3.5 奇异值分解

我们学过了许多分解，但是我想要告诉大家的是，奇异值分解是线性代数最好也是我们最后一个了解的分解形式！

奇异值分解

矩阵的奇异值分解通常简称为 **SVD**。这是对矩阵而言“最终且最佳”的分解方式：

$A = U\Sigma V^T$（当然本来应该是$A = U\Sigma V^{-1}$，由于是正交矩阵所以可以写成转置形式后文有说这个）

其中：

- U 是正交矩阵，
- Σ 是对角矩阵，
- V 也是正交矩阵。

在这种分解中，矩阵 A 可以是任意形式的矩阵。我们知道，如果 A 是对称正定矩阵，那么它的特征向量是正交的，我们可以将其写成 $A = Q\Lambda Q^T$。这是 SVD 的一个特例，此时 $U = V = Q$。对于更一般的矩阵 A，SVD 需要两个不同的矩阵 U 和 V。

我们还学习过另一种分解形式：$A = S\Lambda S^{-1}$，其中 S 是由 A 的 n 个线性无关特征向量组成的矩阵。然而，这里的 S 不一定是正交的；而 SVD 中的 U 和 V 则一定是正交矩阵。

工作原理（几何直观）

我们可以将矩阵 A 看作一个线性变换：将行空间中的向量 v_1 变换为列空间中的向量 $u_1 = Av_1$。SVD 的核心思想就是：**找到一个行空间中的正交基，使得它在 A 的作用下变换为列空间中的另一个正交基**：

$$Av_i = \sigma_i u_i$$

找到一个行空间中的正交基并不难——Gram-Schmidt 正交化过程可以立刻给出。但一般而言，我们并不能保证矩阵 A 会将这个正交基变换为另一个正交基。

你可能会问：那矩阵 A 及其转置 A^T 的零空间（nullspaces）怎么办呢？——没关系，对角矩阵 Σ 中的零元素会自动处理这些零空间中的向量。

矩阵语言描述（代数形式）

问题的核心在于：为矩阵 A 的行空间找到一个标准正交基 $v_1, v_2, \ldots, v_r$，使得：

$$A\begin{bmatrix} v_1 & v_2 & \dots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

其中，$u_1, u_2, \dots, u_r$ 是矩阵 A 的列空间的标准正交基。一旦我们加入了零空间中的向量，这个方程就变成了：

$$AV = U\Sigma$$

我们可以将 $v_1, \dots, v_r$ 和 $u_1, \dots, u_r$ 分别扩展为整个空间的标准正交基。由于 $v_{r+1}, \dots, v_n$ 位于 A 的零空间中，因此对应的 $\sigma_{r+1}, \dots, \sigma_n$ 必然为零。
也就是说：

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

但是当A是方阵，那么扩充后Σ的形式就是如上所示！但是如果A是矩形矩阵，那么Σ就不是方阵了，而是存在一个斜对角线的！

矩阵 U 和 V 的列分别构成了矩阵 A 的列空间和行空间的基。一般而言，$U \neq V$，但如果矩阵 A 是正定的，我们就可以对行空间和列空间使用相同的基！
然而只要是列满秩矩阵或者可逆矩阵，那么他们的奇异值就不会出现0

接下来我们介绍如何去计算SVD

计算方法（如何实际求 SVD）

假设矩阵 A 是一个可逆的 2×2 矩阵：

$$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$$

我们希望找到：

- 行空间 $\mathbb{R}^2$ 中的向量 v_1, v_2，
- 列空间 $\mathbb{R}^2$ 中的向量 u_1, u_2，
- 正数 σ_1, σ_2，

