

摘要

SVD(Singular Value Decomposition, 奇异值分解) 是线性代数中既优雅又强大的工具, 它揭示了矩阵最本质的变换. 使用 SVD 对矩阵进行分解, 能得到代表矩阵最本质变化的矩阵元素. 这就好比一个合数能表示为若干质数之积, 分解合数能得到表示该合数的质因数; 复杂周期信号可以表示为若干简单的正弦波和余弦波之和, 使用傅里叶变换能得到表示该信号的简单波; 复杂矩阵所代表的线性变换可由若干个简单矩阵所代表的线性变换组合起来, 使用 SVD 能找到这些简单矩阵.

关键词: 矩阵论, 数据压缩, 数据降维, 机器学习

SVD 的数学涵义

矩阵在线性代数系统中是一个核心的概念, 其从不同的角度出发都能拥有丰富的内涵. 对于矩阵 $\mathbf{A}_{m \times n}$, 当其参与运算

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (1)$$

时, 我们可以从以下三个角度看待其角色:

1. 矩阵 \mathbf{A} 是线性方程组 (1) 的系数组成的矩阵, 其每一行是 (1) 中每一个方程式的系数部分, 通过分析矩阵的秩 $\text{rank}(\mathbf{A})$ 和其极大线性无关组的情况, 我们可以了解 (1) 的解的情况, 同时, 对于使用高斯消元法等进行求解也比较方便;
2. 当 $m \geq n \wedge \text{rank}(\mathbf{A}) = n$ 时, 矩阵 \mathbf{A} 是 R^n 空间中的一个基, 在这个基上面, 有向量 $\vec{x} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$, 而此向量在标准正交基上表示为 $\vec{b} = [\mathbf{b}_1, \dots, \mathbf{b}_m]^\top$, 此时 (1) 隐含着—个基变换的关系, 即 $\mathbf{A}\mathbf{x} = \mathbf{I}\mathbf{b}\mathbf{I}$ 为标准正交基;
3. 矩阵 \mathbf{A} 本身表示一个线性变换, (1) 表示其对向量 \vec{x} 进行线性变换得到向量 \vec{b} 的过程.

上述的关于矩阵的各种角色与我们阐述 SVD 有什么关系呢? 当我们将矩阵视为一种线性变换时, SVD 可以帮我们揭示组成该线性变换的最本质的变换, 具体地, SVD 揭示了这样的一个事实: **对于任意的矩阵 \mathbf{A} , 我们总能找到一组单位正交基, 使得 \mathbf{A} 对其进行变换之后, 得到的向量组仍然是正交的.** 这样的表述还是相当地晦涩, 我们不妨在二维平面中举一个例子.

设有矩阵 \mathbf{A} , 其对单位正交基 \vec{v}_1, \vec{v}_2 进行线性变换, 得到的向量仍然是彼此正交的, 即

$\mathbf{A}\vec{v}_1, \mathbf{A}\vec{v}_2$ 仍然是正交的. 设 $\mathbf{A}\vec{v}_1, \mathbf{A}\vec{v}_2$ 方向上的单位向量是 $\vec{\mu}_1, \vec{\mu}_2$, 长度是 σ_1, σ_2 , 则我们可得

$$\mathbf{A}\vec{v}_1 = \sigma_1 \vec{\mu}_1 \quad (2)$$

$$\mathbf{A}\vec{v}_2 = \sigma_2 \vec{\mu}_2 \quad (3)$$

现在利用矩阵 \mathbf{A} 对向量 \vec{x} 进行线性变换. 我们先将向量 \vec{x} 在单位正交基 \vec{v}_1, \vec{v}_2 上进行表示, 即

$$\begin{aligned} \vec{x} &= [\mathbf{v}_1, \mathbf{v}_2] \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\ &= [\mathbf{v}_1, \mathbf{v}_2] \cdot \begin{bmatrix} v_1^\top \vec{x} \\ v_2^\top \vec{x} \end{bmatrix} \end{aligned} \quad (4)$$

