



[线代] 浅写奇异值分解 (SVD)

近来繁忙，游目骋怀于大千世界，此文成于横加公路一日。

补充前置知识

阅读本文的你应当具备初等代数、线性代数的基本知识。

时间有限，只是随便写点的基础定义，具体的应用之后有时间在ML专栏写。

么正矩阵/酉矩阵

我们知道矩阵转置相当于将矩阵以对角线翻转，将一个 $m \times n$ 的矩阵 M 转置成 $n \times m$ 的矩阵，记做 M^T 。而共轭转置则是在复数域的情况下，转置的同时对矩阵中的元素取复共轭。对于**取复数共轭**的矩阵，我们记作 M^* ，而矩阵 M 的**共轭转置**记作 M^H （在量子力学中常用 $M^\dagger$ ，但是这里是线性代数）。

一个矩阵 M 为酉矩阵的充要条件是它的共轭转置矩阵等于其逆矩阵，即 $M^H = M^{-1}$ 。其数学定义是 $M^H M = M M^H = I_n$ ，实际上，酉矩阵可以理解为正交矩阵在复数域的推广。

一些显而易见的性质，考虑 n 阶的酉矩阵 M, N ：

- M 的行列式的模长为1
- 酉矩阵的所有特征值都是模长为1的复数
- MN 也是酉矩阵
- 对于一个范数 $\|\cdot\|$ ，任意矩阵 A ，若满足 $\|A\| = \|MAN\|$ ，则我们称 $\|\cdot\|$ 为**酉不变范数**。一类代表是 L2范数 和 Frobenius 范数。

对于两个复方阵 P, Q ，考虑酉矩阵 M 若存在

$$P = MQM^H$$

则我们称 P, Q 是**酉等价的**。而若 P, Q 是实数域的方阵，则特别地称它们**正交等价**。

酉矩阵的集合在复数域上构成一个群，称为**酉群**。

特殊正交群

考虑一个交换环 R ，并且 $2^{-1} \in R$ ，用 $M_n(R)$ 表示所有元素属于 R 的 n 阶方阵集合，对于任意的 n 阶方阵 Q ，**正交群** $O(n, R)$ 定义为

$$O(n, R) = \{W \in M_n(R) | Q^\top Q = I_n\}$$

配备矩阵乘法得到的群。而其**特殊正交群**则是正交群中行列式为 1 的元素构成的**正规子群**，表示为 $SO(n, R)$ 。复数域的情况，这被称为**特殊酉群**。

实数域 $\mathbb{R}^{n \times n}$ 上的全体正交矩阵构成一特殊正交群 $SO(n, \mathbb{R}^{n \times n})$ ；复数域 $\mathbb{C}^{n \times n}$ 上全体酉矩阵构成一特殊酉群 $SU(n, \mathbb{C}^{n \times n})$ 。实数域 $\mathbb{C}^{n \times n}$ 上全体酉矩阵构成一特殊正交群 $SO(n, \mathbb{C}^{n \times n})$

特征值

对于一个 n 阶的方阵 M ，一个非零向量 v ，存在标量 λ ，满足

$$Mv = \lambda v$$

则我们称 v 是特征向量， λ 是它的特征值。有同一特征值的特征向量加上零向量，形成一个线性空间称作特征空间。

对于这个 M ，我们寻找一组特征向量，这些特征向量两两正交（它们的内积为0），并且这些特征向量是单位向量，当矩阵作用在这组正交单位向量上时，每个特征向量的长度可能会改变，但它们之间的正交性仍然保持不变。

我们所谓的对角化，是将一个矩阵转化为对角矩阵的过程。对于 M ，存在一个由特征向量构成的可逆正交矩阵 P 使得 P^{-1} 与 M 相乘得到一个对角矩阵 D ，也就是

$$D = P^{-1}AP$$

实现对角化的前提是，原矩阵要有 n 个线性无关的特征向量，并且对于每个特征值 λ 对应的特征向量是构成矩阵 P 的列向量。

奇异值

考虑一个 $m \times n$ 的矩阵 M ，我们可以构建两个对称的方阵 $A_1 = MM^\top$ 或者 $A_2 = M^\top M$ ，它们具有相同的非零特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ ，其中 $r = \text{rank}(M)$ ， M 的奇异值定义为特征值的平方根

$$\sigma_i = \sqrt{\lambda_i}$$

有限维情况的谱定理

概括地说，谱定理给出了算子或者矩阵可以对角化的条件，但我们在线性代数中研究的主要都是有限维的线性空间，而谱定理比较广泛，适用于无限维空间中的自伴算子。所以本文我们仅讨论有限维的情况，无限维还是希尔伯特空间什么的，就留给以后的泛函分析文章吧。

我们将满足 $MM^H = M^H M$ 的矩阵 M 称为**正规矩阵**。

自伴随矩阵，或称厄米矩阵，指的是共轭对称的方阵（有一个等价条件是 $H^H = H$ ，如果是实矩阵的情况，则等价于 $H^T = H$ ）。容易注意到它是特殊的正规矩阵。每个**正规矩阵** H 都可以被酉矩阵对角化，即存在酉矩阵 U 和对角矩阵 N 使得

$$H = UNU^H$$

但是，正规矩阵的特征值可以是实数或者复数。而自伴随矩阵 H 的特征值都是实数，意味着自伴随矩阵的特征向量可以被选择为标准正交的，因此存在**标准正交基** V ，由 H 的特征向量组成。

奇异值分解

奇异值分解（Singular Value Decomposition, SVD）是线性代数中一种重要的矩阵分解方法。我的领域多用于在 PCA 中进行特征提取、降维以及提高模型的泛化能力。

