3 Vector Norm 向量范数

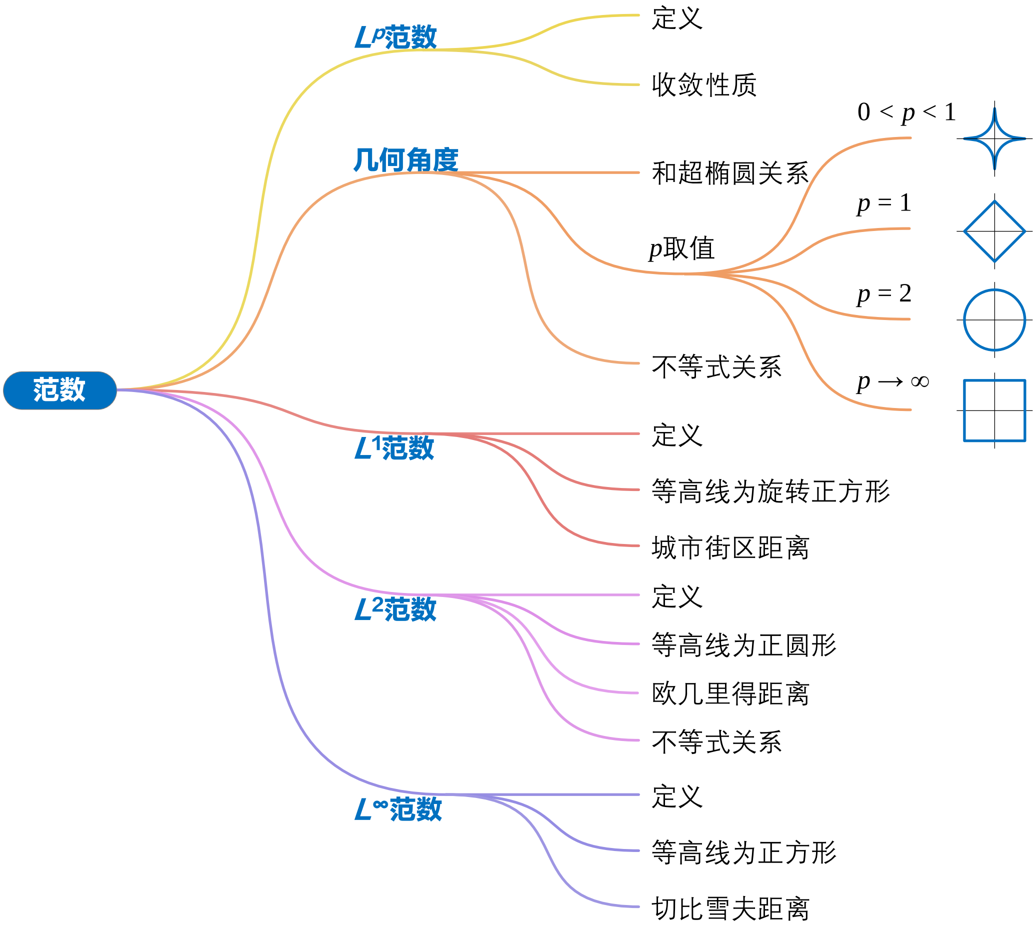
欧几里得距离的延伸

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.linsapce()` 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


范数
L^p范数
定义
收敛性质
几何角度
和超椭圆关系
p取值
$0 < p < 1$
$p = 1$
$p = 2$
$p \to \infty$
不等式关系
L^1范数
定义
等高线为旋转正方形
城市街区距离
L^2范数
定义
等高线为正圆形
欧几里得距离
不等式关系
L^∞范数
定义
等高线为正方形
切比雪夫距离

3.1 L^p范数：L^2范数的推广

本书前文介绍了 L^2 范数，本章将其推广到 L^p 范数。

给定如下列向量 $\boldsymbol{x}$：

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}^{\mathrm{T}} \tag{1}$$

向量 $\boldsymbol{x}$ 的 L^p 范数定义为：

$$\|\boldsymbol{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_D|^p\right)^{1/p} = \left(\sum_{i=1}^{D} |x_i|^p\right)^{1/p} \tag{2}$$

一般情况，p 取正数。

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

向量 $\boldsymbol{x}$ 的模便是 **L^2 范数** (L2-norm)，也叫 2-范数、欧几里得距离、欧氏距离，定义为：

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots x_D^2} = \left(\sum_{i=1}^{D} x_i^2\right)^{\frac{1}{2}} \tag{3}$$