由 (2), (3), (4), 我们有

$$\begin{aligned}
\because \mathbf{A}\vec{x} &= \mathbf{A} \cdot [v_1, v_2] \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= [\mathbf{A}v_1, \mathbf{A}v_2] \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= [\sigma_1\mu_1, \sigma_2\mu_2] \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= [\mu_1, \mu_2] \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= \mathbf{U}\Sigma\mathbf{V}^\top \cdot \vec{x} \\
\therefore \mathbf{A} &= \mathbf{U}\Sigma\mathbf{V}^\top
\end{aligned} \tag{5}$$

至此, 我们由 " 对于任意的矩阵 \mathbf{A} , 我们总能找到一组单位正交基, 使得 \mathbf{A} 对其进行变换之后, 得到的向量组仍然是正交的 ", 即 (2)(3) 出发, 得到了矩阵 \mathbf{A} 最终的分解形式 (6). (6) 表达了这样一个事实, 对于任意的矩阵 \mathbf{A} , 我们总可以将其分解为一个酉矩阵 \mathbf{U} , 一个对角矩阵 Σ 和另一个酉矩阵的转置 \mathbf{V}^\top 的乘积, 这便是 SVD 的核心内容.

SVD 的几何涵义

现在我们知道, 对于任意的矩阵 \mathbf{A} , 我们总可以将其分解为一个酉矩阵 \mathbf{U} , 一个对角矩阵 Σ 和另一个酉矩阵的转置 \mathbf{V}^\top 的乘积, 即等式 (6) 所表述的内容. $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top$ 表示矩阵 \mathbf{A} 所代表的线性变换可以由更简单的旋转, 拉伸变换进行合成. 这些更简单的变换是怎么进行生效的呢? 我们还是在二维平面中举例说明.

当使用矩阵 \mathbf{A} 对向量 \vec{x} 进行变化时, 我们可以先将向量 \vec{x} 在单位正交基 \vec{v}_1, \vec{v}_2 上进行表示, 即 (4) 所表述. 我们不妨令 $\xi_1 = v_1^\top \vec{x}, \xi_2 = v_2^\top \vec{x}$, 则 ξ_1, ξ_2 是向量 \vec{x} 在单位正交基 \vec{v}_1, \vec{v}_2 上的坐标, 即

$$\begin{aligned}
\vec{x} &= [v_1, v_2] \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= [v_1, v_2] \cdot \begin{bmatrix} v_1^\top \vec{x} \\ v_2^\top \vec{x} \end{bmatrix} \\
&= [v_1, v_2] \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}
\end{aligned} \tag{7}$$

由 (6), (7) 我们有

$$\begin{aligned}
\mathbf{A}\vec{x} &= \mathbf{U}\Sigma\mathbf{V}^\top \cdot \vec{x} \\
&= [\mu_1, \mu_2] \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot \vec{x} \\
&= [\mu_1, \mu_2] \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} \cdot [v_1, v_2] \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}
\end{aligned} \tag{8}$$

现在我们仔细地来分析 (8) 中各矩阵的具体操作效果.

