SVD 分解

矩阵的特征向量和特征值

概念

基本公式: $A \cdot x = \lambda \cdot x$

x 是矩阵 A 的特征向量 λ 是矩阵 A 的特征值

x 经过矩阵 A 的线性变换后没有变换方向，只是经过了 λ 标量的拉伸 / 压缩操作

计算过程

$$[A] \cdot x - \lambda \cdot x = 0$$

在后项加入一个单位矩阵 E 并不会影响结果

$$[A] \cdot x - \lambda \cdot E \cdot x = 0$$

$mention$: E 是一个对角矩阵，且对角线上的值均为 1 ，任何矩阵与 E 相乘都会得到原矩阵本身, 形如： $E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

已知，特征向量 x 不会为 0

$$([A] - \lambda \cdot E) = 0$$

将 A 和 对应的 E 带入上述表达式，即可计算出 λ 的值（特征值）

基于计算得到的 λ 的值即可计算出多个对应的特征向量 x

对于矩阵 A 而言可能会计算得到多个特征值 λ

同样也就会产生多个特征向量 x

假设存在二维矩阵 A 为：

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$$

计算矩阵 A 的特征向量和特征值

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} 1-\lambda & 2 \\ 3 & -4-\lambda \end{bmatrix} = 0$$

$$(1-\lambda)(-4-\lambda) - 2 \times 3 = 0$$

$\lambda = 2$ 或 $\lambda = -5$

$\lambda = 2$ 时 $\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \times \begin{bmatrix} m \\ n \end{bmatrix} = 2 \begin{bmatrix} m \\ n \end{bmatrix}$ 所以 $m = 2n, x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

背景知识：

对于矩阵 A 而言，什么情况下是他的特征向量是线性无关的：

1. 矩阵 A 有 n 个不同的特征值，其中 n 是矩阵 A 的维数，每个特征向量对应于一个不同的特征值。在这种情况下，每个特征向量是线性无关的，因为它们对应于不同的特征值。

2. 矩阵 A 是对称矩阵，即 A 等于其转置矩阵的矩阵，且所有的特征值都是实数。在这种情况下，A 的特征向量也是实数向量，并且对应于不同的特征值的特征向量是正交的。因此，它们是线性无关的。

于是，假设我们求出了矩阵 A 的 n 个特征值$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

以及这 n 个特征值所对应的特征向量$\{w_1, w_2, \ldots, w_n\}$

如果满足，**特征向量之间线性无关**，

使用 W 表示 n 个特征向量组成的 n x n 维度的矩阵(基于背景知识 1)

