

# 3 向量范数

欧几里得距离的延伸



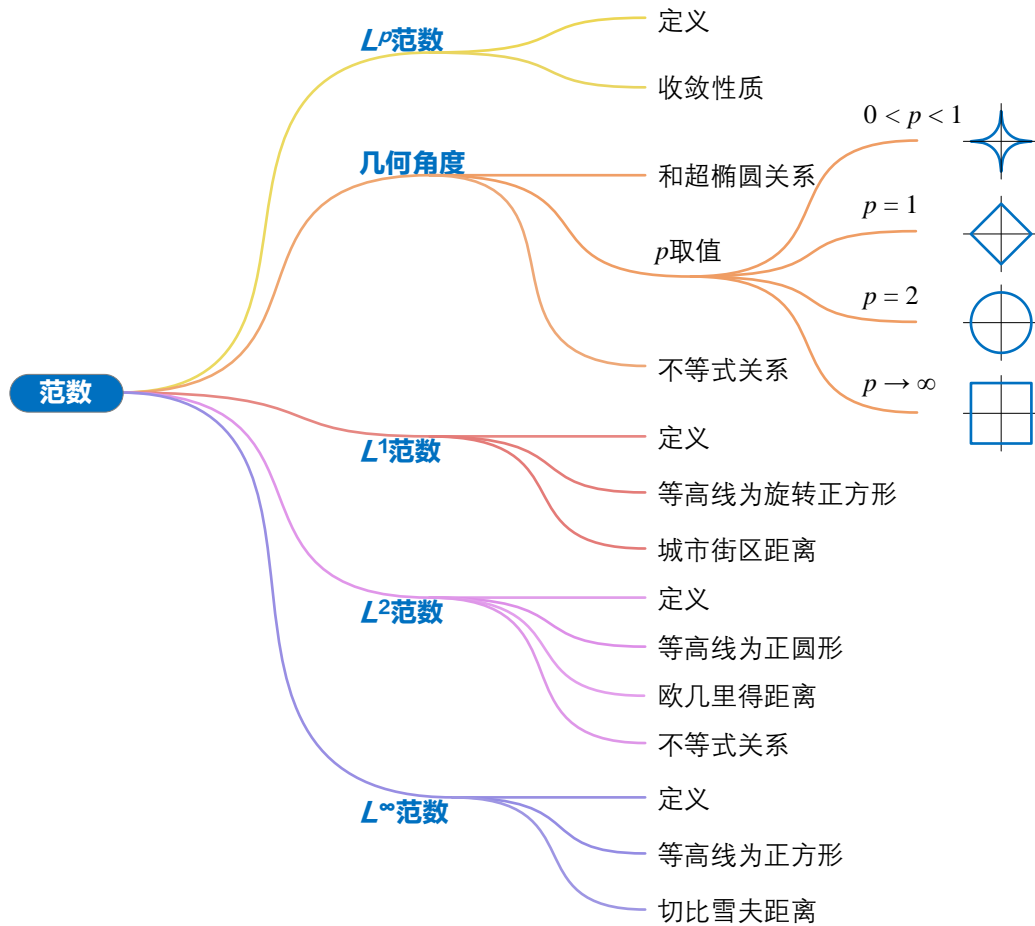
数学领域，遇到理解不了的概念别怕，用习惯就好了。

*In mathematics, you don't understand things. You just get used to them.*

—— 约翰·冯·诺伊曼 (Johann von Neumann) | 理论计算机科学与博弈论奠基者 | 1903 ~ 1957



- ▶ `matplotlib.pyplot.axhline()` 绘制水平线
- ▶ `matplotlib.pyplot.axvline()` 绘制竖直线
- ▶ `matplotlib.pyplot.contour()` 绘制等高线图
- ▶ `matplotlib.pyplot.contourf()` 绘制填充等高线图
- ▶ `numpy.abs()` 计算绝对值
- ▶ `numpy.linalg.norm()` 默认计算  $L^p$  范数
- ▶ `numpy.linspace()` 指定的间隔内返回均匀间隔数组
- ▶ `numpy.maximum()` 计算最大值
- ▶ `numpy.meshgrid()` 生成网格化数据



## 3.1 $L^p$ 范数: $L^2$ 范数的推广

上一章介绍了  $L^2$  范数,  $L^2$  范数代表向量的长度, 也叫向量的模, 等价于欧几里得距离。本章将  $L^2$  范数推广到  $L^p$  范数。

给定如下列向量  $\mathbf{x}$ :

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_D]^T \quad (1)$$

向量  $\mathbf{x}$  的  $L^p$  范数定义为:

$$\|\mathbf{x}\|_p = \left( |x_1|^p + |x_2|^p + \cdots + |x_D|^p \right)^{1/p} = \left( \sum_{j=1}^D |x_j|^p \right)^{1/p} \quad (2)$$

一般情况,  $p \geq 1$ ; 虽然  $p < 1$  时, 上式有定义, 但是不能称之为范数。容易判断,  $L^p$  范数非负, 即  $\|\mathbf{x}\|_p \geq 0$ 。  $L^p$  范数代表“距离”, 也是一种“向量  $\rightarrow$  标量”的运算规则。

**⚠ 注意**, (2) 中  $|x_j|$  计算  $x_j$  的绝对值。另外, 很多教材将  $L^p$  范数记做,  $L_p$  范数, 或  $p$ -范数。

当  $p = 2$  时, 向量  $\mathbf{x}$  的  $L^p$  范数便是  **$L^2$  范数** (L2-norm), 也叫 2-范数, 具体定义为:

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_D^2} = \left( \sum_{j=1}^D x_j^2 \right)^{\frac{1}{2}} \quad (3)$$

(3) 中  $\|\mathbf{x}\|_2$  的下角标常被省略, 也就是说  $\|\mathbf{x}\|$  默认为  $L^2$  范数。

特别地, 当  $p$  趋向  $+\infty$  时, 对应的范数记成  $L^\infty$ 。  $L^\infty$  范数定义为:

$$\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_D|) \quad (4)$$

即,  $\|\mathbf{x}\|_\infty$  为  $|x_i|$  中的最大值。

### 大小关系

举个例子, 如图 1 所示, 给定向量  $\mathbf{x}$ :

$$\mathbf{x} = [1 \quad 2 \quad 3]^T \quad (5)$$

向量  $\mathbf{x}$  的  $L^1$  范数是图 1 中三个坐标值的绝对值之和, 也就是图 1 长方体边长之和:

$$\|\mathbf{x}\|_1 = |1| + |2| + |3| = 6 \quad (6)$$

$L^2$  范数是图 1 向量  $\mathbf{x}$  的长度:

$$\|\mathbf{x}\|_2 = \left( |1|^2 + |2|^2 + |3|^2 \right)^{1/2} = (14)^{1/2} \approx 3.742 \quad (7)$$