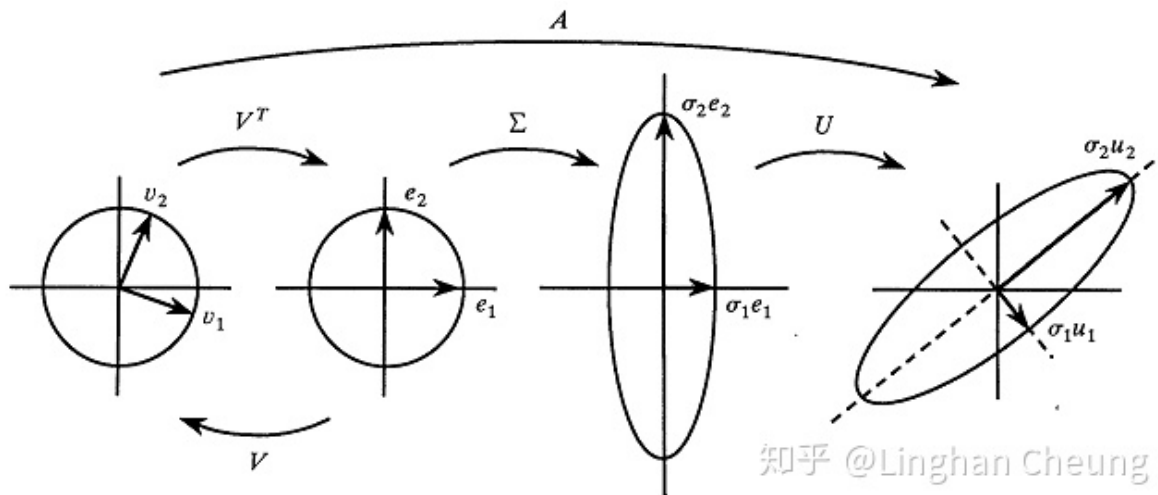
$$\mathbf{A}\vec{x} = \underbrace{[\mu_1, \mu_2]}_{\text{旋转}} \cdot \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}_{\text{拉伸}} \cdot \underbrace{\begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix}}_{\text{旋转}} \cdot \underbrace{[v_1, v_2]}_{\text{单位正交基 } \mathbf{V}} \cdot \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}}_{\text{x 在 } \mathbf{V} \text{ 坐标}} \quad (9)$$

$\underbrace{\hspace{10em}}_{\text{对单位正交基 } \mathbf{V} \text{ 进行旋转, 使之变为标准正交基 } \mathbf{I}}$
 $\underbrace{\hspace{10em}}_{\text{对标准正交基进行拉伸}}$
 $\underbrace{\hspace{10em}}_{\text{对拉伸后的正交基进行旋转}}$

如 (9) 所示, 矩阵 \mathbf{A} 对向量 \vec{x} 进行线性变换, 其先将向量 \vec{x} 用单位正交基 \mathbf{V} 进行表示. 然后使用酉矩阵 \mathbf{V}^\top 进行旋转, 由酉矩阵的性质我们可知 $\mathbf{V}\mathbf{V}^\top = \mathbf{V}^\top\mathbf{V} = \mathbf{I}$, 所以旋转之后我们可得到标准正交基 \mathbf{I} . 然后使用矩阵 Σ 对标准正交基 \mathbf{I} 进行拉伸, 使得 $x - axis, y - axis$ 分别拉伸 σ_1, σ_2 倍的长度. 最后再使用酉矩阵 \mathbf{U} 对拉伸之后的正交基进行旋转, 得到最终的基, 从而得到最终的向量为

$$\begin{aligned} \mathbf{A}\vec{x} &= [\sigma_1\mu_1, \sigma_2\mu_2] \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \\ &= \xi_1\sigma_1\mu_1 + \xi_2\sigma_2\mu_2 \end{aligned} \quad (10)$$

上述过程可表示为下图



通过 SVD, 我们找到了能代表矩阵 \mathbf{A} 作为线性变换时最本质的操作. 而 σ_1, σ_2 就是所谓的奇异值, 表示对标准正交基各个轴进行拉伸的程度.

SVD 的求解过程

上述关于 SVD 在二维平面上的结论可以轻易地推广到多维情况. 那 SVD 具体如何求解呢? 由 (6) 我们知道 SVD 能使矩阵 \mathbf{A} 进行分解, 现在由 (6) 出发我们来构造矩阵 $\mathbf{U}, \Sigma, \mathbf{V}$, 具体地, 我们有

$$\begin{aligned} \mathbf{A}\mathbf{A}^\top &= \mathbf{U}\Sigma\mathbf{V}^\top\mathbf{V}\Sigma^\top\mathbf{U}^\top \\ &= \mathbf{U}\Sigma^2\mathbf{U}^\top \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{A}^\top\mathbf{A} &= \mathbf{V}\Sigma^\top\mathbf{U}^\top\mathbf{U}\Sigma\mathbf{V}^\top \\ &= \mathbf{V}\Sigma^2\mathbf{V}^\top \end{aligned} \quad (12)$$

