StudyThinking

卷三 数据学习笔记

作者：latalealice

日期：2025/04/24

目　录

序言

Notes on “LINEAR ALGEBRA AND LEARNING FROM DATA” by GILBERT STRANG 📕.

Thank you for reading this notes ❤️,

I hope you like it 😄.

第一章 线性代数的要点

本章要学习的五个基本问题:

$$Ax = b, Ax = \lambda x, Av = \sigma u, \text{Minimize } \frac{|Ax|^2}{\|x\|^2}$$

这些问题表面是常规计算任务(如“求 x”“分解 A=列 $\times$ 行”),但我们的目标远不止求解,而是理解其本质.例如:

- $Ax = b$是否有解?关键在于“向量 b是否位于 A的列空间?”
- 深入探索“空间”这一看似简单的概念,并使其高效解决问题的关键工具.

特征值与奇异值的深层差异:

- 特征方程 $Ax = \lambda x$:无需外部向量 b,仅研究矩阵 A自身的性质.**特征向量 x的方向经 A变换后保持不变**,此时复杂的矩阵关系简化为标量运算(如 $A^2x = \lambda^2 x$,微分方程中的 $e^{At}x = e^{it}x$).掌握所有 x与 λ,即可解决一切线性问题.
- 奇异值方程 $Av = \sigma u$:涉及两个向量 v和 u,通常对应矩形数据矩阵 A.奇异值分解将其拆解为基本单元 σuv^T(列向量 u与行向量 v^T 的乘积),揭示数据矩阵的核心结构.

最小化与分解的应用意义 通过优化(如最小二乘法求最佳 x)和分解(如 PCA 提取主成分 v_1),构建了数据拟合的代数框架.

掌握核心概念的广阔应用,一旦理解列空间、零空间、特征向量与奇异向量,便能驾驭诸多领域:

- 最小二乘法
- 傅里叶变换
- 统计学中的 LASSO
- 神经网络中的随机梯度下降

1.1 基于矩阵列向量的 Ax乘法运算

开篇将探讨矩阵与向量的乘法 Ax、矩阵的列空间以及秩的概念. 分别采用矩阵 A的三行和两列进行 Ax乘法运算:

$$\text{采用行向量与 } x \text{ 的内积: } \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1+3x_2 \\ 2x_1+4x_2 \\ 3x_1+7x_2 \end{bmatrix}$$

$$\text{采用列向量的线性组合: } \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$$

两种方法结果一致.

第一种(逐行计算)得到三个内积,由于点号表示法,这些结果也被称为“点积”.

$$\textbf{row} \cdot \textbf{column} = \boldsymbol{(2,3) \cdot (x_1, x_2) = 2x_1 + 3x_2}$$

计算 Ax三个分量的方法只是底层运算方式,而非对原理的理解,理解原理要更高维度的向量视角.

向量视角将 Ax视为 a_1和 a_2的“线性组合”,这正是线性代数的核心运算.构建 a_1和 a_2的线性组合包含两个步骤:

1. 用“标量” x_1和 x_2分别乘列向量 a_1和 a_2
2. 将向量相加得到 $x_1a_1 + x_2a_2 = Ax$

因此, Ax本质上是 A的列向量的线性组合——这是最根本的认知.

这种思维将人引向 A的列空间.关键是要考虑列向量的所有可能组合.允许 x_1和 x_2取任意实数时,该空间就包含了所有可能向量 x对应的 Ax.由此我们得到无限多个输出向量 Ax,这些输出结果可以从几何角度直观呈现.

在此示例中,每个 Ax 都是三维空间 $\mathbb{R}^3$ 中的一个向量.**所有组合 $Ax = x_1a_1 + x_2a_2$ 会构成完整三维空间中的一个平面.**该平面包含 $a_1 = (2,2,3)$ 方向的完整直线,因为所有 x_1a_1 向量都被包含在内.平面还包含 a_2 方向的所有 x_2a_2 向量直线.同时,它也包含来自一条直线上任一向量与另一条直线上任一向量相加的和.这种加法运算填充出了一个包含这两条直线的无限延伸平面,但并未填满整个三维空间 $\mathbb{R}^3$.

定义:矩阵列向量的所有线性组合构成了 A 的列空间(column space).

在本例中,列空间是一个平面.该平面包含:

- 原点 $(0,0,0)$(当 $x_1 = x_2 = 0$ 时生成)
- 向量 $(5,6,10) = a_1 + a_2$
- 向量 $(-1,-2,-4) = a_1 - a_2$

所有形如 $x_1a_1 + x_2a_2$ 的组合都属于这个列空间.也许可以百分百地确定 rand(3,1)生成的向量肯定不在这个列空间内:

当且仅当方程组 $Ax = b$ 有解 (x_1, x_2),向量 $b = (b_1, b_2, b_3)$ 属于 A 的列空间.