向量  $\mathbf{x}$  的  $L^3$  范数可以通过下式求得：

$$\|\mathbf{x}\|_3 = \left(|1|^3 + |2|^3 + |3|^3\right)^{1/3} = 36^{1/3} \approx 3.302 \quad (8)$$

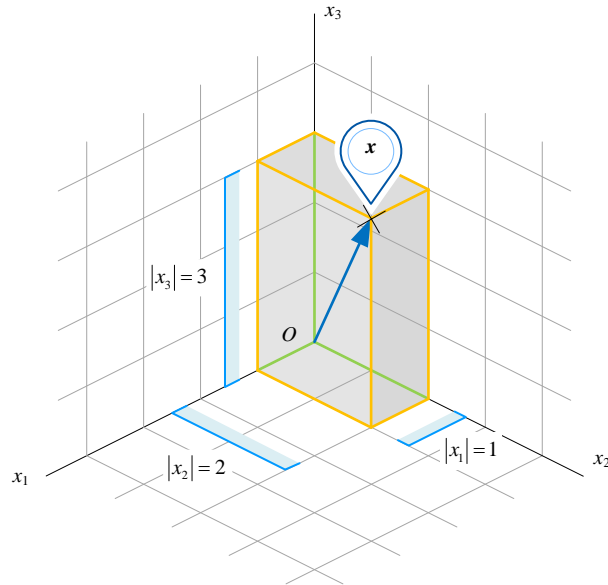


图 1. 向量  $\mathbf{x}$  在三维直角坐标系的位置

类似地，计算向量  $\mathbf{x}$  的  $L^4$  范数：

$$\|\mathbf{x}\|_4 = \left(|1|^4 + |2|^4 + |3|^4\right)^{1/4} = 98^{1/4} \approx 3.1463 \quad (9)$$

向量  $\mathbf{x}$  的  $L^\infty$  范数是图 1 中  $x_1$ 、 $x_2$ 、 $x_3$  三者绝对值中最大值：

$$\|\mathbf{x}\|_\infty = \max(|1|, |2|, |3|) = 3 \quad (10)$$

图 2 所示图像为  $L^p$  范数随  $p$  变化。对于  $\mathbf{x} = [1, 2, 3]^T$ ， $L^p$  范数随  $p$  增大而减小，最后收敛于 3。

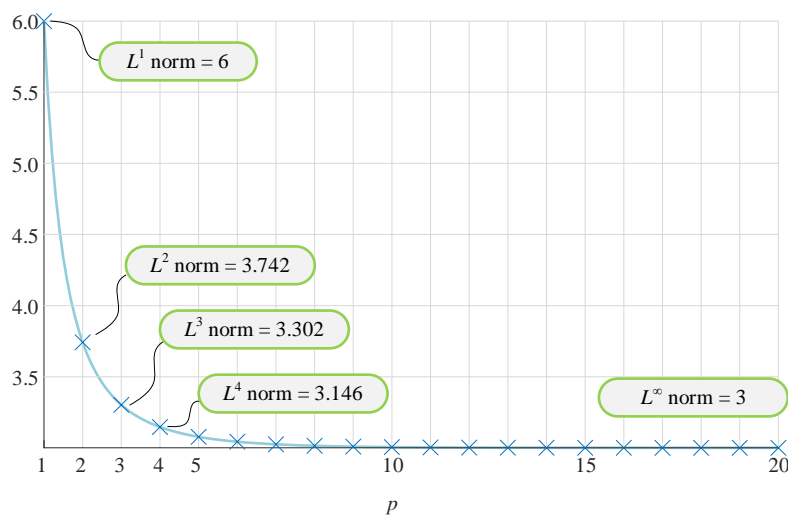


图 2.  $L^p$  范数随  $p$  变化

在数据科学、机器学习算法中， $L^p$  范数扮演重要角色，比如距离度量、正则化 (regularization)。下一节开始，我们就从几何图像入手，深入分析  $L^p$  范数性质。

## 3.2 $L^p$ 范数和超椭圆的联系

给定列向量  $\mathbf{x} = [x_1, x_2]^T$ ， $\mathbf{x}$  的  $L^p$  范数为：

$$\|\mathbf{x}\|_p = \left(|x_1|^p + |x_2|^p\right)^{1/p} \quad (11)$$

当  $p$  一定时，将 (11) 写成二元函数  $f(x_1, x_2)$ ：

$$f(x_1, x_2) = \left(|x_1|^p + |x_2|^p\right)^{1/p} \quad (12)$$

大家可能早已发现上式和《数学要素》一册讲过的超椭圆有着千丝万缕的联系。图 3 所示为  $p$  取不同值时， $f(x_1, x_2)$  函数对应曲面等高线变化。图中，暖色系代表函数  $f(x_1, x_2)$  更大数值，冷色系对应  $f(x_1, x_2)$  较小数值。

