奇异值分解（SVD）

漫漫成长
算法工程师

以下内容来自刘建平Pinard-博客园的学习笔记，总结如下：

奇异值分解(Singular Value Decomposition，以下简称SVD)是在机器学习领域广泛应用的算法，它不光可以用于降维算法中的特征分解，还可以用于推荐系统，以及自然语言处理等领域。是很多机器学习算法的基石。本文就对SVD的原理做一个总结，并讨论在在PCA降维算法中是如何运用运用SVD的。

1. 回顾特征值和特征向量

首先回顾下特征值和特征向量的定义如下：

$$Ax = \lambda x$$

其中 A 是一个 $n \times n$ 矩阵， x 是一个 n 维向量，则 λ 是矩阵 A 的一个特征值，而 x 是矩阵 A 的特征值 λ 所对应的特征向量。

求出特征值和特征向量有什么好处呢？ 就是我们可以将矩阵A特征分解。如果我们求出了矩阵A的n个特征值 $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ ，以及这 n 个特征值所对应的特征向量 $w_1, w_2, ..., w_n$ ，

那么矩阵A就可以用下式的特征分解表示：

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

其中W是这n个特征向量所张成的n×n维矩阵，而Σ为这n个特征值为主对角线的n×n维矩阵。

一般我们会把W的这n个特征向量标准化，即满足 $||w_i||_2 = 1$ ，或者 $w_i^T w_i = 1$ ，此时W的

n个特征向量为标准正交基，满足 $W^T W = I$ ，即 $W^T = W^{-1}$ ，也就是说W为酉矩阵。

这样我们的特征分解表达式可以写成

$$A = W\Sigma W^T$$

注意到要进行特征分解，矩阵A必须为方阵。

那么如果A不是方阵，即行和列不相同时，我们还可以对矩阵进行分解吗？答案是可以，此时我们的SVD登场了。

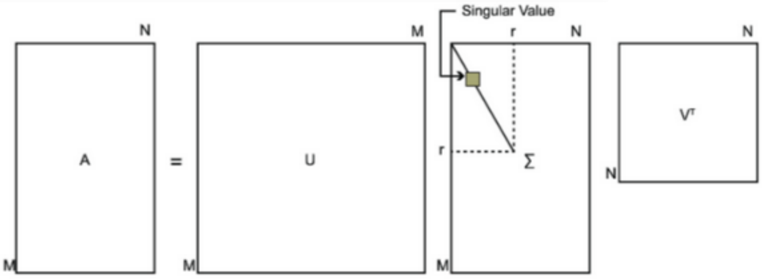
2. SVD的定义

SVD也是对矩阵进行分解，但是和特征分解不同，SVD并不要求要分解的矩阵为方阵。假设我们的矩阵A是一个m×n的矩阵，那么我们定义矩阵A的SVD为：

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

其中 U 是一个 $m \times m$ 的矩阵， Σ 是一个 $m \times n$ 的矩阵，除了主对角线上的元素以外全为0，主对角线上的每个元素都称为奇异值， V 是一个 $n \times n$ 的矩阵。 U 和 V 都是酉矩阵，即满足

$U^T U = I, V^T V = I$ 。下图可以很形象的看出上面SVD的定义：



那么我们如何求出SVD分解后的U,Σ,V这三个矩阵呢?

如果我们将A的转置和A做矩阵乘法，那么会得到n×n的一个方阵 A^TA 。既然 A^TA 是方阵，那么我们就可以进行特征分解，得到的特征值和特征向量满足下式:

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

这样我们就可以得到矩阵 A^TA 的n个特征值和对应的n个特征向量v了。将 A^TA 的所有特征向量张成一个n×n的矩阵V，就是我们SVD公式里面的V矩阵了。一般我们将V中的每个特征向量叫做A的右奇异向量。

如果我们将A和A的转置做矩阵乘法，那么会得到m×m的一个方阵 AA^T 。既然 AA^T 是方阵，那么我们就可以进行特征分解，得到的特征值和特征向量满足下式:

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

这样我们就可以得到矩阵 AA^T 的m个特征值和对应的m个特征向量u了。将 AA^T 的所有特征向量张成一个m×m的矩阵U，就是我们SVD公式里面的U矩阵了。一般我们将U中的每个特征向量叫做A的左奇异向量。

U和V我们都求出来了，现在就剩下奇异值矩阵Σ没有求出了.

