0 范数、1 范数、2 范数有什么区别?

0 范数、1 范数、2 范数有什么区别?

魏通
A lucky person ~

谢 @邱乾方 指出错误，已改正。

以下分别列举常用的向量范数和矩阵范数的定义。

- **向量范数**

1-范数：

$||x||_1 = \sum_{i=1}^{N} |x_i|$，即向量元素绝对值之和，matlab调用函数norm(x, 1) 。

2-范数：

$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$，Euclid范数（欧几里得范数，常用计算向量长度），即向量元素绝对值的平方和再开方，matlab调用函数norm(x, 2)。

∞ -范数：$||\mathbf{x}||_\infty = \max_i |x_i|$，即所有向量元素绝对值中的最大值，matlab调用函数norm(x, inf)。

$-\infty$ -范数：$||\mathbf{x}||_{-\infty} = \min_i |x_i|$

，即所有向量元素绝对值中的最小值，matlab调用函数norm(x, -inf)。

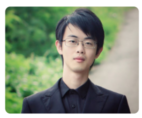
p-范数：$||\mathbf{x}||_p = (\sum_{i=1}^{N} |x_i|^p)^{\frac{1}{p}}$

，即向量元素绝对值的p次方和的1/p次幂，matlab调用函数norm(x, p)。

- **矩阵范数**

1-范数：$||A||_1 = \max_j \sum_{i=1}^{m} |a_{i,j}|$

，列和范数，即所有矩阵列向量绝对值之和的最大值，matlab调用函数norm(A, 1)。



2-范数：$||A||_2=\sqrt{\lambda_1}$， $\lambda < br/ >$ 为 A^TA 的最大特征值。

，谱范数，即A'A矩阵的最大特征值的开平方。matlab调用函数norm(x, 2)。

∞-范数：$||A||_\infty = \max \sum_{j=1}^{N} |a_{ij}|$

，行和范数，即所有矩阵行向量绝对值之和的最大值，matlab调用函数norm(A, inf)。

F-范数：$||A||_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2\right)^{\frac{1}{2}}$

，Frobenius范数，即矩阵元素绝对值的平方和再开平方，matlab调用函数norm(A, 'fro')。

核范数：$||A||_*=\sum_{i=1}^{n}\lambda_i, \lambda_i$ 是A的奇异值。

即奇异值之和。

编辑于 2016-06-16

凌空
渐行渐远

114 人赞同了该回答

各类范数从机器学习的角度会比较好理解一些，所以我从这个角度说一下，本人为初学者，如有错误，还请不吝指正。

我们从最简单的最小二乘线性模型开始。最开始，最小二乘的loss(需优化的目标函数)如下：

$$E_D(\mathbf{w})=\frac{1}{2}\sum_{n=1}^{N}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$

式中，tn是目标变量，xn是观测量（自变量），\phi 是基函数（后期推导与核化相关），是w是参数。此式有闭式解，解为：

$$\mathbf{w}_{\mathrm{ML}}=\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}$$

但是我们都知道，矩阵求逆是一个病态问题，即矩阵并不是在所有情况下都有逆矩阵。所以上述

式子在实际使用时会遇到问题。为了解决这个问题，可以求其近似解。可以用SGD(梯度下降法)求一个近似解，或者加入正则项（2范数）。加入正则项是我们这里要说的。加入2范数的正则项可以解决这个病态问题，并且也可以得到闭式解，在实际使用时要比用SGD快，并且加入正则化后的好处并不仅仅是这些。加入正则项（2范数）的loss如下：

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

其闭式解为：

$$\mathbf{w} = \left(\lambda\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}.$$

此式在 \lambda 不为零时，总是有解的，所以是一个非病态的问题，这在实际使用时很好。除了这一点，2范数的正则项还有其他好处，比如控制方差和偏差的关系，得到一个好的拟合，这里就不赘述了，毕竟这里讲的是范数，有兴趣可以参阅相关资料。