▲ 再次请大家特别注意， $0 < p < 1$  时，(11) 不能叫范数，因为不满足次可加。

$p = 1$  时， $f(x_1, x_2)$  函数的等高线为旋转正方形：

$$f(x_1, x_2) = |x_1| + |x_2| \quad (13)$$

$p = 2$  时， $f(x_1, x_2)$  函数等高线为正圆：

$$f(x_1, x_2) = \sqrt{x_1^2 + x_2^2} \quad (14)$$

$p = +\infty$  时， $f(x_1, x_2)$  函数等高线为正方形：

$$f(x_1, x_2) = \max(|x_1|, |x_2|) \quad (15)$$

如图 4 所示， $L^p$  范数取定值  $c$  时，即  $L^p = c$ ，随着  $p$  增大，等高线一层层包裹。

从相反角度，对于同一向量， $p$  增大， $L^p$  范数减小。请大家注意如下不等式关系：

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \quad (16)$$



Bk4\_Ch3\_01.py 绘制图 3 所示等高线。

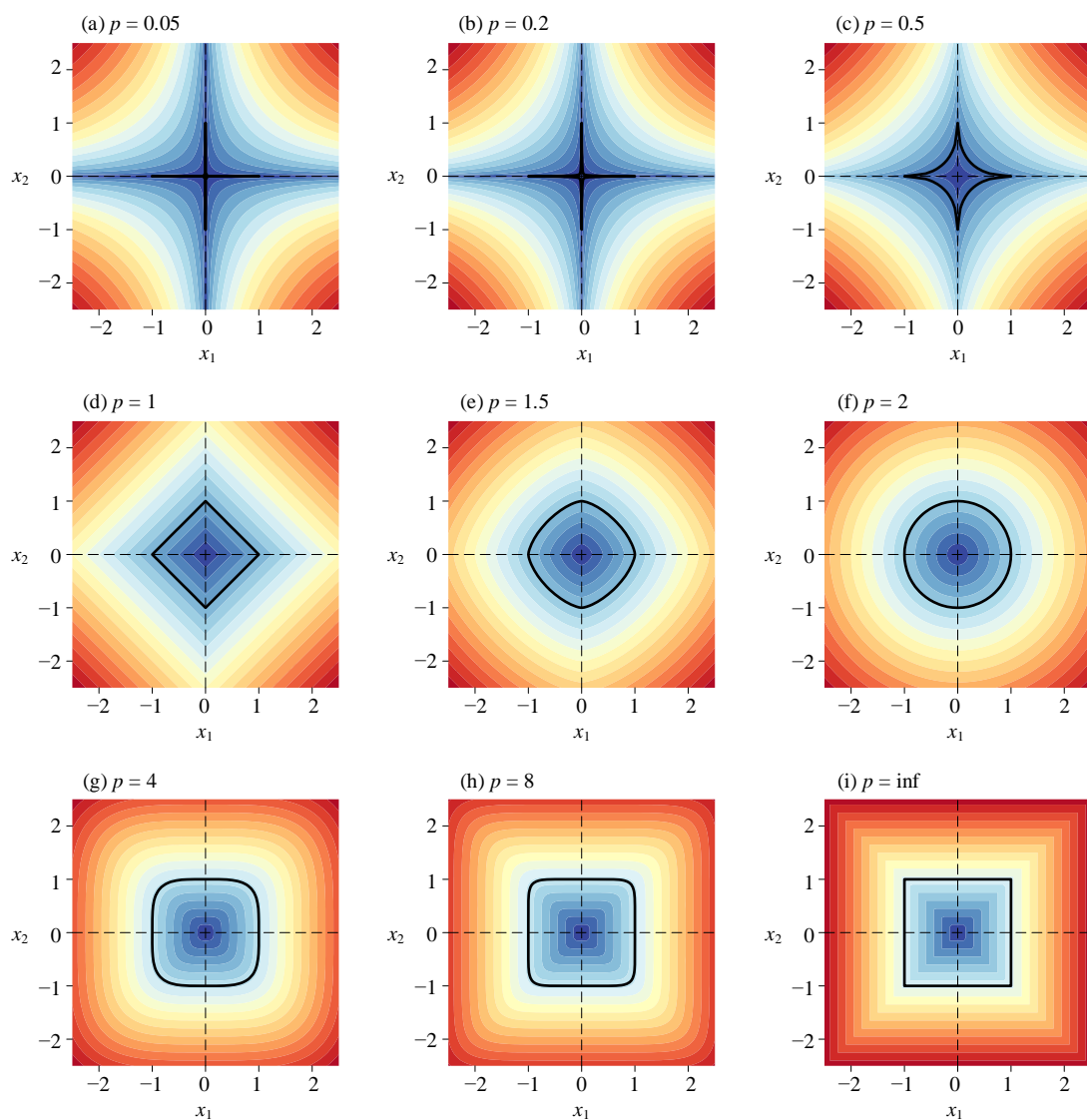
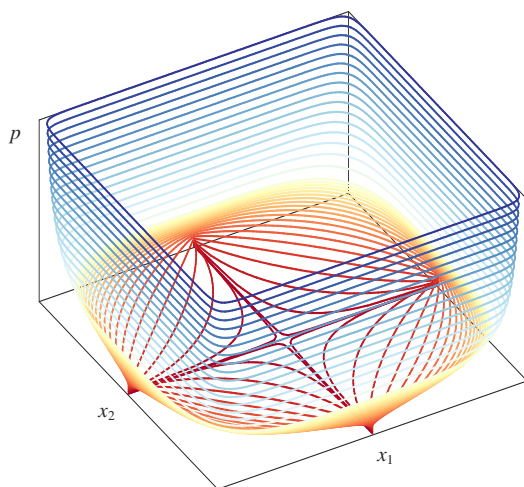
图 3.  $p$  取不同正数时, 二元函数等高线

图 4. 随着  $p$  增大, 等高线一层层包裹

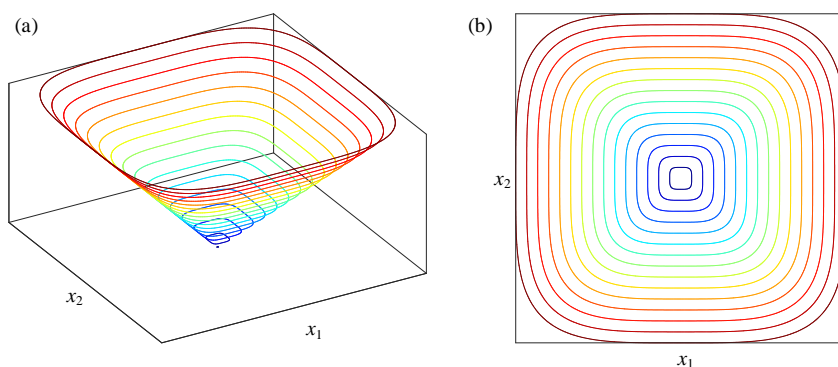
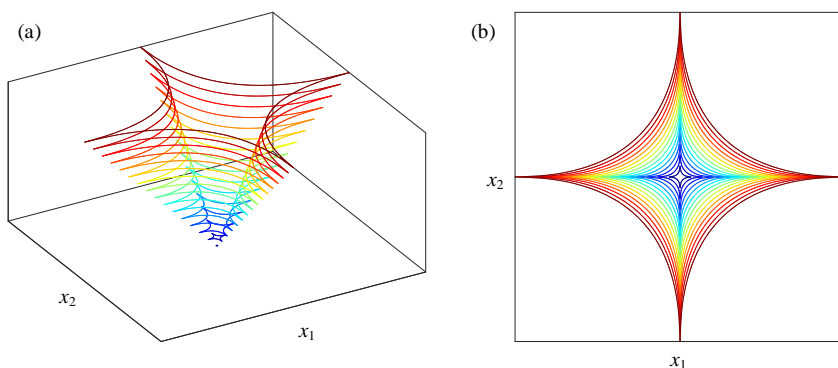
## 凸凹性

$p \geq 1$  时,  $L^p$  范数等高线形状为凸 (convex), 比如图 5。这是范数的一个重要性质——次可加性 (subadditivity), 也叫三角不等式 (triangle inequality):