由于Σ除了对角线上是奇异值其他位置都是0，那我们只需要求出每个奇异值σ就可以了。

我们注意到:

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

这样我们可以求出我们的每个奇异值，进而求出奇异值矩阵Σ。

上面还有一个问题没有讲，就是我们说 A^TA 的特征向量组成的就是我们SVD中的V矩阵，而

AA^T 的特征向量组成的就是我们SVD中的U矩阵，这有什么根据吗？这个其实很容易证明，我们以V矩阵的证明为例。

$$A = U\Sigma V^T \Rightarrow A^T = V\Sigma U^T \Rightarrow A^TA = V\Sigma U^TU\Sigma V^T = V\Sigma^2V^T$$

上式证明使用了 $U^U = I, \Sigma^T = \Sigma$ 。可以看出 A^TA 的特征向量组成的的确就是我们SVD中的V矩阵。类似的方法可以得到 AA^T 的特征向量组成的就是我们SVD中的U矩阵。

进一步我们还可以看出我们的特征值矩阵等于奇异值矩阵的平方，也就是说特征值和奇异值满足如下关系:

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

这样也就是说，我们可以不用 $\sigma_i = \frac{Av_i}{u_i}$ 来计算奇异值，也可以通过求出 A^TA 的特征值取平方根来求奇异值。

这里我们用一个简单的例子来说明矩阵是如何进行奇异值分解的。我们的矩阵A定义为：

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

首先求出 A^TA 和 AA^T

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

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

进而求出 A^TA 的特征值和特征向量：

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

接着求出 AA^T 的特征值和特征向量：

$$\lambda_1=3;u_1=\begin{pmatrix}1/\sqrt{6}\\2/\sqrt{6}\\1/\sqrt{6}\end{pmatrix};\lambda_2=1;u_2=\begin{pmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{pmatrix};\lambda_3=0;u_3=\begin{pmatrix}1/\sqrt{3}\\-1/\sqrt{3}\\1/\sqrt{3}\end{pmatrix}$$

利用 $Av_i=\sigma_iu_i,i=1,2$ 求奇异值：

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

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

也可以用 $\sigma_i=\sqrt{\lambda_i}$ 直接求出奇异值为 $\sqrt{3}$ 和1.

最终得到A的奇异值分解为：

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

4. SVD的一些性质

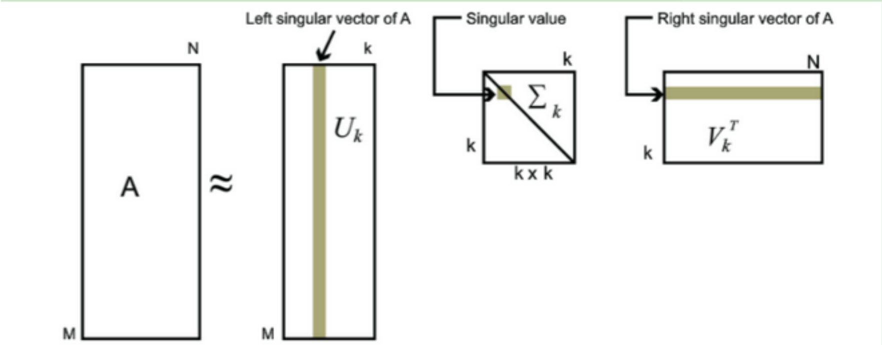
对于奇异值,它跟我们特征分解中的特征值类似，在奇异值矩阵中也是按照从大到小排列，而且奇异值的减少特别的快，在很多情况下，前10%甚至1%的奇异值的和就占了全部的奇异值之和的99%以上的比例。

也就是说，我们也可以用最大的k个的奇异值和对应的左右奇异向量来近似描述矩阵。

也就是说：

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T\approx U_{m\times k}\Sigma_{k\times k}V_{k\times n}^T$$

其中k要比n小很多，也就是一个大的矩阵A可以用三个小的矩阵 $U_{m\times k},\sum_{k\times k},V_{k\times n}^T$ 来表示。如下图所示，现在我们的矩阵A只需要灰色的部分的三个小矩阵就可以近似描述了。