这一数学本质揭示了列空间 $\mathbf{C}(A)$ 的核心特征:解 x 指明了如何将 b 表示为列向量的线性组合 $x_1a_1 + x_2a_2$.对于不属于该列空间的 b,这种线性表出是不可能的.

$$A_2 = \begin{bmatrix} 2 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 7 & 10 \end{bmatrix}, A_3 = \begin{bmatrix} 2 & 3 & 1 \\ 2 & 4 & 1 \\ 3 & 7 & 1 \end{bmatrix}$$

矩阵 A_2 的列空间与原先保持同一平面.新增的第三列 $(5,6,10)$ 是第一列与第二列之和,因此该列向量 a_3 已存在于原平面中,并未拓展新的空间维度.引入这个"线性相关"的列向量,列空间依然被限制在原始的二维平面内.

矩阵 A_3 的列空间是整个三维空间 $\mathbb{R}^3$.新增的第三列(1,1,1)不在原列空间 $\mathbf{C}(A)$ 的平面内,使得列空间 $\mathbf{C}(A_3)$ 得以扩展. xy 平面与一个穿出该平面的向量 $(x_3, y_3, z_3)(z_3 \neq 0)$,它们的线性组合能生成 $\mathbb{R}^3$ 中的所有向量.

$\mathbb{R}^3$ 中所有可能的列空间(维度 0 至 3):

- 零子空间:仅含零向量(0,0,0)
- 直线:所有 x_1a_1 向量的集合
- 平面:所有 $x_1a_1 + x_2a_2$ 向量的集合
- 全空间 $\mathbb{R}^3$:所有 $x_1a_1 + x_2a_2 + x_3a_3$ 向量的集合

关键约束条件:向量 a_1, a_2, a_3 必须**线性无关**(唯一能生成零向量的组合 $0a_1 + 0a_2 + 0a_3$).因此:

- 单个向量 a_1 张成直线
- a_1 与 a_2 张成平面
- a_1, a_2, a_3 可生成 $\mathbb{R}^3$ 中任意向量 b

(注:零向量是所有子空间的共有元素)

线性代数标准表述:

1. $\mathbb{R}^3$ 中三个线性无关列向量构成可逆矩阵: $AA^{-1} = A^{-1}A = I$
2. 方程 $Ax = 0$ 仅有零解 $x = (0,0,0)$,此时 $Ax = b$ 存在唯一解 $x = A^{-1}b$

构建列空间的关键在于: 从 A 的 n 个列向量中筛选出最大线性无关组构成矩阵 C,实现 $A = CR$ 的因式分解(C 包含原始列向量, R 记录组合关系).由此证明线性代数第一基本定理,同时揭示矩阵的秩与子空间维度的本质联系.

$\boxed{\textbf{对于 } m\times n \textbf{ 矩阵 } A \textbf{ 有(对应线性映射 } T:\mathbb{R}^n\to\mathbb{R}^m\textbf{):nullity(}A\textbf{)+rank(}A\textbf{)=}n}$

基础构造算法具体步骤:

1. 若 A 第 1 列非零向量,纳入 C
2. 若 A 第 2 列不是第 1 列的倍数,纳入 C
3. 若 A 第 3 列不能表示为前两列的线性组合,纳入 C

$\vdots$

最终得到的 C 将包含 r 个列向量($r\le n$),这些列构成 A 列空间的一组基.被剔除的列均可表示为 C 中基向量的线性组合.

子空间的基是一组线性无关的向量:空间中的所有向量都可表示为该基向量的线性组合.

数值 r 称为矩阵 A 的“秩”,同时也是矩阵 C 的秩.它表示线性无关列的数量.

矩阵的秩等于其列空间的维数.

矩阵 C 通过第三个矩阵 R 与 A 相关联: $A=CR$.它们的维度关系为: $(m\times n)=(m\times r)(r\times n)$

$$A=\begin{bmatrix}1&3&8\\1&2&6\\0&1&2\end{bmatrix}=\begin{bmatrix}1&3\\1&2\\0&1\end{bmatrix}\begin{bmatrix}1&0&2\\0&1&2\end{bmatrix}=CR$$

当 C 乘以 R 的第一列时,将生成 C 和 A 的第一列;

当 C 乘以 R 的第二列时,将生成 C 和 A 的第二列;

当 C 乘以 R 的第三列时,将得到 2 倍第一列加 2 倍第二列的组合.