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \quad (17)$$

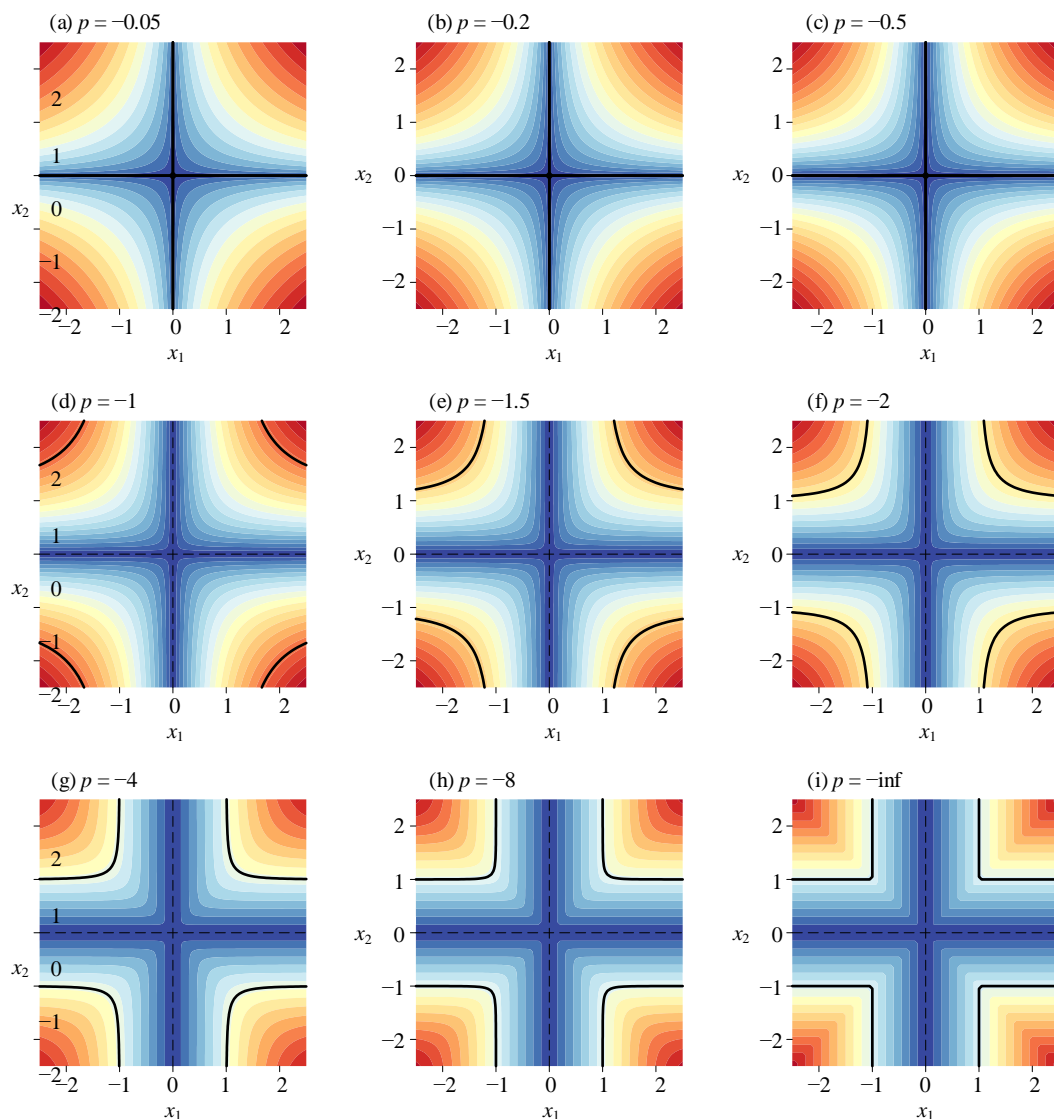
上式叫做闵可夫斯基不等式 (Minkowski inequality)。

$0 < p < 1$  时, (2) 对应等高线形状如图 6, 它非凸也非凹。严格来说,  $0 < p < 1$  时, (2) 虽然有定义, 但是不能称之为范数。这是因为,  $0 < p < 1$  时, (2) 不满足次可加性, 即违反三角不等式。

图 5.  $p = 4$ ,  $L^p$  范数等高线图像图 6.  $p = 0.5$ ,  $L^p$  范数等高线图像

## $p$ 为负数

$p$  取负数时, (12) 也有定义, 但是我们不能称之为范数。当图 7 所示为  $p$  取不同负数时, (12) 中函数等高线形状变化。

图 7.  $p$  取不同负数时，函数等高线变化

### 3.3 $L^1$ 范数：旋转正方形

本节探讨  $L^1$  范数几何特征。向量  $\mathbf{x}$  的  $L^1$  范数定义为：

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_D| = \sum_{j=1}^D |x_j| \quad (18)$$

当  $D = 2$  时，向量  $\mathbf{x}$  的  $L^1$  范数为：

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| \quad (19)$$

本 PDF 文件为作者草稿，发布目的为方便大家在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：[jiang.visualize.ml@gmail.com](mailto:jiang.visualize.ml@gmail.com)



(19) 中  $L^1$  范数等于 1 时，得到解析式：

$$|x_1| + |x_2| = 1 \quad (20)$$

下面，我分成几种情况展开 (20)，并绘制图像。

### 几何图形

观察 (20) 可以发现， $x_1$  和  $x_2$  的取值范围均为  $[-1, 1]$ ， $x_1$  和  $x_2$  符号可正可负。为了去掉绝对值符号，分四种情况考虑，得到如下展开式：

$$\begin{cases} x_1 + x_2 = 1 & 0 < x_1 < 1, 0 < x_2 < 1 \\ -x_1 + x_2 = 1 & -1 < x_1 < 0, 0 < x_2 < 1 \\ x_1 - x_2 = 1 & 0 < x_1 < 1, -1 < x_2 < 0 \\ -x_1 - x_2 = 1 & -1 < x_1 < 0, -1 < x_2 < 0 \end{cases} \quad (21)$$