由于这个重要的性质，SVD可以用于PCA降维，来做数据压缩和去噪。也可以用于推荐算法，将用户和喜好对应的矩阵做特征分解，进而得到隐含的用户需求来做推荐。同时也可以用于NLP中的算法，比如潜在语义索引（LSI）。

下面我们就对SVD用于PCA降维做一个介绍。

5. SVD用于PCA

PCA降维，需要找到样本协方差矩阵 X^TX 的最大的d个特征向量，然后用这最大的d个特征向量张成的矩阵来做低维投影降维。可以看出，在这个过程中需要先求出协方差矩阵 X^TX ，当样本数多样本特征数也多的时候，这个计算量是很大的。

注意到我们的SVD也可以得到协方差矩阵 X^TX 最大的d个特征向量张成的矩阵，但是SVD有个好处，有一些SVD的实现算法可以不求先求出协方差矩阵 X^TX ，也能求出我们的右奇异矩阵V。也就是说，我们的PCA算法可以不用做特征分解，而是做SVD来完成。这个方法在样本量很大的时候很有效。实际上，scikit-learn的PCA算法的背后真正的实现就是用的SVD，而不是我们我们认为的暴力特征分解。

另一方面，注意到PCA仅仅使用了我们SVD的右奇异矩阵，没有使用左奇异矩阵，那么左奇异矩阵有什么用呢?

假设我们的样本是m×n的矩阵X，如果我们通过SVD找到了矩阵 XX^T 最大的d个特征向量张成的m×d维矩阵U，则我们如果进行如下处理：

$$X'_{d\times n} = U^T_{d\times m} X_{m\times n}$$

可以得到一个d×n的矩阵X ‘,这个矩阵和我们原来的m×n维样本矩阵X相比，行数从m减到了k，可见对行数进行了压缩。

左奇异矩阵可以用于行数的压缩。

右奇异矩阵可以用于列数即特征维度的压缩，也就是我们的PCA降维。

6. SVD小结

SVD作为一个很基本的算法，在很多机器学习算法中都有它的身影，特别是在现在的大数据时代，由于SVD可以实现并行化，因此更是大展身手。

SVD的缺点是**分解出的矩阵解释性往往不强**，有点黑盒子的味道，不过这不影响它的使用。

编辑于 2017-10-12 12:38

数据降维 机器学习

写下你的评论...

182 条评论

默认 最新

刘思黎

试着解释一下SVD矩阵：在我看来，SVD是对数据进行有效特征整理的过程。首先，对于一个m×n矩阵A，我们可以理解为其有m个数据，n个特征，（想象成一个n个特征组成的坐标系中的m个点），然而一般情况下，这n个特征并不是正交的，也就是说这n个特征并不能归纳这个数据集的特征。SVD的作用就相当于是一个坐标系变换的过程，从一个不标准的n维坐标系，转换为一个标准的k维坐标系，并且使这个数据集中的点，到这个新坐标系的欧式距离为最小值（也就是这些点在这个新坐标系中的投影方差最大化），其实就是一个最小二乘的过程。进一步，如何使数据在新坐标系中的投影最大化呢，那么我们就需要让这个新坐标系中的基尽可能的不相关，我们可以用协方差来衡量这种相关性。A^T·A中计算的便是n×n的协方差矩阵，每一个值代表着原来的n个特征之间的相关性。当对这个协方差矩阵进行特征分解之后，我们可以得到奇异值和右奇异矩阵，而这个右奇异矩阵则是一个新的坐标系，奇异值则对应这个新坐标系中每个基对于整体数据的影响大小，我们这时便可以提取奇异值最大的k个基，作为新的坐标，这便是PCA的原理。

2020-01-10 · 热评 回复 408

究极无敌超大水果 ▸ **周论**

右奇异矩阵应该是固定点的数量，压缩冗余的特征，左奇异矩阵是固定这么多特征，去掉冗余的点。

2021-04-07 回复 21

周论

左奇异矩阵是不是就是把m看作特征，n看作数据的个数，原理是一样的对吧，只不过是样本数据的不同表示方法😄

2021-03-11 回复 8

查看全部 26 条回复 >

00木水

公式评论里面挂了，这里详细描述下，就是2.SVD的定义里的A转置A那里的证明，一个是U乘U转置为单位阵，这里公式打错，一个是中间的矩阵sigma是m*n维度，它的转置是n*m，并不相等，但它们相乘是对角方阵。

