[1、对称矩阵（Symmetric Matrices） 2](#_Toc239466360)

[2、方形矩阵 2](#_Toc842348550)

[3、单位矩阵 2](#_Toc1262132045)

[4、矩阵的特征值 3](#_Toc799266598)

[5、海森矩阵（Hessian matrix） 4](#_Toc1840889602)

[6、正定矩阵 4](#_Toc314739062)

[7、样本中心化和标准化 4](#_Toc1982525936)

[8、协方差矩阵 7](#_Toc1187906276)

[10、矩阵的奇异值 7](#_Toc1057973062)

[11、奇异值分解 8](#_Toc1432149727)

[12、酉矩阵 14](#_Toc423750082)

矩阵有关

# 1、对称矩阵（Symmetric Matrices）

对称矩阵是指元素以[主对角线](https://baike.baidu.com/item/%E4%B8%BB%E5%AF%B9%E8%A7%92%E7%BA%BF" \t "/home/260158/文档\\x/_blank)为对称轴，对应相等的矩阵。 在[线性](https://baike.baidu.com/item/%E7%BA%BF%E6%80%A7" \t "/home/260158/文档\\x/_blank)代数中，对称矩阵是一个方形矩阵，其转置矩阵和自身相等。

**--->性质：**

**--->应用：**

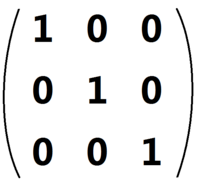
# 2、方形矩阵

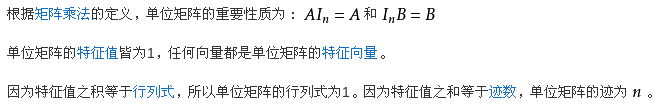
**--->**行，列数相等

2017-12-11 16-29-28 的屏幕截图

# 3、单位矩阵

在矩阵的乘法中，有一种矩阵起着特殊的作用，如同数的乘法中的1，这种矩阵被称为单位矩阵。它是个方阵，从左上角到右下角的对角线（称为主对角线）上的元素均为1。除此以外全都为0。



**性质：  
**

**应用：**

高等代数中，在求解相应的矩阵时若添加单位矩阵然后通过初等变换进行求解往往可以使问题变得简单。

# 4、矩阵的特征值

设 A 是n阶方阵，如果存在数m和非零n维列向量 x，使得 Ax=mx 成立，则称 m 是矩阵A的一个特征值（characteristic value)或本征值（eigenvalue)。非零n维列向量x称为矩阵A的属于（对应于）特征值m的特征向量或本征向量，简称A的特征向量或A的本征向量。

式Ax=λx也可写成( A-λE)X=0。这是n个未知数n个方程的齐次线性方程组，它有非零解的充分必要条件是系数行列式| A-λE|=0。

# 5、海森矩阵（Hessian matrix）

# 6、正定矩阵

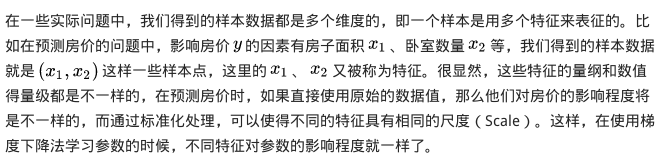
# 7、样本中心化和标准化

在回归问题和一些机器学习算法中，以及训练神经网络的过程中，通常需要对原始数据进行中心化（Zero-centered或者Mean-subtraction）处理和标准化（Standardization或Normalization）处理。