只需在 R 中填入正确的系数,通过矩阵 C 各列的线性组合就能重构出 A 的所有列.

$\boxed{R=\text{rref}(A)=\boldsymbol{A}\textbf{“的行最简阶梯形”}}$

矩阵中线性无关列的数量等于线性无关行的数量.

这个秩定理适用于所有矩阵.在线性代数中,列与行总是相伴相生.虽然 m 行包含与 n 列相同的数值 a_{ij},但构成的是不同的向量.

该定理通过 $A=CR$ 分解得到证明.若从行视角而非列视角观察:矩阵 R 具有 r 个行向量,通过与 C 相乘可得到这些行向量的线性组合.由于 $A=CR$, A 的所有行向量均可由 R 的 r 个行向量生成.且这 r 个行向量线性无关,故构成 A 的行空间基.

因此,矩阵 A 的列空间与行空间具有相同的维度 r——其基向量分别为 C 的 r 个列向量和 R 的 r 个行向量. $(m\times n)=(m\times r)(r\times n)$

1.2 矩阵乘法 AB 的列空间视角

矩阵乘法 $AB=C$ 中的每个元素,都是由行向量与列向量的内积计算得出,矩阵 A 的第 i 行与矩阵 B 的第 j 列相乘,得到矩阵 C 中的元素 c_{ij}:

$$\begin{bmatrix}\cdot&\cdot&\cdot\\a_{21}&a_{22}&a_{23}\\\cdot&\cdot&\cdot\end{bmatrix}\begin{bmatrix}\cdot&\cdot&b_{13}\\\cdot&\cdot&b_{23}\\\cdot&\cdot&b_{33}\end{bmatrix}=\begin{bmatrix}\cdot&\cdot&\cdot\\\cdot&\cdot&c_{23}\\\cdot&\cdot&\cdot\end{bmatrix}$$

$$c_{ij}=\textstyle\sum_{k=1}^{n}a_{ik}b_{kj}$$

矩阵乘法 AB 的另一种计算方式是: A 的列向量乘以 B 的行向量.

一个 $m\times 1$ 矩阵(列向量 u)乘以 $1\times p$ 矩阵(行向量 v^T),将得到一个 $m\times p$ 矩阵:

$$uv^T=\begin{bmatrix}2\\2\\1\end{bmatrix}\begin{bmatrix}3&4&6\end{bmatrix}=\begin{bmatrix}6&8&12\\6&8&12\\3&4&6\end{bmatrix}$$

矩阵 uv^T 的列空间是一维的:即沿着向量 u 方向的直线.列空间的维度(线性无关列的数量)就是矩阵的秩.所有非零矩阵 uv^T 的秩都为 1

同时:矩阵 uv^T 的行空间是通过向量 v 的直线.矩阵 A 的行空间就是其转置矩阵 A^T 的列空间 $\mathbf{C}\left(A^T\right)$.将 uv^T 转置(行列互换)得到矩阵 vu^T:

$$\left(uv^T\right)^T=\begin{bmatrix}6&6&3\\6&8&4\\12&12&6\end{bmatrix}=\begin{bmatrix}3\\4\\6\end{bmatrix}\begin{bmatrix}2&2&1\end{bmatrix}=vu^T$$

此为线性代数第一基本定理最清晰的示例:

行秩=列秩

r 个线性无关列 $\Leftrightarrow$ r 个线性无关行

矩阵 AB 可表示为所有秩 1 矩阵的和:

$$\boxed{AB=\sum uv^T}$$

采用 A 的列向量与 B 的行向量相乘的方法,若 $a_1,a_2,...,a_n$ 为矩阵 A 的 n 个列向量,则矩阵 B 必须有对应的 n 个行向量 $b_1^*,b_2^*,...,b_n^*$,乘积 AB 可表示为所有列向量 a_k 与行向量 b_k^* 的外积之和:

$$\boxed{AB=\begin{bmatrix}|&&|\\a_1&\cdots&a_n\\|&&|\end{bmatrix}\begin{bmatrix}-&b_1^*&-\\&\vdots&\\-&b_n*&-\end{bmatrix}=\sum a_kb_k^*}$$

这是一个 2×2 的示例,展示 $n=2$ 个部分(列乘以行)以及它们的和 AB.

$$\begin{bmatrix}1&0\\3&1\end{bmatrix}\begin{bmatrix}2&4\\0&5\end{bmatrix}=\begin{bmatrix}2&4\\6&12\end{bmatrix}+\begin{bmatrix}0&0\\0&5\end{bmatrix}=\begin{bmatrix}2&4\\6&17\end{bmatrix}$$

得到 2,4,6,12 需要 4 次乘法.再得到 0,0,0,5 又需要 4 次乘法.总共是 2^3=8 次乘法.当 A 和 B 都是 $n\times n$ 矩阵时,总是需要进行 n^3 次乘法.而当 $AB=(m\times n)$ 矩阵乘以 $(n\times p)$ 矩阵时,需要进行 mnp 次乘法: n 个秩为 1 的矩阵,每个矩阵的大小都是 $m\times p$.