2018-01-20 回复 65

张亿锋

我也发现了，你说的是对的

2018-02-14 回复 3

王同学

我就说有问题。

2021-07-13 回复 1

展开其他 2 条回复 >

舒适星球

这个先算U V，再算中间矩阵的方法是错的！！！浪费我一个下午加晚上排bug！！！气死我了！！！（应该先算u或v再求中间矩阵推出v或u，按这种算法特征向量正负都无法确定）希望懂的人赞上去 别让这个文章害人了!

2021-05-23 回复 55

HK416

兄啊，这是一个推导过程，来解释这些左右矩阵的含义的。而且也提到了，实际现在的算法并不是通过算协方差矩阵的特征值，因为这样非常慢。实际上MATLAB，numpy里面有大量算法可以直接输出SVD的左右矩阵以及奇异值，速度也很可观......你要做算法或者码程序，应该要参考讲算法的文章，而不是这种。

2021-06-01 回复 15

舒适星球 ▸ **HK416**

这算法错的没毛病吧 🤔大学生要做编程训练的 坑害了我是事实🤣

2021-06-01 回复 3

查看全部 11 条回复 >

知乎用户9YJDJs

1里面A必须为实对称矩阵，实对称矩阵才有正交基让其对角化。一般的n * n矩阵，比如Jordan block是没法对角化的

2018-04-07 回复 46

2018-12-20
回复 5

论分析
这里的A默认是实对称阵或复正规阵吧🤔
2019-12-13
回复 3

展开其他 1 条回复 >

曲面参数化
分解出的矩阵解释性往往不强，深有体会。
2017-10-09
回复 33

Jeffery ‣ **JunMa**
我一直不懂为什么分解出来的那三个矩阵有这样的物理含义。。
2017-11-08
回复 11

让我缓一缓
本质是基变换造成的"点"的坐标表示方式不同，想要解释清楚，还是得理解矩阵论的思想。我认为基变换是本质，点（向量，数据）的坐标的变化是外在体现
2021-10-09
回复 4

查看全部 6 条回复 >

helloworld
UTU=I,ΣTΣ=Σ2 这里写错了
2019-07-07
回复 19

天明
就在找你
2021-01-06
回复 3

秋离隽
很棒的文章，让我这个傻子都看懂了SVD
2020-03-08
回复 8

忍冬之声
我这个傻子还没看懂 是不是没救了
2021-05-18
回复 3

知乎用户7RHVBo
那为啥SVD这么定义,你懂吗?
2021-09-23
回复 2

展开其他 1 条回复 >

知乎用户0H6wvp
解释了怎么做,没有解释为什么这么做
2019-03-04
回复 8

YoungHojan ‣ **见证骑士的时刻**
你口气挺大😅
2022-01-06
回复 8

见证骑士的时刻
解释了你能听懂吗
2021-12-13
回复 赞

我天吾啊
包括这篇文章，都说使用SVD的优势是，不要求必须是方阵，可以不用计算协方差的特征值。可是你们在讲怎么计算SVD那两个矩阵的时候都是用协方差的特征值计算的，看了很多文章都搞不懂这点，而且PCA计算协方差矩阵特征值也不要求是方阵
2018-11-06
回复 8

刘亮 ‣ **计算机没前途**
我来回应一下吧，这里提的计算方法只是为了讲清楚这三个矩阵的含义，并且方便给出证明这么分解是对的。具体这么通过不计算协方差的特征值来求解 SVD 是有很多后续研究的，不是一篇简单的科普教程能讲清楚的，不求甚解的话，就记住高级的算法可以绕过这一步直接得到三个矩阵，如果实在感兴趣的话可以看一下 NIPS 2007 的文章 Fast

回复 44

计算机没前途
这些人参考来参考去，也不正面回应一下
2020-02-06 回复 1

查看全部 7 条回复

probe
写的很清楚了，感谢。但有一个问题，第1节里，一般n*n矩阵的特征向量不一定正交，不是因为特征向量标准化后就是正交基了。实对称矩阵的不同特征值对应的特征向量必正交，比如PCA里的协方差矩阵就是实对称矩阵。还有一个疑问，
AA^T
和
A^TA
的特征值数量不同啊，一个是m个，一个是n个，难道只取其中相同的就行了吗？
2018-11-02 回复 7