**目的：**通过中心化和标准化处理，得到均值为0，标准差为1的服从标准正态分布的数据。

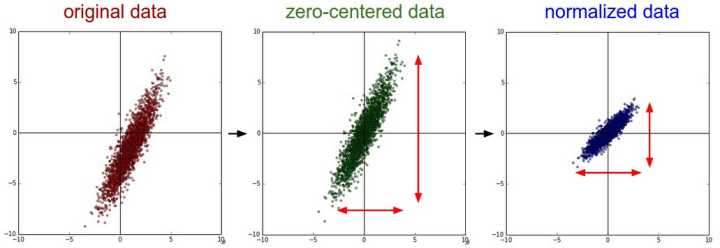
计算过程由下式表示：2017-12-20 09-18-43 的屏幕截图

下面解释一下为什么需要使用这些数据预处理步骤。



---->简言之，当原始数据不同维度上的特征的尺度（单位）不一致时，需要标准化步骤对数据进行预处理。

---->下图中以二维数据为例：左图表示的是原始数据；中间的是中心化后的数据，数据被移动大原点周围；右图将中心化后的数据除以标准差，得到为标准化的数据，可以看出每个维度上的尺度是一致的（红色线段的长度表示尺度）。

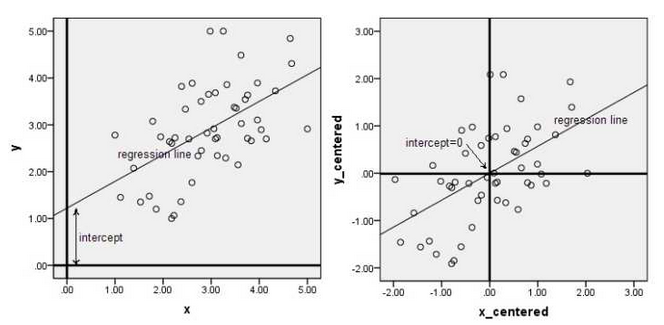


其实，在不同的问题中，中心化和标准化有着不同的意义，

---->比如在训练神经网络的过程中，通过将数据标准化，能够加速权重参数的收敛。

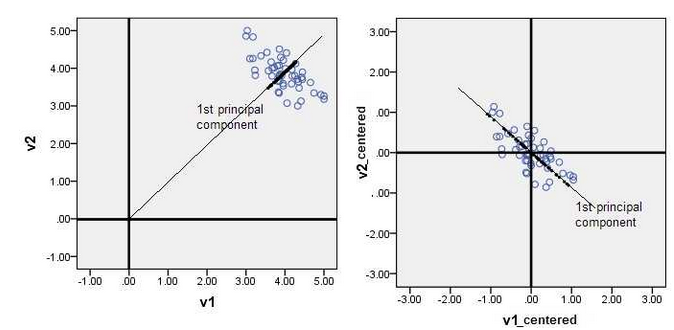
---->另外，对于主成分分析（PCA）问题，也需要对数据进行中心化和标准化等预处理步骤。

下面两幅图是数据做中心化（centering）前后的对比，可以看到其实就是一个平移的过程，平移后所有数据的中心是（0，0）



在做PCA的时候，我们需要找出矩阵的特征向量，也就是主成分（PC）。比如说找到的第一个特征向量是a = [1, 2]，a在坐标平面上就是从原点出发到点（1，2）的一个向量。

如果没有对数据做中心化，那算出来的第一主成分的方向可能就不是一个可以“描述”（或者说“概括”）数据的方向了。还是看图比较清楚。



黑色线就是第一主成分的方向。只有中心化数据之后，计算得到的方向才能比较好的“概括”原来的数据。

# 8、协方差矩阵

9、奇异矩阵  
 奇异矩阵是线性代数的概念，就是如果一个矩阵对应的行列式等于0，则该矩阵称为奇异矩阵。

如何判断一个矩阵是否是奇异阵呢？

（1）看这个矩阵是不是方阵（即行数和列数相等的矩阵。若行数和 列数不相等，那就谈不上奇异矩阵和非奇异矩阵）。

（2）看此方阵的行列式|A|是否等于0，若等于0，称矩阵A为奇异 矩阵；若不等于0，称矩阵A为非奇异矩阵。

（3）由|A|≠0可知矩阵A可逆，可以得出另外一个重要结论：逆矩 阵就是非奇异矩阵，非奇异矩阵也是可逆矩阵。如果A为奇异 矩阵，则AX=0有无穷解，AX=b有无穷解或者无解。如果A为非 奇异矩阵，则AX=0有且只有唯一零解，AX=b有唯一解。

（4）如果A(n×n)为奇异矩阵<=>A的秩Rank(A)<n.