行乘以列共需 mnp 次内积运算,每次运算包含 n 次乘法,总计 mnp 次乘法.

列乘以行共需 n 次外积运算,每次运算包含 mp 次乘法,总计 mnp 次乘法.

它们实际上是相同的乘法运算 $a_{ik}b_{kj}$,只是顺序不同.外积方法在数据科学中至关重要,简而言之:寻求矩阵 A 的最主要成分——正是秩为 1 的矩阵 uv^T.应用线性代数的一个核心主题:

将矩阵 A 分解为 CR 形式,并考察 $A=CR$ 中的各个组成部分 $c_kr_k^*$

将矩阵 A 分解为 CR 是矩阵相乘 $CR=A$ 的逆过程,分解过程耗时更长,特别是当分解涉及特征值或奇异值时,这些数值包含矩阵 A 的内部信息,在分解前是不可见的.以下是五种重要分解形式:

$$\boxed{A=LU,A=QR,,,A=U\Sigma V^T}$$

以下是这些分解方法的详解：

1. $A=LU$ 来自消元法.通过行变换将 A 转换为 U,并可从 U 还原回 A. L 是下三角矩阵, U 是上三角矩阵.
2. $A=QR$ 来自列向量 a_1 至 a_n 的正交化过程("Gram-Schmidt"方法). Q 具有标准正交列 $(Q^TQ=I)$, R 是上三角矩阵.
3. $S=QAQ^T$ 来自对称矩阵 $S=S^T$ 的特征值 $\lambda_1,...,\lambda_n$.特征值位于 Λ 的对角线上,标准正交特征向量构成 Q 的列.
4. $A=X\Lambda X^{-1}$ 是 $n\times n$ 矩阵 A(具有 n 个线性无关特征向量)的对角化. A 的特征值位于 Λ 对角线,特征向量构成 X 的列.
5. $A=U\Sigma V^T$ 是任意矩阵 A(无论方阵与否)的奇异值分解.奇异值 $\sigma_1,...,\sigma_r$ 位于 Σ 中,标准正交奇异向量构成 U 和 V 的列.

这里以最经典的第 3 种分解来距离,这种特殊的分解 QAQ^T 从对称矩阵 S 出发.该矩阵具有标准正交的特征向量 $q_1, ..., q_n$.这些互相垂直的特征向量(点积=0)构成 Q 的列向量. S 和 Q 堪称线性代数中的“OK 组合”:

$$\boxed{\begin{gathered} S = S^T, s_{ij} = s_{ji} \\ Q^T = Q^{-1}, q_i \cdot q_j = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ else} \end{cases} \end{gathered}}$$

对角矩阵 Λ 包含实数特征值 λ_1 至 λ_n.每个实对称矩阵 S 都有 n 个标准正交特征向量 q_1 至 q_n.当这些特征向量与 S 相乘时,其方向保持不变,仅按特征值 λ 进行缩放:

$$\boxed{Sq = \lambda q}$$

$$\boxed{SQ = S\begin{bmatrix} | & & | \\ q_1 & ... & q_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1 q_1 & ... & \lambda_n q_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ q_1 & ... & q_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = Q\Lambda}$$

将 $SQ = Q\Lambda$ 乘以 $Q^{-1} = Q^T$ 可得 $S = Q\lambda^T$,这就是对称矩阵的分解形式.每个特征值 λ_k 和对应的特征向量 q_k 共同构成了 S 的一个秩为 1 的分量 $\lambda_k q_k q_k^T$.

1.3 四大基本子空间

1.4 消元法与 $A = LU$ 分解

1.5 正交矩阵与子空间

1.6 特征值与特征向量

1.7 对称正定矩阵

1.8 SVD 中的奇异值与奇异向量

1.9 主成分分析与最佳低秩矩阵

1.10 瑞利商与广义特征值

1.11 向量/函数/矩阵的范数

1.12 矩阵与张量分解:正定与稀疏