加入正则项后一般情况下的loss为：

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$$

好了，我们终于可以专注于范数了。不同范数对应的曲线如下图：

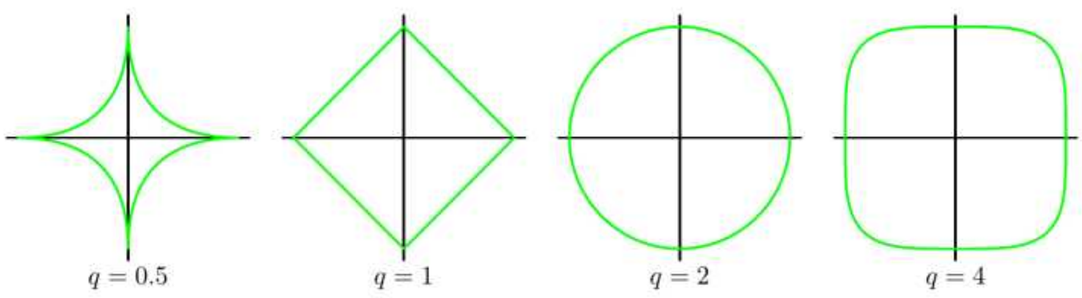
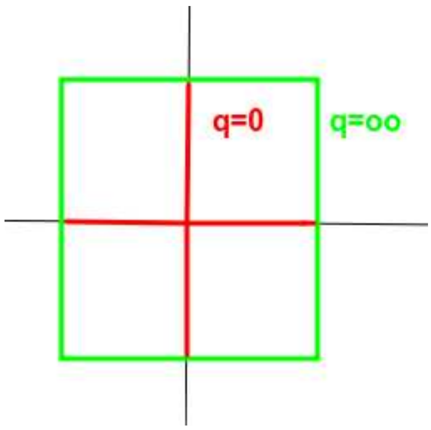


Figure 3.3 Contours of the regularization term in (3.29) for various values of the parameter q.

（图来源于参考书PRML）

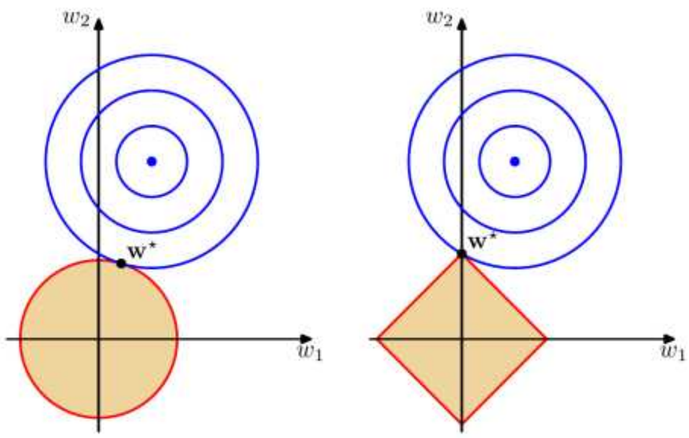
上图中，可以明显看到一个趋势，即q越小，曲线越贴近坐标轴，q越大，曲线越远离坐标轴，并且棱角越明显。那么 q=0 和 q=oo 时极限情况如何呢？猜猜看。



bingio, 你猜对了，答案就是十字架和正方形。除了图形上的直观形象，在数学公式的推导中，q=0 和 q=oo 时两种极限的行为可以简记为非零元的个数和最大项。即0范数对应向量或矩阵中非零元的个数，无穷范数对应向量或矩阵中最大的元素。具体推导可以参考维基百科。至此为止，那么他们用在机器学习里有什么区别呢？

以1范数和2范数为例：

Figure 3.4 Plot of the contours of the unregularized error function (blue) along with the constraint region (3.30) for the quadratic regularizer $q = 2$ on the left and the lasso regularizer $q = 1$ on the right, in which the optimum value for the parameter vector $\mathbf{w}$ is denoted by $\mathbf{w}^\star$. The lasso gives a sparse solution in which $w_1^\star = 0$.