如果A(n×n)为非奇异矩阵<=> A满秩，Rank(A)=n.

# 10、矩阵的奇异值

设A是一个mXn矩阵，称正半定矩阵A‘A的特征值的非负平方根为矩阵A的奇异值，其中A‘表示矩阵A的共扼转置矩阵.

# 11、奇异值分解

　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

我们首先回顾下特征值和特征向量的定义如下：

Ax=λx

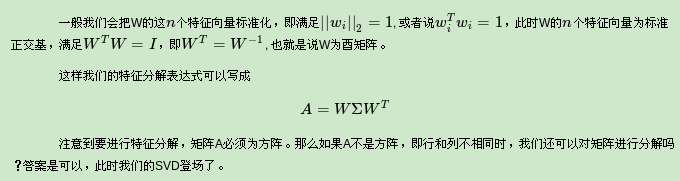
　其中A是一个n×n的矩阵，x是一个n维向量，则我们说λ是矩阵A的一个特征值，而x是矩阵A的特征值λ所对应的特征向量。

　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λ1≤λ2≤...≤λn,以及这n

个特征值所对应的特征向量{w1,w2,...wn}，那么矩阵A就可以用下式的特征分解表示：

2018-03-21 10-19-05 的屏幕截图

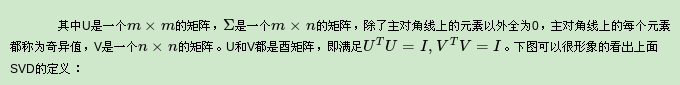
其中W是这n个特征向量所组成的n×n维矩阵，而Σ为这n个特征值为主对角线的n×n维矩阵。

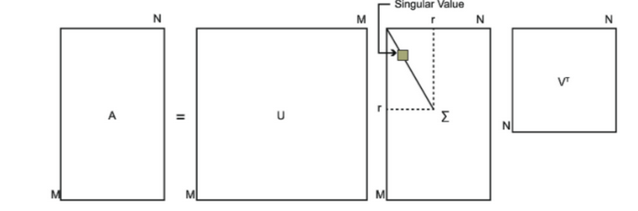


（2）SVD的定义

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：

2018-03-21 10-26-31 的屏幕截图

****

****

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ATA。既然ATA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：

**2018-03-21 10-30-27 的屏幕截图**

这样我们就可以得到矩阵ATA的n个特征值和对应的n个特征向量v了。将ATA的所有特征向量组成一个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。

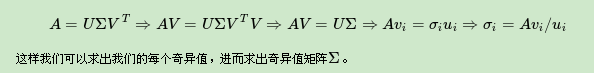
如果我们将A和A的转置做矩阵乘法，那么会得到m×m

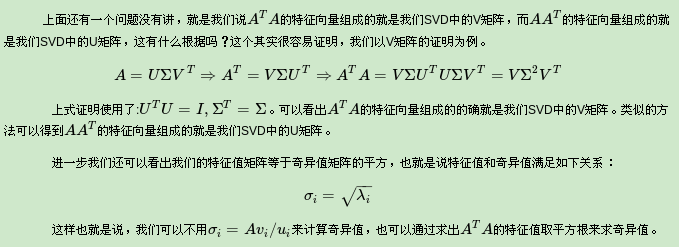
的一个方阵AAT。既然AAT是方阵，那么我们就可以进行特征分解，得到的特征值和特征向量满足下式：