使用 Σ 表示 n 个特征值为主对角线的 n x n 维矩阵（对角矩阵）

那么我们可以得到矩阵 A 可以通过下面的表示进行分解

$$AW = W\Sigma$$
$$A = W\Sigma W^{-1}$$

一般我们会把 W 中的 n 个特征向量进行标准化处理，即确保

$||w_i||_2 = 1$ 或 $w_i^T w_i = 1$

此时 W 的 n 个特征向量构成标准正交基

满足$W^T W = I$ 也就是 $W^T = W^{-1}$

也称 W 叫做 **unitary matrix**

将上式带入$A = W\Sigma W^{-1}$ 就会得到 $A = W\Sigma W^T$

也就是 A 矩阵可以由 特征向量组成的矩阵**乘以**特征值构成的对角矩阵**乘以**特征向量组成的矩阵的转置构成

字面意思就是，矩阵可以通过其所有特征向量和特征值共同表达

但是若需要对 A 进行特征分解，基于上面的假设，矩阵 A 就要求是长宽一致的方阵，

如果 A 不是方阵，行和列的长度不同时，就需要使用 SVD 进行特征分解了

SVD 的定义

SVD 不要求需要被分解的矩阵是方阵，

假设矩阵 A 是一个 m x n 维度的矩阵

那么定义矩阵 A 的 SVD ：

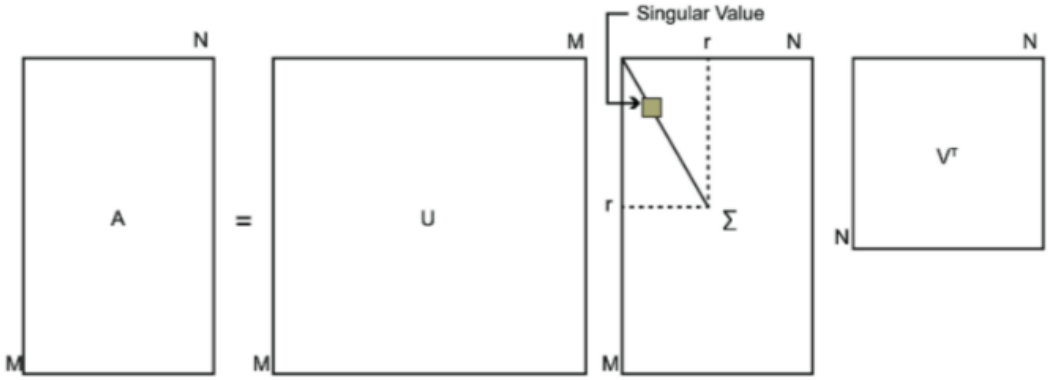
$$A = U\Sigma V^T$$

U 是一个 mxm 维度的矩阵

Σ是一个 mxn 维度的矩阵（奇异值矩阵），且为主对角矩阵（除了主对角线上的元素之外都是 0），**主对角线上的每一个元素都称为奇异值，**

V 是一个 n x n 的矩阵

U 和 V 都是 unitary matrix，满足$U^TU = I, V^TV = I$



求解 SVD

对于不构成方阵的矩阵，求解的过程是将其转化为方阵后再进行矩阵的特征值/特征向量计算

论证：A 的转置和 A 进行矩阵相乘后的到的方阵，其特征向量组成的就是 SVD 公式中的 V 矩阵

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

上式的证明中使用了基本的设定，$U^TU = I$ **和** $\Sigma^T\Sigma = \Sigma^2$

推导可以得到 A 和 A 的转置相乘得到的方阵，其特征向量就是 SVD 中的 V 矩阵

首先将 A 和 A 的转置做矩阵乘法，得到一个 mxm 的方阵，对这个方阵进行特征分解，可以得到其特征向量和特征值满足：

$$(AA^T)u_i = \lambda_i u_i$$

论证：A 和 A 的转置进行矩阵相乘后的到的方阵，其特征向量组成的就是 SVD 公式中的 U 矩阵

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

上式的证明中使用了基本的设定，$V^TV = I$ 和 $\Sigma^T\Sigma = \Sigma^2$

推导可以得到 A 的转置和 A 相乘得到的方阵，其特征向量就是 SVD 中的 U 矩阵

对于 A 和 A 的转置相乘得到的方阵，存在 m 个特征向量 u（长度为 m），也就存在 m 个特征值 λ

将所有特征向量 u 组成一个 *mxm* 的矩阵，构成 U, 就是 SVD 公式中的 U 矩阵，每一个 U 中的特征向量又称为 A 的左奇异向量

同样的思路，

对 A 的转置和 A 做矩阵乘法，得到一个 nxn 的方阵，对这个方阵进行特征分解，可以得到其特征向量和特征值满足：

$$(A^TA)v_i = \lambda_i v_i$$

对于 A 的转置和 A 相乘得到的方阵，存在 n 个特征向量 v（长度为 n），也就存在 n 个特征值 λ

将所有特征向量 v 组成一个 *nxn* 的矩阵，构成 V，就是 SVD 公式中的 V 矩阵，每一个 V 中的特征向量又称为右奇异向量

在已知 V 和 U 可以求得的情况下，求解奇异值矩阵Σ

鉴于奇异值矩阵式一个主对角线上存在奇异值，其余位置均为 0 的矩阵，所以只需要求解出每一个奇异值 σ 即可

$$A = U\Sigma V^T$$
$$AV = U\Sigma V^TV$$
$$AV = U\Sigma$$
$$Av_i = \sigma_i u_i$$
$$\sigma_i = Av_i \div u_i$$

基于上述公式，即可求解每一个u_i 和 v_i 对应的奇异值

SVD 的计算举例

假设矩阵 A 为

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

首先求解出 A^TA 和 AA^T 这两个方阵

$$A^TA=\begin{pmatrix}0&1&1\\1&1&0\end{pmatrix}\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix}$$

$$AA^T=\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}\begin{pmatrix}0&1&1\\1&1&0\end{pmatrix}=\begin{pmatrix}1&1&0\\1&2&1\\0&1&1\end{pmatrix}$$