（图来自PRML）

上图中，蓝色的圆圈表示原问题可能的解范围，橘色的表示正则项可能的解范围。而整个目标函数（原问题+正则项）有解当且仅当两个解范围相切。从上图可以很容易地看出，由于2范数解范围是圆，所以相切的点有很大可能不在坐标轴上（感谢评论区＠临熙指出表述错误），而由于1范数是菱形（顶点是凸出来的），其相切的点更可能在坐标轴上，而坐标轴上的点有一个特点，其

只有一个坐标分量不为零，其他坐标分量为零，即是稀疏的。所以有如下结论，1范数可以导致稀疏解，2范数导致稠密解。那么为什么不用0范数呢，理论上它是求稀疏解最好的规范项了。然而在机器学习中，特征的维度往往很大，解0范数又是NP-hard问题，所以在实际中不可行。但是用1范数解是可行的，并且也可以得到稀疏解，所以实际稀疏模型中用1范数约束。

至此，我们总结一下，在机器学习中，以0范数和1范数作为正则项，可以求得稀疏解，但是0范数的求解是NP-hard问题; 以2范数作为正则项可以得到稠密解，并且由于其良好的性质，其解的定义很好，往往可以得到闭式解，所以用的很多。

另外，从距离的角度说一下范数。1范数对应街区距离，2范数对应大家熟知的欧式距离，无穷范数对应棋盘距离（切比雪夫距离）。

(1) 欧氏（Euclidean）距离

$$D_E(p,q) = [(x-s)^2 + (y-t)^2]^{1/2}$$

(2) 城区（city-block）距离

$$D_4(p,q) = |x-s| + |y-t|$$

(3) 棋盘（chessboard）距离

$$D_8(p,q) = \max(|x-s|, |y-t|)$$

			3			
	2.8	2.2	2	2.2	2.8	
	2.2	1.4	1	1.4	2.2	
3	2	1	0	1	2	3
	2.2	1.4	1	1.4	2.2	
	2.8	2.2	2	2.2	2.8	
			3			

(a)

D_E

			3			
		3	2	3		
	3	2	1	2	3	
3	2	1	0	1	2	3
	3	2	1	2	3	
		3	2	3		
			3			

(b)

D_4

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	0	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

(c)

D_8 43

参考资料:

1. PRML: Pattern recognition and machine learning, Christopher M. Bishop.

2. USTC 图像分析课程ppt，统计学习课程ppt.

编辑于 2017-10-13

▲ 114 ▼ 8 条评论 分享 收藏 感谢 ··· 收起

JI Weiwei

327 人赞同了该回答

你是问向量范数还是矩阵范数?

要更好的理解范数，就要从函数、几何与矩阵的角度去理解，我尽量讲的通俗一些。

我们都知道，函数与几何图形往往是有对应的关系，这个很好想象，特别是在三维以下的空间内，函数是几何图像的数学概括，而几何图像是函数的高度形象化，比如一个函数对应几何空间上若干点组成的图形。

但当函数与几何超出三维空间时，就难以获得较好的想象，于是就有了映射的概念，映射表达的就是一个集合通过某种关系转为另外一个集合。通常数学书是先说映射，然后再讨论函数，这是因为函数是映射的一个特例。

为了更好的在数学上表达这种映射关系，（这里特指线性关系）于是就引进了矩阵。这里的矩阵就是表征上述空间映射的线性关系。而通过向量来表示上述映射中所说的这个集合，而我们通常所说的基，就是这个集合的最一般关系。于是，我们可以这样理解，一个集合（向量），通过一种映射关系（矩阵），得到另外一个几何（另外一个向量）。

那么向量的范数，就是表示这个原有集合的大小。

而矩阵的范数，就是表示这个变化过程的大小的一个度量。

那么说到具体几几范数，其不过是定义不同，一个矩阵范数往往由一个向量范数引出，我们称之为算子范数，其物理意义都如我上述所述。

以上符合知乎回答问题的方式。

接下来用百度回答方式:

0范数，向量中非零元素的个数。

1范数，为绝对值之和。

2范数，就是通常意义上的模。

发布于 2014-05-14

▲ 327 ▼ 18 条评论 分享 收藏 感谢 ···

柯均堂

14 人赞同了该回答

几何意义不考虑，单从分析角度。

总而言之都是P范数，零范数即是当p趋于零，可以证明这时候的极限 $(x_1^p+x_2^p+\cdots+x_n^p)^{\frac{1}{p}}$ 恰好是向量 $x=(x_1,x_2,\cdots,x_n)^T$ 非零元素的个数。而无穷范数则是当 p 趋于无穷的时候的范数。

关于这些极限的证明，数学分析的开篇就介绍了，我想这也是为什么当初学习数学分析的时候为什么老是有各种奇形怪状的极限要证明存在性的原因之一。

但严格来说，零范数并不满足数乘率(i.e. $|\alpha x|\neq|\alpha|\,|x|$)。之所以最近被多次提到，也是因为压缩感知的火热

发布于 2016-09-13

▲ 14 ▼ 1 条评论 分享 收藏 感谢 ···

孙满
微信公众号建设中:掌门狗说

8 人赞同了该回答

再引申一点吧，如果有理解不到位的情况，还请各位前辈指正。

0、1、2范数通常会和ML里的各个模型的正则化联系起来。一般我们训练的模型总是想既能很好地描绘训练数据，又不至于过拟合，因此我们希望这个模型又不会很复杂(过拟合，习得模型参数过多)，出现不fit验证数据的情况。因此需要选择既能够描绘训练数据（对应经验风险→小），又能够尽量不那么复杂的模型（对应模型参数简单→正则化项也小）。

而正则化项通常使用**模型参数向量***w*的0、1、2范式表示。

- L0 w分量尽量稀疏 如 (0,a,0,0,b,0,0)
- L1 效果同上
- L2 w分量取值尽量均衡、稠密，即小而趋近于0 如(0.3,0.5,-0.3,0.1,-0.2,0.3,-0.3)

有些详细的讨论移步评论区~

编辑于 2017-11-02

▲ 8 ▼ 6 条评论 分享 收藏 感谢 ···

匿名用户

7 人赞同了该回答

x 的 0 范数：x 到零点的汉明距离

x 的 1 范数：x 到零点的曼哈顿距离

x 的 2 范数：x 到零点的欧氏距离

...

x 的 n 范数：x 到零点的 n 阶闵氏距离

x 的无穷范数：x 到零点的切比雪夫距离

编辑于 2017-06-22

▲ 7 ▼ 添加评论 分享 收藏 感谢 ···

陈俊
学生

5 人赞同了该回答

刚学了矩阵论。。。其实也不知道正不正确。就是我是类比了以前学过的平均数概念，平均数不是有几何平均数，算术平均数巴拉巴拉一堆么，但是都是用来度量一个整体的平均水平，只不过是用在了不同的地方。
范数我觉得也是这样，什么一范数二范数也是用来度量一个整体，比如两个个班的人比较高度，你可以用班里面最高的人（向量无穷范数）去比较，也可以用班里所有人的身高总和比较（向量

一范数），也可以求平均（几何平均？忘记了。。）（类似向量二范数）。
所以我认为范数就是一种度量方式。至于矩阵范数和向量范数，你可以把向量范数认为是长度为一的矩阵范数的特例，然后由于经常用，所以人们就单独把向量范数拿出来专门探究他的性质（当然，在推导过程里面一般都是通过研究简单的，如向量，在扩展到困难的，如矩阵）
声明，以上都是我的主观臆测，很可能不对，如果错了请告诉我，谢啦～

发布于 2016-10-29

▲ 5 ▼ 1 条评论 分享 收藏 感谢 ···

Panfeng Li
Come here for knowledge

1 人赞同了该回答

“为了更好的在数学上表达这种映射关系，（这里特指线性关系）于是就引进了矩阵。这里的矩阵就是表征上述空间映射的线性关系”，楼上所说的这个之前体会不深。
向量的零范数即为其稀疏度sparsity

编辑于 2016-02-25

▲ 1 ▼ 添加评论 分享 收藏 感谢 ···

写回答

3 个回答被折叠（为什么？）