由实对称矩阵必可正交对角化, 即

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top \quad (13)$$

其中矩阵 \mathbf{Q} 为酉矩阵, 即满足 $\mathbf{Q}^\top = \mathbf{Q}^{-1}$. 矩阵 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ 为矩阵 \mathbf{A} 的特征值所组成的对角矩阵. 而矩阵 $\mathbf{A}\mathbf{A}^\top, \mathbf{A}^\top\mathbf{A}$ 是实对称矩阵, 矩阵 \mathbf{U}, \mathbf{V} 是酉矩阵, 矩阵 $\mathbf{\Sigma}^2$ 是对角矩阵, 所以由 (11), (12), (13), 我们对矩阵 $\mathbf{A}\mathbf{A}^\top, \mathbf{A}^\top\mathbf{A}$ 进行正交对角化, 即可得到矩阵 \mathbf{U}, \mathbf{V} . 再由

$$\begin{aligned}\therefore \mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \\ \mathbf{A}\mathbf{V} &= \mathbf{U}\mathbf{\Sigma} \\ \mathbf{A}\mathbf{v}_i &= \sigma_i \mu_i \\ \therefore \sigma_i &= \frac{\mathbf{A}\mathbf{v}_i}{\mu_i}\end{aligned}\tag{14}$$

由 (14) 我们便可得到矩阵 $\mathbf{\Sigma}$. 由 (11), (12), (13), (14) 即可得完整的 SVD 分解.

SVD 的具体应用

除了前文所述, SVD 揭示了矩阵进行线性变换时最本质的变换, 使我们能了解矩阵的具体操作, 这是 SVD 最直接的应用. 除此之外, SVD 还有许多其他方面的应用, 下面举例说明.

压缩

许多存储在计算机中的数据都是以矩阵的形式存在的, 进行合理的矩阵压缩能把存储矩阵所占的空间缩减下来. 例如图像, 事实上一个灰度图像就是一个矩阵, 矩阵中的每个元素就是灰度图像的像素值. 如果有灰度图 $\mathbf{A}_{m \times n}$, 由 (6) 我们有

$$\begin{aligned}\mathbf{A}_{m \times n} &= \mathbf{U}_{m \times m} \cdot \mathbf{\Sigma}_{m \times n} \cdot \mathbf{V}_{n \times n}^\top \\ &= \sum_{i=1}^n \sigma_i \mu_i \mathbf{v}_i^\top\end{aligned}\tag{15}$$

奇异值 $\sigma_i, i = 1, \dots, n$ 有一定的大小关系, 我们不妨设 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$, 取前 k 个分量, 则由 (15) 可知, 若一个像素为 1 字节, 原始图像需 $m \times n$ 字节的存储空间, 而使用 SVD 分解后只需 $k \times (1 + m + n)$ 字节的存储空间, 以此达到压缩图像 (矩阵) 的目的.

降维

数据降维在机器学习, 数据挖掘等领域是一个重要的技术, 通过数据降维可以挖掘数据的关键信息, 降低运算的成本. 使用 SVD 进行降维的核心思想是, 通过对 feature 向量 (如机器学习中的数据向量) 所组成的矩阵 \mathbf{X} 进行分解, 可直接得到降维后的 feature 向量矩阵, 这其实就是 PCA (主成分分析, Principal component analysis) 的过程 [5], 具体地, 令矩阵 \mathbf{X} 是若干 feature 向量 \vec{x} 所组成的 feature 向量矩阵且矩阵 \mathbf{X} 已经进行去均值处理, 则我们有

$$\begin{aligned}\max \quad & \text{tr}(\mathbf{W}^\top \mathbf{X}\mathbf{X}^\top \mathbf{W}) \\ \text{s.t.} \quad & \mathbf{W}^\top \mathbf{W} = \mathbf{I}\end{aligned}\tag{16}$$