根据 (21) 定义的四个一次函数解析式，可以得到图 8 所示图形。

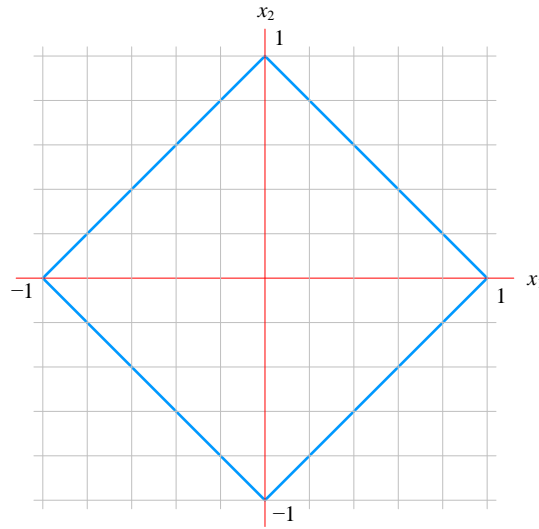
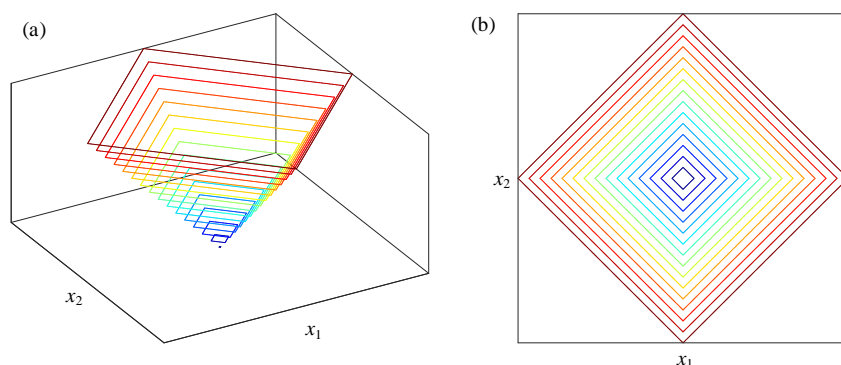


图 8.  $|x_1| + |x_2| = 1$  解析式图像

图 9 所示为如下函数的等高线图像：

$$f(x_1, x_2) = |x_1| + |x_2| \quad (22)$$

图 9 (b) 中每一条等高线上的点距离原点有相同的  $L^1$  范数。

图 9.  $p = 1$  时,  $L^p$  范数等高线图像

$L^1$  范数也叫**城市街区距离** (city block distance), 也称**曼哈顿距离** (Manhattan distance)。

如图 10 所示, 一个城市街区布局方方正正, 从  $A$  点到  $B$  点的行走距离不可能是两点的直线距离, 即欧氏距离。图中给出的行走路径类似  $L^1$  范数。

此外,  $L^1$  范数等高线存在“尖点”, 这个尖点将会在**套索回归** (LASSO regression) 的  $L^1$  正则项中起到重要作用。

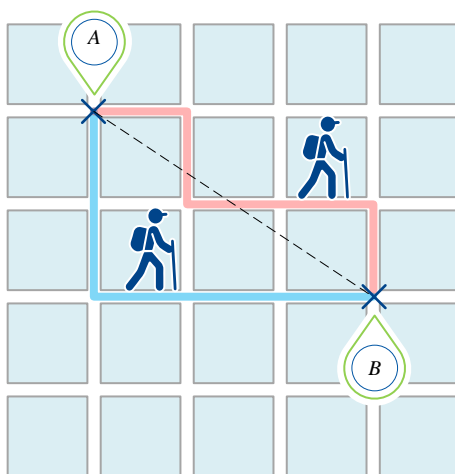


图 10. 城市街区距离

## 3.4 $L^2$ 范数: 正圆

本节探讨  $L^2$  范数形状。向量  $\mathbf{x}$  的  $L^2$  范数定义为:

$$\|\mathbf{x}\|_2 = (x_1^2 + x_2^2 + \cdots + x_D^2)^{1/2} = \left( \sum_{j=1}^D |x_j|^2 \right)^{1/2} \quad (23)$$

本 PDF 文件为作者草稿, 发布目的为方便大家在移动终端学习, 终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有, 请勿商用, 引用请注明出处。

代码及 PDF 文件下载: <https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教, 本书专属邮箱: [jiang.visualize.ml@gmail.com](mailto:jiang.visualize.ml@gmail.com)

特别地，当  $D=2$  时，向量  $\mathbf{x}$  的  $L^2$  范数为：

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2} \quad (24)$$

从距离度量角度， $L^2$  范数为欧几里得距离。

(24) 中  $L^2$  范数等于 1 时，对应如下平面图形解析式：

$$x_1^2 + x_2^2 = 1 \quad (25)$$

图 11 所示为 (25) 图像。

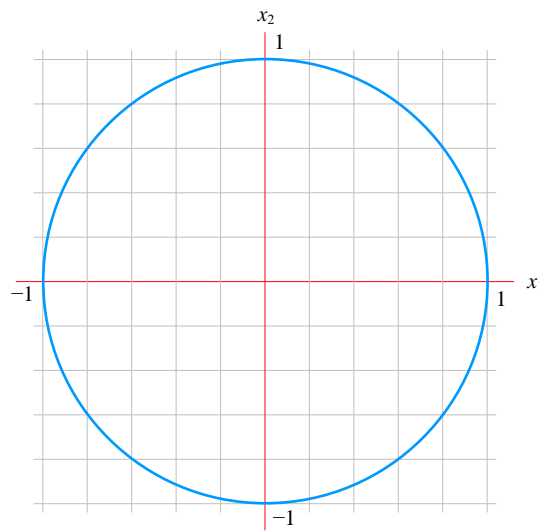


图 11.  $x_1^2 + x_2^2 = 1$  解析式图像

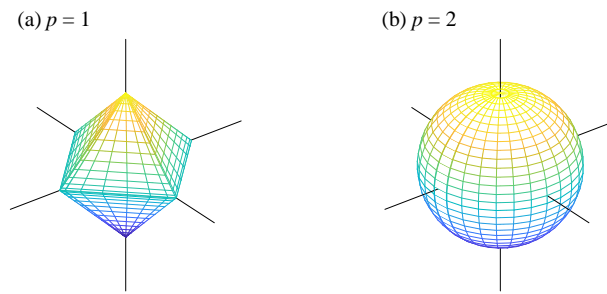
另外，实践中也经常使用  $L^2$  范数的平方，比如，

$$\|\mathbf{x}\|_2^2 = x_1^2 + x_2^2 \quad (26)$$

再次强调范数、向量内积、矩阵乘法关系，对于列向量  $\mathbf{x}$ ，以下运算等价，结果都是标量：

$$\|\mathbf{x}\|_2^2 = \|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x} \quad (27)$$