求解 A^TA 的特征向量和特征值

$$\lambda_1=3;\ v_1=\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix};\quad \lambda_2=1;\quad v_2=\begin{pmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}$$

求解 AA^T 的特征向量和特征值

$$\lambda_1=3;\quad u_1=\begin{pmatrix}1/6\\2\sqrt{6}\\1/6\end{pmatrix};\quad \lambda_2=1;\quad \mu_2=\begin{pmatrix}1/\sqrt{2}&0\\0&\\-1/\sqrt{2}&\end{pmatrix};\quad \lambda_3=0;\quad a_3=\begin{pmatrix}1&\sqrt{3}\\-1/\sqrt{3}&\\1&\sqrt{3}\end{pmatrix}$$

分别利用 $Av_i=\sigma_iu_i,i=1,2$(存在两套特征向量)，求解奇异值 σ_1 和 σ_2

$$\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\1\sqrt{2}\end{pmatrix}=\sigma_1\begin{pmatrix}1\sqrt{6}\\2\sqrt{6}\\1\sqrt{6}\end{pmatrix}\rightarrow\sigma_1=\sqrt{3}$$

$$\begin{pmatrix}0&1\\1&1\\1&0\end{pmatrix}\begin{pmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}=\sigma_2\begin{pmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{pmatrix}\rightarrow\sigma_2=1$$

最终使用 u 组成 U，v 组成 V，σ 组成 Σ

$$A=U\Sigma V^T=\begin{pmatrix}1\sqrt{6}&1/\sqrt{2}&1/\sqrt{3}\\2\sqrt{6}&0&-1\sqrt{3}\\1/\sqrt{6}&-1\sqrt{2}&-1/\sqrt{3}\end{pmatrix}\begin{pmatrix}\sqrt{3}&0\\0&1\\0&0\end{pmatrix}\begin{pmatrix}1/\sqrt{2}&1/\sqrt{2}\\-1/\sqrt{2}&1/\sqrt{2}\end{pmatrix}$$

SVD 的应用

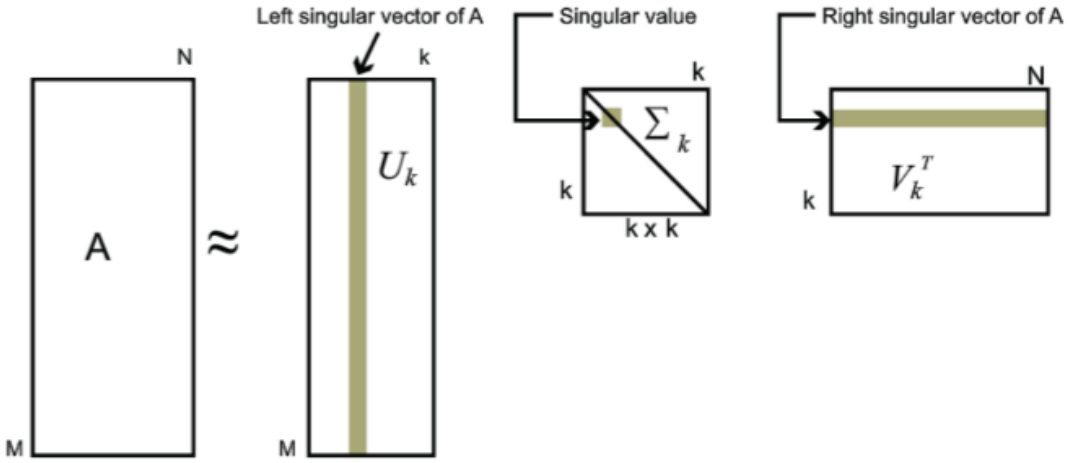
对于奇异值矩阵而言，矩阵中的奇异值也是从左上到右下按照从大到小的顺序进行排列的

基于实验计算，奇异值的递减速度很快，如果使用标量分布进行计算，前 10%的奇异值就占据了全部奇异值之和 99% 以上的比例

换句话来说，前 10%的奇异值矩阵及其对应的特征向量（矩阵）是可以近似描述原矩阵的特征信息的

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}\approx U_{m\times k}\Sigma_{k\times k}V^T_{k\times n}$$

