**函数矩阵：**

对x求导即矩阵每个元素均对x进行求导。

双重导数直接是双重求导。

纯量函数对矩阵的求导：相当于对矩阵每一个元素求导。

同样有

函数矩阵对矩阵变量的导数

若列向量对行向量，结果是一个矩阵

Y=AX

矩阵对矩阵求导会导致维度扩宽。

列向量对横向量求导会扩宽为矩阵。

**向量和矩阵范数：**从范数可以导出向量与向量、矩阵与矩阵之间的距离。

范数满足：（1）正定性（2）正齐次性（3）三角不等式

三种向量长度度量方式（1）欧式长度（2）用最长的一边的长度（3）用两边长度之和

其中L0范数指向量中非零元素的个数，L1范数：向量中每个元素绝对值的和，L2范数：向量元素绝对值的平方和再开平方。

在机器学习中： L2正则化通过权重衰减，保证了模型的简单，提高了泛化能力。

L0范数指的是向量中非零元素的个数，L0正则化就是限制非零元素的个数在一定的范围，这很明显会带来稀疏。

**相似矩阵：**  
原理，特征值。

相似不变量：行列式、特征多项式、特征值、迹

1）特征值：

    如果说一个向量v是方阵A的特征向量，将一定可以表示成下面的形式：

[image](http://images.cnblogs.com/cnblogs_com/LeftNotEasy/201101/201101192226321862.png)

    这时候λ就被称为特征向量v对应的特征值，一个矩阵的一组特征向量是一组正交向量。特征值分解是将一个矩阵分解成下面的形式：

[image](http://images.cnblogs.com/cnblogs_com/LeftNotEasy/201101/201101192226321023.png)

特征值分解可以得到特征值与特征向量，特征值表示的是这个特征到底有多重要，而特征向量表示这个特征是什么，可以将每一个特征向量理解为一个线性的子空间，我们可以利用这些线性的子空间干很多的事情。不过，特征值分解也有很多的局限，比如说变换的矩阵必须是方阵。

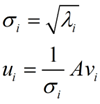
2）奇异值：

    下面谈谈奇异值分解。特征值分解是一个提取矩阵特征很不错的方法，但是它只是对方阵而言的，在现实的世界中，我们看到的大部分矩阵都不是方阵，比如说有N个学生，每个学生有M科成绩，这样形成的一个N \* M的矩阵就不可能是方阵，我们怎样才能描述这样普通的矩阵呢的重要特征呢？奇异值分解可以用来干这个事情，奇异值分解是一个能适用于任意的矩阵的一种分解的方法：

    假设A是一个N \* M的矩阵，那么得到的U是一个**N \* N的方阵**（里面的向量是正交的，**U里面的向量称为左奇异向量**），Σ是一个**N \* M的矩阵**（除了对角线的元素都是0，**对角线上的元素称为奇异值**），**V’(V的转置)是一个N \* N的矩阵**，里面的向量也是正交的，**V里面的向量称为右奇异向量**），从图片来反映几个相乘的矩阵的大小可得下面的图片

[image](http://images.cnblogs.com/cnblogs_com/LeftNotEasy/201101/201101192226335092.png)

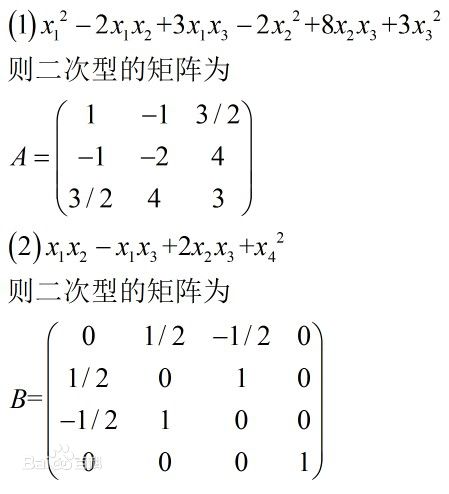
那么奇异值和特征值是怎么对应起来的呢？首先，我们将一个矩阵A的转置 \* A，将会得到一个方阵，我们用这个方阵求特征值可以得到：[image](http://images.cnblogs.com/cnblogs_com/LeftNotEasy/201101/201101192226349618.png)    这里得到的v，就是我们上面的右奇异向量。此外我们还可以得到：

[](http://images.cnblogs.com/cnblogs_com/LeftNotEasy/201101/201101192226344111.png)    这里的σ就是上面说的奇异值，u就是上面说的左奇异向量。

若n阶方阵A的行列式不为零，即 |A|≠0，则称A为非奇异矩阵或满秩矩阵，否则称A为奇异矩阵或降秩矩阵。

一个矩阵半正定当且仅当它的每个特征值大于或等于零。

**二次型和二次曲面：**



二次型：f(X)=XTAX

A与B合同：存在可逆矩阵C,B=CTAC

N元二次型的规范型，标准型。

正交变换化实二次型为标准型。

正定二次型，半正定二次型

f(X)=XTAX >= 0

特征值非负