图 12 所示为当  $D=3$  时， $p$  分别取 1 和 2 时， $L^p$  范数对应的几何体。

图 12.  $p = 1, 2$ ,  $D = 3$  时,  $L^p$  范数对应的几何体

本系列丛书《数学要素》中简单讨论过向量范数在岭回归和套索回归的应用。岭回归引入的是  $L^2$  正则项，套索回归引入  $L^1$  正则项。

我们这里在介绍另外一种正则化回归——弹性网络回归 (elastic net regression)。弹性网络回归以不同比例同时引入  $L^1$  和  $L^2$  正则项。图 13 所示，正则化曲面是  $L^1$  和  $L^2$  范数曲面按不同比例叠加。图 13 中正则化部分既有  $L^1$  的“尖点”，也有  $L^2$  的凸曲面。

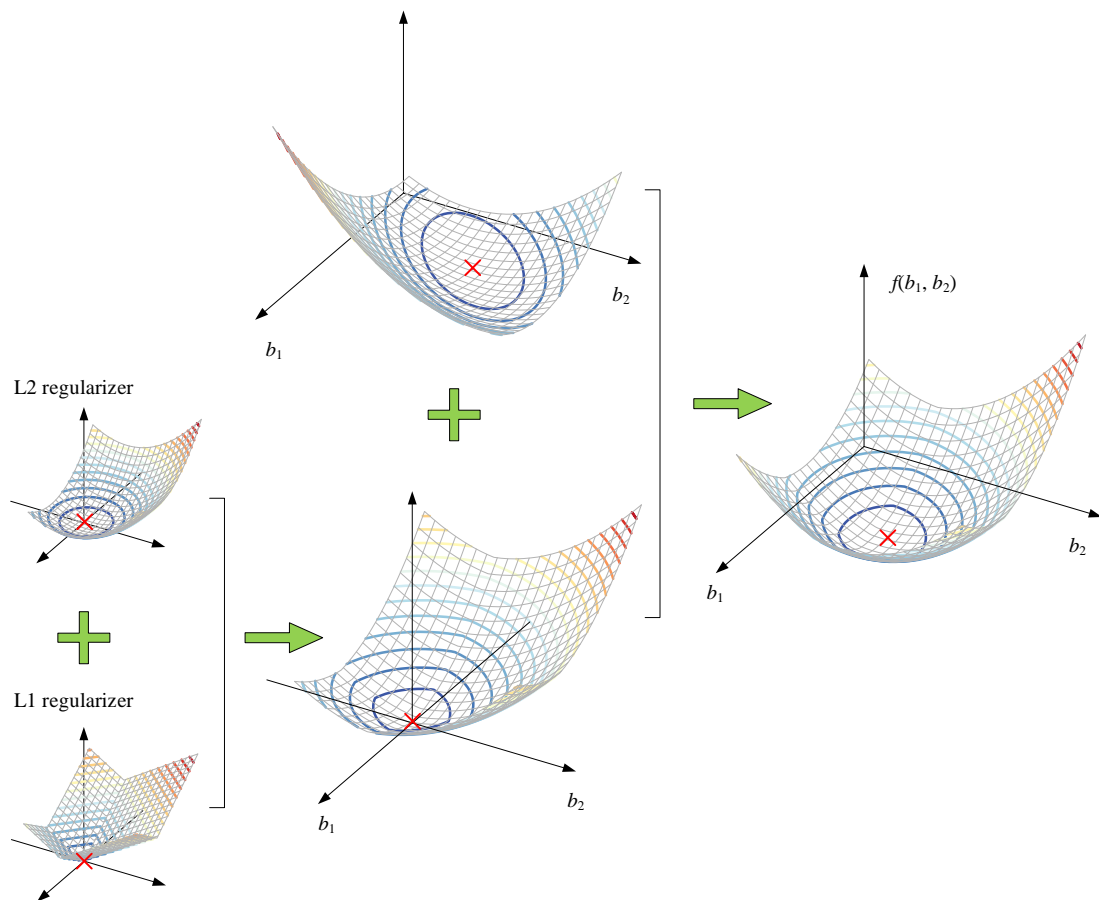


图 13. 弹性网络回归参数曲面

## 不等式

相信大家都知道，三角形两边之和大于第三边。应用到向量  $L^2$  范数，对应如下不等式：

$$\|u\|_2 + \|v\|_2 \geq \|u+v\|_2 \quad (28)$$

举个例子，给定向量  $u$  和  $v$ ：

$$u = [4 \ 3]^T, \quad v = [-2 \ 4]^T \quad (29)$$

向量  $u$  和  $v$  两者之和为：

$$u+v = [4 \ 3]^T + [-2 \ 4]^T = [2 \ 7]^T \quad (30)$$

图 14 所示为向量  $u$  和  $v$  以及  $u+v$  在平面上的关系。

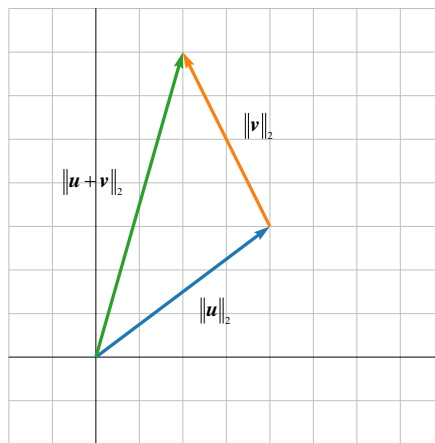


图 14. 向量  $u$  和  $v$  以及两者之和

$u$  和  $v$  的  $L^2$  范数分别为：

$$\|u\|_2 = \sqrt{4^2 + 3^2} = 5, \quad \|v\|_2 = \sqrt{(-2)^2 + 4^2} = \sqrt{20} \approx 4.4721 \quad (31)$$

$u$  和  $v$  的  $L^2$  范数和为：

$$\|u\|_2 + \|v\|_2 \approx 9.4721 \quad (32)$$

$u+v$  的  $L^2$  范数为：

$$\|u+v\|_2 = \sqrt{2^2 + 7^2} = \sqrt{54} \approx 7.2801 \quad (33)$$

显然，(28) 成立。请大家自行验证，满足  $p \geq 1$  时，当  $p$  取不同值时， $L^p$  范数都满足这种三角不等式关系。



Bk4\_Ch3\_02.py 绘制图 14 图 12。

## 3.5 $L^\infty$ 范数：正方形

向量  $\mathbf{x}$  的  $L^\infty$  范数的定义为：

$$\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_D|) \quad (34)$$

上式也叫做**切比雪夫距离** (Chebyshev distance)。

当特征数  $D = 2$  时，向量  $\mathbf{x}$  的  $L^\infty$  范数定义为：

$$\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|) \quad (35)$$

当  $L^\infty$  范数等于 1 时，可以得到如下平面图形解析式：

$$\max\{|x_1|, |x_2|\} = 1 \quad (36)$$

借助《数学要素》第 8、9 章讲解的圆锥曲线知识，我们一起推导 (36) 解析式对应的图像。

### 几何图形

观察 (36) 可以发现， $x_1$  和  $x_2$  的取值范围均为  $[-1, 1]$ ， $x_1$  和  $x_2$  符号可正、可负。分情况讨论，得到解析式：

$$\begin{cases} |x_1| = 1 & |x_1| > |x_2| \\ |x_2| = 1 & |x_2| > |x_1| \end{cases} \quad (37)$$