其中 k 的取值可以是 n 的 10%，也就实现了矩阵的降维操作



降维的思路可以用于 PCA 的降维，可以用于数据的压缩和提取，放在 NLP 中，对共现矩阵进行操作，可以实现单词与单词之间的语义提取

举例来说，SVD 不仅仅适用于共现矩阵，也适用于 TF-IDF 和 BOW 生成的向量空间模型

使用 SVD 可以找到文档与词汇之间的相关性和文档与文档之间的相关信息

通过特征向量和特征值的方式，表达存在高度相关性的词语，这些词语可以表达文档的主题

从语言学的角度来说，从 1950-1990 年间一直有科学家研究 '**单词之间的语义相似性与上下文使用的词语和表达方式的相似性存在正比关系**'

比如在 LSA（latent semantic analysis 潜在语义分析中），首先使用 TF-IDF 构建单词与文档的矩阵，再使用 SVD 进行降维操作

Example

假设我们存在三句话，使用 window_size= 1(关注中心词的上一个词和下一个词)

1. I enjoy flying
2. I like NLP
3. I like deep learning

I = enjoy(1 次), like(2 次)

enjoy = I (1 次), flying(2 次)

flying = enjoy(1 次)

like = I(2 次), NLP(1 次), deep(1 次)

NLP = like(1 次)

deep = like(1 次), learning(1 次)

learning = deep(1 次)

	NLP	flying	I	like	deep	learning	enjoy
NLP	0.0	0.0	0.0	1.0	0.0	0.0	0.0
flying	0.0	0.0	0.0	0.0	0.0	0.0	1.0
I	0.0	0.0	0.0	2.0	0.0	0.0	1.0
like	1.0	0.0	2.0	0.0	1.0	0.0	0.0
deep	0.0	0.0	0.0	1.0	0.0	1.0	0.0
learning	0.0	0.0	0.0	0.0	1.0	0.0	0.0
enjoy	0.0	1.0	1.0	0.0	0.0	0.0	0.0

这里为我们的矩阵 A 是一个 7x7 的矩阵，对其进行 SVD 分解得到矩阵 U（在没有进行特征值矩阵选取的基础上）为：

	0	1	2	3	4	5	6
NLP	-3.471975e-01	0.000000e+00	-1.942887e-01	0.000000e+00	0.000000e+00	-4.183768e-01	8.164966e-01
flying	-1.396622e-01	0.000000e+00	6.739877e-01	-9.992007e-16	-4.440892e-16	5.996402e-01	4.082483e-01
enjoy	1.249001e-16	-3.687707e-01	-2.220446e-16	-8.041284e-01	4.662464e-01	-1.110223e-16	-5.551115e-17
I	-8.340573e-01	-5.277047e-18	2.854104e-01	-7.948171e-16	-1.162003e-15	-2.371134e-01	-4.082483e-01
deep	-4.053355e-01	0.000000e+00	-6.530953e-01	0.000000e+00	0.000000e+00	6.396638e-01	-1.110223e-16
like	1.665335e-16	-9.167567e-01	-2.775558e-17	2.318040e-01	-3.253062e-01	5.551115e-17	1.110223e-16
learning	-1.318390e-15	-1.535102e-01	1.221245e-15	5.473979e-01	8.226726e-01	3.885781e-16	-5.551115e-17

在进行特征值选取的情况下，假设我们选取了 3 个特征值

我们的 U 矩阵变为

	0	1	2
NLP	-3.471975e-01	0.000000e+00	-1.942887e-01
flying	-1.396622e-01	0.000000e+00	6.739877e-01
enjoy	1.249001e-16	-3.687707e-01	-2.220446e-16
I	-8.340573e-01	-5.277047e-18	2.854104e-01
deep	-4.053355e-01	0.000000e+00	-6.530953e-01
like	1.665335e-16	-9.167567e-01	-2.775558e-17
learning	-1.318390e-15	-1.535102e-01	1.221245e-15

这也是我们基于共现矩阵 SVD 之后得到的词向量矩阵

每一行代表了选定词的词向量