一般描述

我们将每个元素都属于 $\mathbb{K}$ （即实数域或复数域）的 $m \times n$ 的矩阵 M 分解成三个矩阵

$$M = U\Sigma V^H$$

其中：

- U ： $m \times m$ 的正交特征向量矩阵（即么正矩阵/酉矩阵，注意这强调的是复数域的情况下）
- Σ ：有 $m \times n$ 个元素的非负实数对角矩阵，其主对角线上的每个元素都称为奇异值，等于 MM^H 或 $M^H M$ 的正特征值的根（注意是实数，所以要求是非负平方根）
- V^H ：一个 $n \times n$ 的酉矩阵 V 的共轭转置

这样的分解我们称之为**奇异值分解**。并且，我们一般将 U 中的列向量称作左奇异向量， V 中的列向量称作右奇异向量，而 Σ 的主对角元 σ_i 被是矩阵的奇异值。由于 V 是正交的。所以我们亦可以将上述分解写成另一个形式

$$MV = U\Sigma$$

另外强调，一个矩阵的奇异值分解并不是唯一的，只是有些时候为了方便，我们会将奇异值从大到小地排序。

当一个方阵可以对角化时，意味着存在一个基底，其中对它的作用可以被表示为一个对角矩阵的作用。这从几何意义上看，我们可以将奇异值分解理解为将原空间中的向量转换到一个新的基底（这个基底由 $M^H M$ 的特征向量构成）；在新基底对向量进行伸缩，每个方向的缩放比例由奇异值确定；将缩放后的向量转换回到目标空间中的新基底（这个基底由 $M M^H$ 的特征向量构成）。因此我们可以说，奇异值分解可以被看作是对角化更一般的推广。

正交性及扩展

前面的定义可能会有点武断，所以我们展开说说。继续之前的定义，接下来，我们将用 u_i 和 v_i 分别表示 U 和 V 的列， $r = \text{rank}(M)$ 。SVD 提供了矩阵 M 的四个子空间的正交基：

- $u_1, u_2, \dots, u_r$ 形成 M 的列空间的正交基
- $v_1, v_2, \dots, v_r$ 形成 M 的行空间的正交基
- $u_{r+1}, \dots, u_m$ 形成 M 的左零空间的正交基
- $v_{r+1}, \dots, v_n$ 形成 M 的零空间的正交基

关注到后两者，对于 $u_{m-r}, \dots, u_m$ 和 $v_{n-r}, \dots, v_n$ ，显然因为整个零空间和左零空间本身就是与行空间和列空间正交的，所以它们与前 r 个基向量是正交的。在构造正交矩阵 U, V 时，将这些向量全部包含在内，以保证它们是方阵，使得 $MV = U\Sigma$ 成立。

最优低秩近似

最优低秩近似是指在给定矩阵和指定秩 k 的情况下，找到一个秩为 k 的矩阵，使其在某种范数下与原矩阵的误差最小。

Eckart–Young–Mirsky 定理

矩阵的低秩近似通过保留最大的前 k 个奇异值及其对应的奇异向量来实现。对于给定的 $k < r$ ，矩阵 M 的最佳秩- k 近似矩阵 M_k 可以表示为

$$M_k = \sum_{i=1}^k \sigma_i u_i v_i^*$$

这个近似在最小化 Frobenius 范数误差方面是最优的，即

$$\|M - M_k\|_F = \min_{\text{rank}(N) \leq k} \|M - N\|_F$$

Eckart–Young–Mirsky 定理的内容即为，对于任何正整数 k ($k \leq \min(m, n)$)，存在一个秩为 k 的矩阵 M_k 使得 M_k 的 Frobenius 范数 $\|M - M_k\|_F$ 是所有秩为 k 的矩阵 N 与 M 之间 Frobenius 范数的最小值。

引用一个我比较喜欢的证明：<https://truenobility303.github.io/Young/>

Eckart-Young-Mirsky 定理及其泛函版本简证

一种贪婪算法

顺便讲讲一种思路。

对于一个 $m \times n$ 的矩阵 M ，其中每个行向量 m_i 都能表示为 n 维空间中的一个点。我们选择一个单位向量 v 来表示一条过原点的直线方向，即每个 m_i 在 v 方向的投影长度为 $|m_i \cdot v|$ ，我们希望找到的是使得所有投影长度的平方和 $|Mv|^2$ 最大的直线，即使得其每个对应点的距离平方和最小的直线。

所以第一个奇异向量定义为

$$v_1 = \arg \max_{\|v\|=1} |Mv|$$

其对应的第一个奇异值为

$$\sigma_1 = \max_{\|v\|=1} |Mv_1|$$

由于我们使用的是投影长度平方和，任何包含 v_1 的 2 维子空间的投影长度平方和等于投影到 v_1 方向的平方和加上在 v_1 垂直方向上的平方和，所以同理地

$$v_2 = \arg \max_{\|v\|=1, v \perp v_1} |Mv|$$

一样，寻找第 k 个奇异向量 v_k 是在所有与之前找到的奇异向量正交的单位向量中，使 $|Mv|^2$ 最大化的过程

$$v_k = \arg \max_{\|v\|=1, v \perp v_1, v_2, \dots, v_{k-1}} |Mv|$$

对应的奇异值为

$$\sigma_k = \max_{\|v\|=1, v \perp v_1, v_2, \dots, v_{k-1}} |Mv_k|$$

参考

* 文章的部分表述我参考了记忆中阅读过的一些资料，因当下不方便查阅，可能相关部分表述和理解会与原文有不小出入。

1. *Chapter 7 - The Singular Value Decomposition (SVD)*