线性代数
与
数据学习

latale alice

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

March 28, 2025

Thank you for reading this notes ❤️,
I hope you like it 😄

目录

1. 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×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:

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

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

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”.

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

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$\boldsymbol{Ax = x_1 a_1 + x_2 a_2}$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$x_1 a_1$向量都被包含在内.平面还包含$a_2$方向的所有$x_2 a_2$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$\mathbb{R}^3$.

$$\boxed{\textbf{定义}:\textbf{矩阵列向量的所有线性组合构成了}\boldsymbol{A}\textbf{的列空间}\text{(column space).}}$$

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$x_1 a_1 + x_2 a_2$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:

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

这一数学本质揭示了列空间 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$x_1 a_1 + x_2 a_2$.对于不属于该列空间的b,这种线性表出是不可能的.

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

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 C(A)的平面内,使得列空间 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$x_1 a_1$向量的集合
- 平面:所有$x_1 a_1 + x_2 a_2$向量的集合
- 全空间$\mathbb{R}^3$:所有$x_1 a_1 + x_2 a_2 + x_3 a_3$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.

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

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

⋮

最终得到的C将包含r个列向量($r\leq n$),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.

$$\boxed{\textbf{矩阵的秩等于其列空间的维数}.}$$

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:$(m\times n)=(m\times r)(r\times n)$

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

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.

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

$$\boxed{\textbf{矩阵中线性无关列的数量等于线性无关行的数量}.}$$

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}:

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

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

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

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