为了进一步展开 (37)，需要分析  $|x_1|$  和  $|x_2|$  大小关系。如果， $|x_1| > |x_2|$ ，不等式两边平方，并整理得到：

$$x_1^2 - x_2^2 > 0 \quad (38)$$

当把大于号  $>$  换成等号  $=$  时，得到下式：

$$x_1^2 - x_2^2 = 0 \quad (39)$$

可以很容发现，(39) 为蜕化双曲线，图形为图 15 所示蓝色线。(38) 所示的不等式区域对应的是图 15 所示阴影区域。

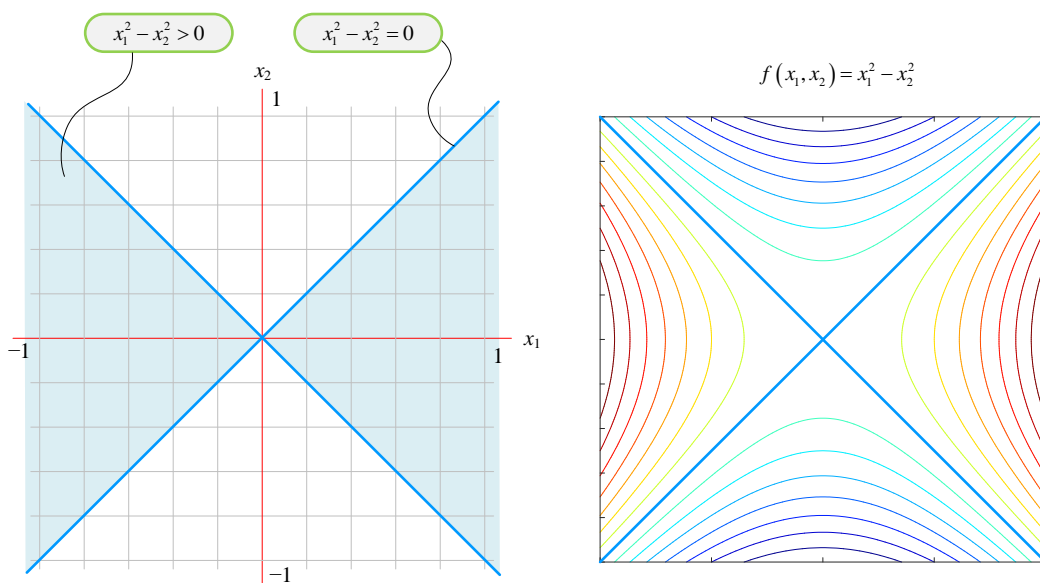


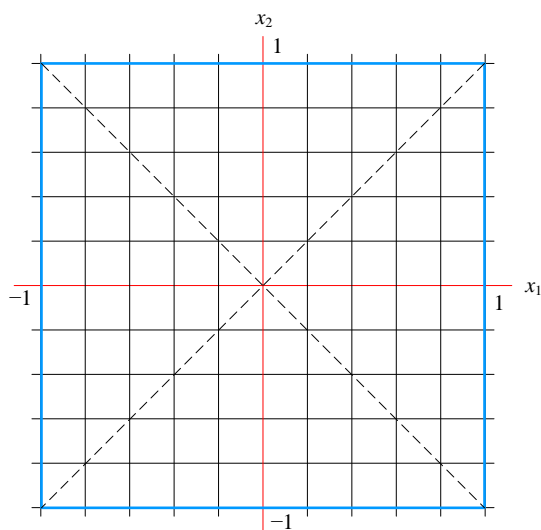
图 15. 蜕化双曲线及不等式区域

根据以上区域划分，改写 (37) 得到：

$$\begin{cases} x_1 = \pm 1 & x_1^2 - x_2^2 > 0 \\ x_2 = \pm 1 & x_1^2 - x_2^2 < 0 \end{cases} \quad (40)$$

由于  $x_1$  和  $x_2$  的取值范围均为  $[-1, 1]$ ，所以在图 15 所示阴影区域中，图像为两条竖直线段 ( $x_1 = \pm 1$ )；类似地，在  $x_1^2 - x_2^2 < 0$  对应区域中，图像为两条水平线段 ( $x_2 = \pm 1$ )。

综合以上分析，可以得到 (36) 对应的图像，具体如图 16 所示。

图 16.  $\max\{|x_1|, |x_2|\} = 1$  解析式图像

## 3.6 再谈距离度量

把 (2) 写成  $\mathbf{x}$  和  $\mathbf{q}$  两个列向量之差的  $L^p$  范数，可以得到：

$$\|\mathbf{x} - \mathbf{q}\|_p = \left( |x_1 - q_1|^p + |x_2 - q_2|^p + \cdots + |x_D - q_D|^p \right)^{1/p} = \left( \sum_{j=1}^D |x_j - q_j|^p \right)^{1/p} \quad (41)$$

其中， $p \geq 1$ ，列向量  $\mathbf{x}$  和  $\mathbf{q}$  分别为：

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_D \end{bmatrix} \quad (42)$$

$\mathbf{q}$  常被称作**查询点** (query point)。

如图 17 所示，(41) 相当于  $D$  维空间中， $\mathbf{x}$  和  $\mathbf{q}$  两点“距离”。距离  $\|\mathbf{x} - \mathbf{q}\|_p$  的取值为  $[0, +\infty)$ 。 $L^p$  范数的  $p$  取不同值时，我们得到不同的距离度量。白话说， $L^p$  范数这个数学工具把向量变成了非负标量，这个标量代表“距离”远近。