2018-03-21 10-33-06 的屏幕截图

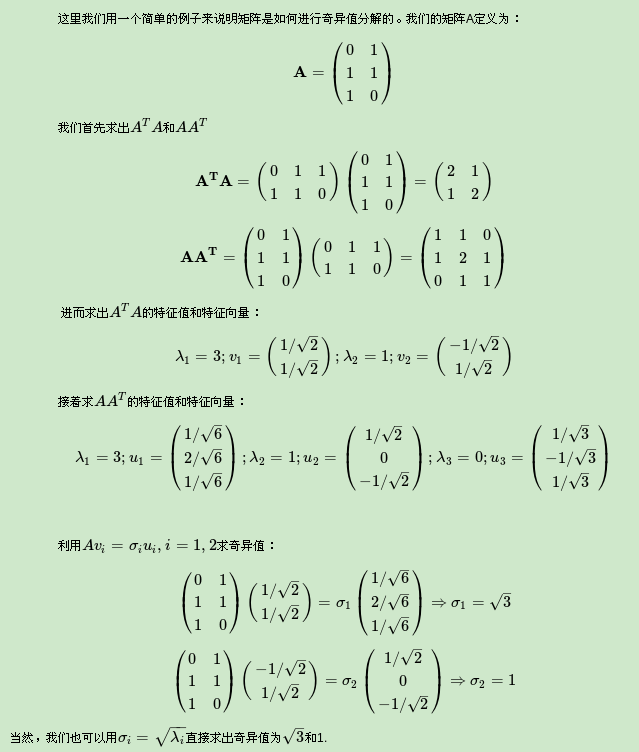
这样我们就可以得到矩阵AAT的m个特征值和对应的m个特征向量u了。将AAT的所有特征向量张成一个m×m的矩阵U，就是我们SVD公式里面的U矩阵了。一般我们将U中的每个特征向量叫做A的左奇异向量。

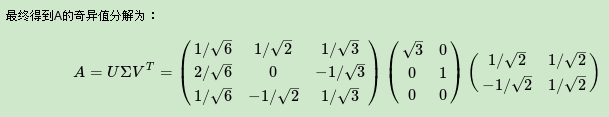
U和V我们都求出来了，现在就剩下奇异值矩阵Σ没有求出了。由于Σ除了对角线上是奇异值其他位置都是0，那我们只需要求出每个奇异值σ就可以了。





（3）SVD计算举例



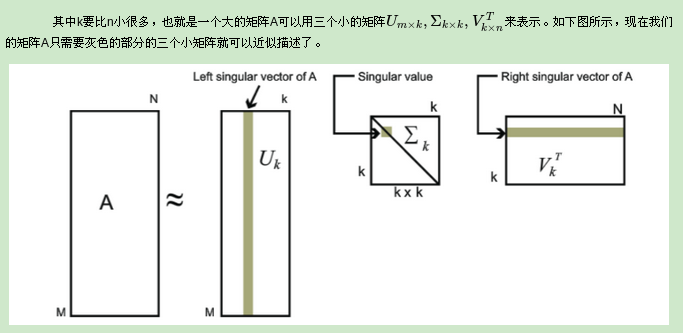


（4）SVD的一些性质

上面几节我们对SVD的定义和计算做了详细的描述，似乎看不出我们费这么大的力气做SVD有什么好处。那么SVD有什么重要的性质值得我们注意呢？

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：

2018-03-21 11-27-18 的屏幕截图



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。

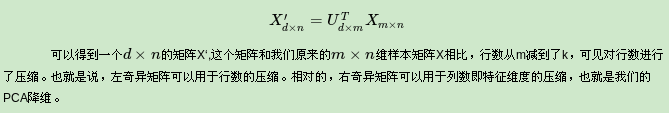
（5）SVD用于PCA

在主成分分析（PCA）原理总结中，我们讲到要用PCA降维，需要找到样本协方差矩阵XTX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XTX，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XTX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XTX，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XXT最大的d个特征向量组成的m×d维矩阵U，则我们如果进行如下处理：



<https://www.zhihu.com/question/22237507> 奇异值的物理意义？

# 12、酉矩阵