刘思黎
他们的非零特征值是相同的。简单证明：A^T·Ax = λ·x，同时左乘A可以得到，(A·A^T)·(Ax) = λ·(Ax)
2020-01-10 回复 14

美国懂王 ‣ **刘思黎**
不是Ax = λ·x？ A^T·Ax = λ·x怎么来的？
2020-08-04 回复 1

查看全部 8 条回复

壮壮
我们的PCA算法可以不用做特征分解，而是做SVD来完成 - 不懂，svd具体是怎么求v和特征值的呢？具体算法说明能推荐一下吗？谢谢
2017-11-27 回复 6

网络罗宾汉
不好意思，我理解错了，作者这里的X是m*d的，m是样本数，d是单个样本向量的维数，和西瓜书是反着的
2019-11-26 回复 4

流年
说的对啊！！！
2021-03-23 回复 赞

seezen
也能求出我们的右奇异矩阵但是SVD有个好处，有一些SVD的实现算法可以不求先求出协方差矩阵 ，也能求出我们的右奇异矩阵V。 有哪些算法。。
2018-09-13 回复 1

Dosea
看了这么多SVD概念，这个看懂了，谢谢🙏
2017-11-16 回复 7

白玉汤
哪些SVD的算法可以不求先求出协方差矩阵
X^TX，也能求出我们的右奇异矩阵？
2018-06-04 回复 6

caa
power method或者贪心算法，前者是估计值，后者可槽性太难但准确
2019-06-23 回复 2

Alvin he ‣ **caa**
这也能抓到你?
2022-01-21 回复 赞

李的摸鱼日记
脑成像一个很普遍的算法里面用过SVD，用于去噪音。
2017-10-09 回复 6

Hello王叔叔

2019-03-29 回复 7

chaselighter
您好，可以推荐一下相关的这篇文章吗？谢谢您
02-20 回复 赞

展开其他 1 条回复 >

我不坏
写的是不错，但在矩阵表示的时候与常识不一样，导致让人困惑，设X为数据，通常为d*n维，其中d为特征维，n为数量维，讲解中刚好相反。
2018-07-21 回复 5

luxcgo
”一般我们会把W的这n个特征向量标准化，即满足 $||w_i||_2 = 1$ ，或者 $w_i^T w_i = 1$ ，此时W的n个特征向量为标准正交基“
是不是只有实对称矩阵才能做这种标准正交的操作，不是任意的矩阵都有这种性质吧？

2019-02-21 回复 3

在NLP的路上了 ‣ **美国懂王**
特征根可以相同，只要满足（具有k重特征值λ，必有k个线性无关的特征向量）就可以了
2021-03-19 回复 1

美国懂王 ‣ **在NLP的路上了**
我明白你说的意思。但我试过几次，特征根相同的话，最后乘起来好像不是最初的矩阵
2021-03-19 回复 赞

展开其他 2 条回复 >

小钳子
我的算法也用到特征值，这个计算量真的是。。。。需要并行处理
2017-10-10 回复 3

木耳机
感谢 🙏
2018-01-07 回复 2

点击查看全部评论 >

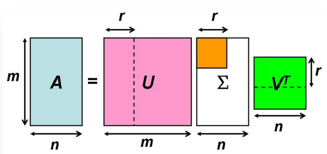
写下你的评论...

文章被以下专栏收录

控制系统设计

推荐阅读

奇异值分解（SVD）
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...



奇异值分解（SVD）

奇异值分解(SVD)
欢迎关注我们的微信公众号“人工智能LeadAI”（ID：atleadai）最近两天都在看奇异值分解及其在推荐系统和图像压缩方面的应用，这部分知识比较散也比较难理解，看代码不是很好懂，所以通过编...
工智能L... 发表于人工智能L...

23、奇异值分解
本章内容包含了：S 的通用计算方法SVI SVD常遇见的疑问在 OpenCV求解SVD的 用SVD的应用场景S 统、协同过滤和相似
曦微