(16) 即为 PCA 的求解过程, 熟悉 PCA 的同学都知道, 这个过程实际上是通过矩阵 $\mathbf{X}\mathbf{X}^\top$ 进行正交对角化求出投影矩阵 \mathbf{W} , 使得投影后的矩阵 $\mathbf{Y} = \mathbf{W}^\top \mathbf{X}$ 各个维度相互独立, 即 $\mathbf{Y}\mathbf{Y}^\top$ 为对角矩阵, 然后取方差最大的若干维以达到降维的效果. 所以我们需要对协方差矩阵 $\mathbf{X}\mathbf{X}^\top$ 进行对角化 $\mathbf{X}\mathbf{X}^\top = \mathbf{W}\mathbf{Y}\mathbf{Y}^\top \mathbf{W}^\top$ 以求出投影矩阵 \mathbf{W} . 而现在我们可以直接对协方差矩阵进行 SVD 分解

$$\begin{aligned}\mathbf{X}\mathbf{X}^T &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T \\ &= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T\end{aligned}\quad (17)$$

(17) 中西矩阵 \mathbf{U} 即为投影矩阵 \mathbf{W} . 由 (17) 可得

$$\begin{aligned}\mathbf{W}^T\mathbf{X} &= \mathbf{U}^T\mathbf{X} \\ &= \mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= \mathbf{\Sigma}\mathbf{V}^T\end{aligned}\quad (18)$$

由 (18) 我们可得投影后的矩阵 $\mathbf{Y} = \mathbf{\Sigma}\mathbf{V}^T$, 此时我们只需

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\quad (19)$$

由 (19) 即可直接求出投影后的向量, 而无需先求得投影矩阵 \mathbf{W} 再进行投影计算. 而后我们便可以取方差最大的若干维, 从而达到降维的目的.

总结

本文从 SVD 的数学原理出发, 推导出其一般形式, 并在此基础上给出了几何解释. 然后介绍了 SVD 的求解方法和具体应用. SVD 是一个非常优雅且实用的方法, 其应用场景颇多, 是构成许多现代计算系统的核心部件.

引用

- [1] Austin, D. (2019). *We Recommend a Singular Value Decomposition*. [online] Ams.org. Available at: <http://www.ams.org/publicoutreach/feature-column/fcarc-svd> [Accessed 27 Feb. 2019].
- [2] Wikipedia contributors. "酉矩阵." *维基百科, 自由的百科全书*. 维基百科, 自由的百科全书, 15 Nov. 2018. Web. 15 Nov. 2018. <<https://zh.wikipedia.org/w/index.php?title=%E9%85%89%E7%9F%A9%E9%98%B5&oldid=52035033>>.
- [3] Cnblogs.com. (2017). *奇异值分解 (SVD) 原理与在降维中的应用 - 刘建平 Pinard - 博客园*. [online] Available at: <https://www.cnblogs.com/pinard/p/6251584.html> [Accessed 4 Mar. 2019].
- [4] 線代啟示錄. (2011). *實對稱矩陣可正交對角化的證明*. [online] Available at: <https://ccjou.wordpress.com/2011/02/09/%E5%AF%A6%E5%B0%8D%E7%A8%B1%E7%9F%A9%E9%99%A3%E5%8F%AF%E6%AD%A3%E4%BA%A4%E5%B0%8D%E8%A7%92%E5%8C%96%E7%9A%84%E8%AD%89%E6%98%8E/> [Accessed 4 Mar. 2019].
- [5] Li, E. (2018). *求pca的时候, 能不能不用协方差矩阵?*. [online] Zhihu.com. Available at: <https://www.zhihu.com/question/39234760/answer/80323126> [Accessed 29 Mar. 2019].