathbf{A}^\dagger$.
4. $(\mathbf{A}^\dagger\mathbf{A})^\top = \mathbf{A}^\dagger\mathbf{A}$.

给定一个秩为ρ的 m×n 阶矩阵 A，它的稀疏奇异值分解可以表示为：

$$\mathbf{A} = \sum_{i=1}^{\rho} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top,$$

它的 Moore-Penrose 伪逆 A† 的稀疏奇异值分解可以表示为：

$$\mathbf{A}^\dagger = \sum_{i=1}^{\rho} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^\top.$$

如果 A 为 n×n 阶满秩矩阵，那么 A† 就等于矩阵 A 的逆。如果 A 为 m×n 阶列满秩矩阵，那么 A†A 就等于 n 阶单位矩阵，AA†为矩阵 A 列上的投影矩阵。如果 A 为满行秩矩阵，那么 AA†就为 m 阶单位矩阵，A†A 为矩阵 A 行上的投影矩阵。

关于两个矩阵乘积的伪逆，有如下特别重要的属性：对于 m×p 阶矩阵 Y1 和 p×n 阶矩阵 Y2，且满足 Rank(Y1)=Rank(Y2)，即秩相等，[9, Theorem 2.2.3] 表明：

$$(\mathbf{Y}_1\mathbf{Y}_2)^\dagger = \mathbf{Y}_2^\dagger\mathbf{Y}_1^\dagger.$$

（我们强调秩相等的条件是非常重要的：因为两个矩阵相乘的逆总是等价于矩阵逆的相乘，但这个推断对于一般的 Moore-Penrose 伪逆 [9] 是不满足的）此外，Moore-Penrose 伪逆的基空间和所有实际的矩阵都有联系。给定一个矩阵 A 和 A 的 Moore-Penrose 伪逆 A†，A†的列空间可以定义为：

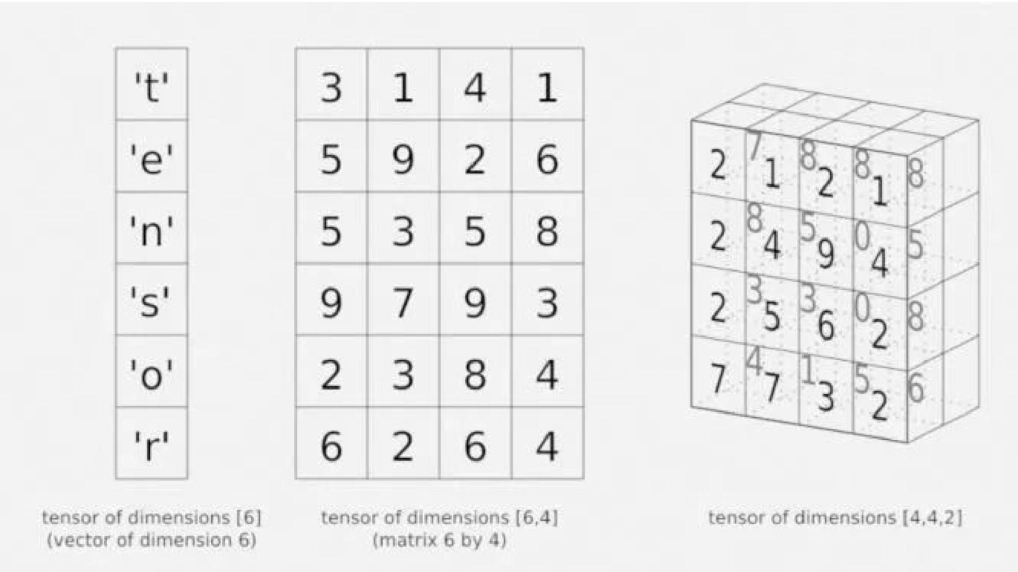
$$\text{range}(\mathbf{A}^\dagger) = \text{range}(\mathbf{A}^\top\mathbf{A}) = \text{range}(\mathbf{A}^\top),$$

A†的列空间和零空间（null space）正交，A†的零空间可以定义为：

$$\text{null}(\mathbf{A}^\dagger) = \text{null}(\mathbf{A}\mathbf{A}^\top) = \text{null}(\mathbf{A}^\top),$$

2.10 张量

三维张量是按照一定规律排列在方格中的数组，其中一个变量数字表示轴。张量有三个索引，其中第一个索引表示行，第二个索引表示列，第三个索引表示轴。例如，V_232 指向第二行、第三列、第二轴的元素，在下图右边的张量中表示 5。



张量是上面谈到的概念中最常用的一个，因为张量是一个多维数组，同时可以是一个向量或者一个矩阵，具体取决于它的索引数量。例如，一阶张量可以表示向量（1 个索引），二阶张量可以表示矩阵（2 个索引），三阶就是张量（3 个索引），更高阶的称为高阶张量（超过 3 个索引）。

《Lectures on Randomized Numerical Linear Algebra》：
https://arxiv.org/pdf/1712.08880.pdf