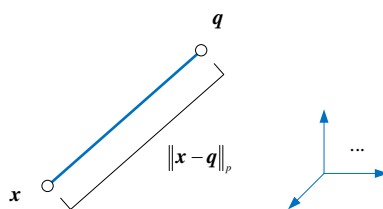


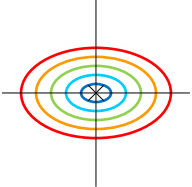
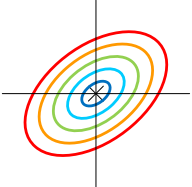
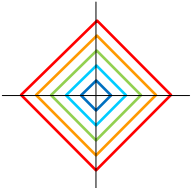
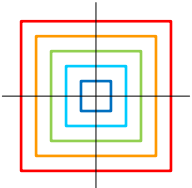
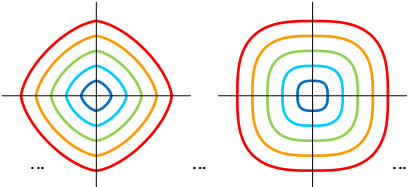
图 17.  $D$  维空间中  $\mathbf{x}$  和  $\mathbf{q}$  之间的“距离”

本系列丛书《数学要素》一册第 7 章给出表 1，表格总结常见距离度量的等距线。有了本章  $L^p$  范数这个数学工具，大家应该理解表 1 中欧氏距离、城市街区距离、切比雪夫距离、闵氏距离的背后的数学思想。

表 1. 常见距离定义及等距线形状，来自《数学要素》

距离度量	平面直角坐标系中等距线形状
欧氏距离 (Euclidean distance)	



标准化欧氏距离 (standardized Euclidean distance)	
马氏距离 (Mahalanobis distance)	
城市街区距离 (city block distance)	
切比雪夫距离 (Chebyshev distance)	
闵氏距离 (Minkowski distance)	

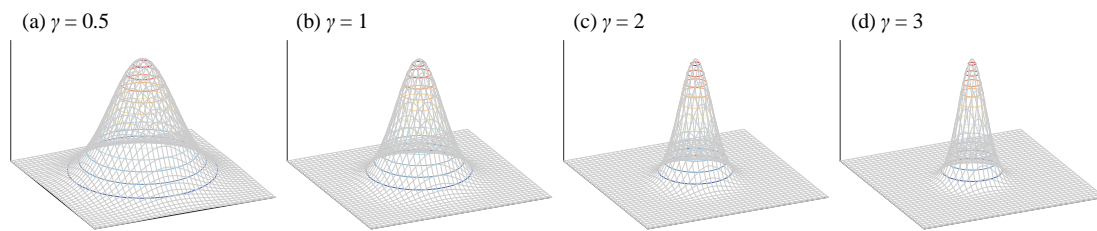
### 高斯核函数：从距离到亲密度

在很多应用场合，我们需要把“距离”转化为“亲密度”，就好比上一章余弦距离和余弦相似度之间的关系。

为了把距离  $\|\mathbf{x} - \mathbf{q}\|_p$  转化成亲密度，我们需要借助复合函数这个工具。本系列丛书《数学要素》一册介绍过高斯函数 (Gaussian function)。二元高斯函数的基本形式为：

$$f(x_1, x_2) = \exp(-\gamma(x_1^2 + x_2^2)) \quad (43)$$

图 18 所示  $\gamma$  对二元高斯核函数形状影响。 $\gamma$  越大坡面越陡峭。

图 18. 高斯核曲面随  $\gamma$  变化

有了  $L^2$  范数，我们就可以定义一个重要的函数——高斯核函数：

$$\kappa_{\text{RBF}}(\mathbf{x}, \mathbf{q}) = \exp(-\gamma \|\mathbf{x} - \mathbf{q}\|_2^2) = \exp(-\gamma \|\mathbf{x} - \mathbf{q}\|^2) \quad (44)$$

其中， $\gamma > 0$ 。高斯核函数也叫径向基核函数 (radial basis function kernel 或 RBF kernel)。不难发现，上式函数的取值范围为  $(0, 1]$ 。当  $\mathbf{x} = \mathbf{q}$  时函数值为 1，函数值无限接近 0，却不能取到 0。

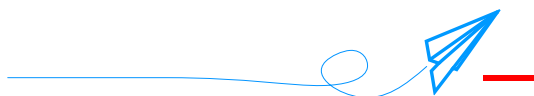
(44) 中  $\|\mathbf{x} - \mathbf{q}\|_2^2$  是  $L^2$  范数平方，即  $\mathbf{x}$  和  $\mathbf{q}$  两点欧几里得距离平方。径向基函数通过复合函数把代表距离的  $\|\mathbf{x} - \mathbf{q}\|_2^2$  变成亲近度。也就是说，距离平方值  $\|\mathbf{x} - \mathbf{q}\|_2^2$  越大，径向基函数越小，代表  $\mathbf{x}$  和  $\mathbf{q}$  越疏远。相反，距离平方值  $\|\mathbf{x} - \mathbf{q}\|_2^2$  越小，径向基函数越大，代表  $\mathbf{x}$  和  $\mathbf{q}$  越靠近。

从  $(\mathbf{x} - \mathbf{q})$  到  $\|\mathbf{x} - \mathbf{q}\|_2$ 、再到  $\exp(-\gamma \|\mathbf{x} - \mathbf{q}\|_2^2)$  是“向量  $\rightarrow$  距离 (标量)  $\rightarrow$  亲近度 (标量)”的转化过程。

(44) 也可以写成：

$$\kappa_{\text{RBF}}(\mathbf{x}, \mathbf{q}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{q}\|^2}{2\sigma^2}\right) \quad (45)$$

表 1 中还有两种重要的距离度量——标准化欧氏距离、马氏距离。想要理解这两种距离度量，需要线性变换、特征值分解、多元统计这几样数学工具，这是本书后续要展开讲解的内容。



本章从几何视角和大家聊了  $L^p$  范数。以下这四副图像总结本章的主要内容。 $L^p$  范数在本系列丛书的应用主要有两大方面：1) 距离度量；2) 正则化。请大家格外注意，只有  $p \geq 1$  时，才叫范数。

此外，请大家注意本章内容和本系列丛书《数学要素》第 7 章的“等距线”和第 9 章的“超椭圆”这两个数学概念的联系。

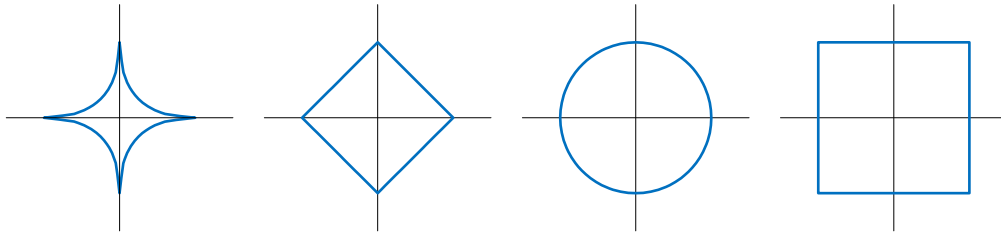


图 19. 总结本章重要内容的四副图