使得：

- v_i 标准正交，

- u_i 标准正交，
- 满足 $Av_i = \sigma_i u_i$。

这是迈向求出正交矩阵 V, U 和对角矩阵 Σ 的关键一步：

$$AV = U\Sigma$$

由于 V 是正交矩阵，我们可以两边同时乘以 $V^{-1} = V^T$，得到：

$$A = U\Sigma V^T$$

我们不直接同时求解 U, V, Σ，而是采用以下步骤：

- 两边转置后乘以 A：

$$A^T A = V\Sigma^T U^T U\Sigma V^T = V\Sigma^2 V^T$$

- 注意到 $A^T A$ 是对称半正定矩阵（一个矩阵的转置乘以他本身，结果至少是一个半正定矩阵，这取决于A是否满秩！），因此它可以被对角化为 $Q\Lambda Q^T$ 的形式。我们可以利用这一点求出 V：
  - V 的列是 $A^T A$ 的特征向量，
  - 特征值就是 σ_i^2（取正平方根）。
- 类似地，为了求出 U，我们考虑矩阵 AA^T。这样就可以消去V！
  是不是感觉很妙，当然这部分的推理大家可能云里雾里，但是无所谓知道他是一个线性变换的过程以及清楚怎么计算即可！

一个具体的 SVD 示例

继续以上矩阵：

$$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$$

我们先计算：

$$A^T A = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$$

求出特征向量和特征值后，我们得到：

- 特征向量（标准正交化后）：

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

- 特征值：

$$\sigma_1^2 = 32, \quad \sigma_2^2 = 18$$

类似地，我们计算 AA^T：

$$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$$

因此，我们得到：

- 特征向量：

$$u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

求解得到的特征值也是32和18，但是这不是巧合哦，BA 与 AB 的特征值相同。所以以后我们无需求两遍特征值！
最终，矩阵 A 的奇异值分解为：

$$A = U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^T$$

但是这里有一个小悬念，就是求u_2的时候我们取的是$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$，我们发现正常的求解思维是取$u_2$为 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$。如果我们取后者发现代数$A = U\Sigma V^T$求出的A不是正确的，这是为什么，我们下一部分再为大家揭晓！

带有零空间（nullspace）的例子

考虑矩阵：

$$A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$$

该矩阵有一个一维零空间，且行空间和列空间也都是一维的。

- 行空间由向量 $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ 的倍数组成。
- 列空间由向量 $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$ 的倍数组成。

我们标准化这些向量：

- 行空间基向量：

$$v_1 = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

- 列空间基向量：

$$u_1 = \frac{1}{\sqrt{80}}\begin{bmatrix}4\\8\end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}$$

计算非零特征值：

- 计算 A^TA 的迹（trace）：

$$A^TA = \begin{bmatrix}4 & 8\\3 & 6\end{bmatrix}\begin{bmatrix}4 & 3\\8 & 6\end{bmatrix} = \begin{bmatrix}80 & 60\\60 & 45\end{bmatrix}$$

- 由于矩阵秩为1，一个特征值为0，另一个特征值为迹（trace）125，因此：

$$\sigma_1^2 = 125, \quad \sigma_1 = \sqrt{125} = 5\sqrt{5}$$

最终，矩阵 A 的奇异值分解为：

$$A = U\Sigma V^T = \begin{bmatrix}\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}}\\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\end{bmatrix}\begin{bmatrix}5\sqrt{5} & 0\\0 & 0\end{bmatrix}\begin{bmatrix}\frac{4}{5} & \frac{3}{5}\\ -\frac{3}{5} & \frac{4}{5}\end{bmatrix}^T$$

发现了吗，尽管行空间列空间维度为1，但是我们仍然扩充到了2（让他们正交就可以完成扩充，一般就是调转然后去一个为负即可），这是弥补零空间或者说的准确一点就是零空间的标准正交基，然后通过0把其消除。

总结与回顾

奇异值分解（SVD）将线性代数中的诸多主题巧妙地结合在一起，包括：

- 正定矩阵，
- 四个基本子空间（行空间、列空间、零空间、左零空间）。
- $v_1, v_2, \dots, v_r$ 是行空间的标准正交基；
- $u_1, u_2, \dots, u_r$ 是列空间的标准正交基；
- $v_{r+1}, \dots, v_n$ 是零空间的标准正交基；
- $u_{r+1}, \dots, u_m$ 是左零空间的标准正交基。

这些是“正确”的基，因为它们满足：

$$Av_i = \sigma_i u_i$$

讨论课

有一个矩阵C为$\begin{bmatrix}5 & 5\\-1 & 7\end{bmatrix}$
首先我们求出$C^TC = \begin{bmatrix}26 & 18\\18 & 74\end{bmatrix}$
然后我们求其的特征值于特征向量（这个矩阵求解起来确实很复杂，大家小心计算）

求得特征值为：20与80。对应的特征向量分别为：$\begin{bmatrix}-\frac{3}{\sqrt{10}}\\ \frac{1}{\sqrt{10}}\end{bmatrix},\begin{bmatrix}\frac{1}{\sqrt{10}}\\ \frac{3}{\sqrt{10}}\end{bmatrix}$

然后这个时候你以为我要接着求CC^T吗，为了避免上面我们会犯的错误，我们转变思路！可以知道$C=U\Sigma V^T$，那么$CV=U\Sigma$（V是正交的！），现在我们求出了V以及Σ。我们就可以求出U了！

$CV=\begin{bmatrix}5&5\\-1&7\end{bmatrix}\begin{bmatrix}-\frac{3}{\sqrt{10}}&\frac{1}{\sqrt{10}}\\ \frac{1}{\sqrt{10}}&\frac{3}{\sqrt{10}}\end{bmatrix}=\begin{bmatrix}-\sqrt{10}&2\sqrt{10}\\ \sqrt{10}&2\sqrt{10}\end{bmatrix}$,而$\Sigma=\begin{bmatrix}2\sqrt{5}&0\\0&4\sqrt{5}\end{bmatrix}$。那么$U=CV\Sigma^{-1}$,求得U=$\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}$

这样可以避免讲座中教授犯的错误！然而我们在讲座中犯错误是因为上面的计算方法，我们是无法分辨向量方向的，或者说无法确定特征向量的符号，这会导致问题出错，现在我们用这个方法就没有任何问题了！

到此求出所有分解内容！

习题课

问题 1

验证：如果我们对 Fibonacci 矩阵

$$A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$$

做奇异值分解 $A=U\Sigma V^T$，则

$$\Sigma=\begin{bmatrix}\dfrac{1+\sqrt{5}}{2}&0\\[2ex]0&\dfrac{\sqrt{5}-1}{2}\end{bmatrix}.$$

解答：

计算

$$A^TA=AA^T=\begin{bmatrix}2&1\\1&1\end{bmatrix}.$$

该矩阵的特征值是方程 $x^2-3x+1=0$ 的根，即

$$\lambda=\frac{3\pm\sqrt{5}}{2}.$$

因此

$$\sigma_1^2=\frac{3+\sqrt{5}}{2},\quad \sigma_2^2=\frac{3-\sqrt{5}}{2}.$$

开平方得

$$\sigma_1 = \frac{1+\sqrt{5}}{2}, \quad \sigma_2 = \frac{\sqrt{5}-1}{2}.$$

于是

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}.$$

问题 2

设矩阵 A 的列向量 $w_1, w_2, \ldots, w_n$ 相互正交，长度分别为 $\sigma_1, \sigma_2, \ldots, \sigma_n$。求 A^TA，并给出 A 的 SVD 中的 U, Σ, V。

解答：

由于 A 的列正交，A^TA 是对角矩阵，其对角线元素为 $\sigma_1^2, \ldots, \sigma_n^2$。（这个在正交矩阵那一讲证明过）。而这样的矩阵的特征值就是对角线元素！

由 $A^TA = V\Sigma^2V^T$ 可知

- Σ^2 的对角线元素是 $\sigma_1^2, \ldots, \sigma_n^2$，
- 因此 Σ 的对角线元素为 $\sigma_1, \ldots, \sigma_n$，
- 且 $V = I$（单位矩阵）。

再由 $A = U\Sigma V^T$ 得

$$U \text{ 的列向量就是 } \frac{w_1}{\sigma_1}, \frac{w_2}{\sigma_2}, \ldots, \frac{w_n}{\sigma_n}.$$

其实我也想告诉大家，奇异值分解是本门线性代数中最后一个重要的内容了！如果大家时间不够，直接跳到第3单元复习和考试部分以及期末复习和考试部分！后面剩余的三节内容都是应用性较大，没时间可以跳过！