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

特别地，当 p 趋向+∞时，对应的范数记成 L^∞。L^∞范数定义为：

$$\|\boldsymbol{x}\|_\infty = \max\left(|x_1|, |x_2|, \ldots, |x_D|\right) \tag{4}$$

即，$\|\boldsymbol{x}\|_\infty$ 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 $\boldsymbol{x}$：

$$\boldsymbol{x} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^{\mathrm{T}} \tag{5}$$

向量 $\boldsymbol{x}$ 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：

$$\|\boldsymbol{x}\|_1 = |1| + |2| + |3| = 6 \tag{6}$$

L^2 范数是图 1 向量 $\boldsymbol{x}$ 的长度：

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

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

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

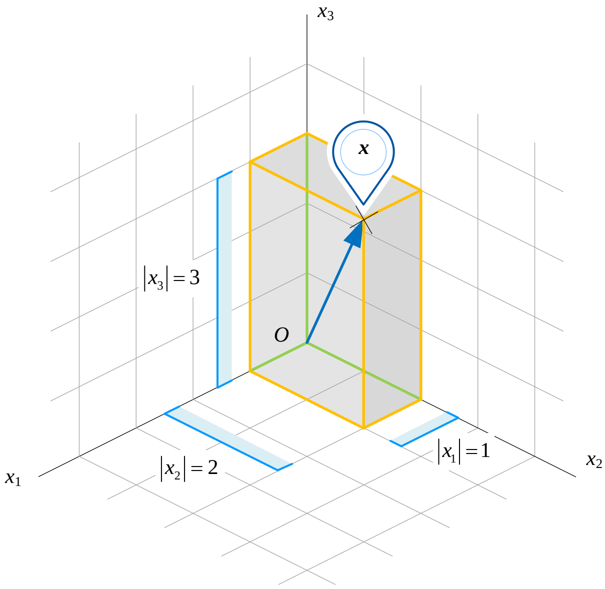


图 1. 向量 $\boldsymbol{x}$ 在三维直角坐标系的位置

类似地，计算向量 $\boldsymbol{x}$ 的 L^4 范数：

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

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

$$\|\boldsymbol{x}\|_\infty = \max\left(|1|, |2|, |3|\right) = 3 \tag{10}$$

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

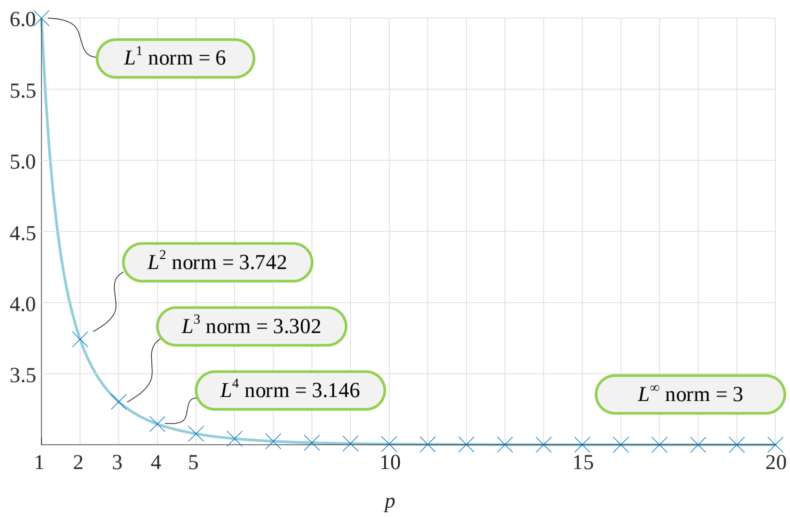


图 2. L^p 范数随 p 变化

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 $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$，$\boldsymbol{x}$ 的 L^p 范数为：

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

当 p 一定时，将 (11) 写成二元函数 $f(x_1, x_2)$：

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

观察 L^p 范数的定义，大家可能早已发现 L^p 范数和《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，也就是 L^p 范数取值变化规律。

$p = 1$ 时，$f(x_1, x_2)$ 函数的等高线为旋转正方形：

$$f(x_1, x_2) = |x_1| + |x_2| \tag{13}$$

$p = 2$ 时，$f(x_1, x_2)$ 函数等高线为正圆：

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

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

$$f(x_1, x_2) = \max\left(|x_1|, |x_2|\right) \tag{15}$$

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：

$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_1 \tag{16}$$

Bk4_Ch3_01.py 绘制图 3 所示等高线。

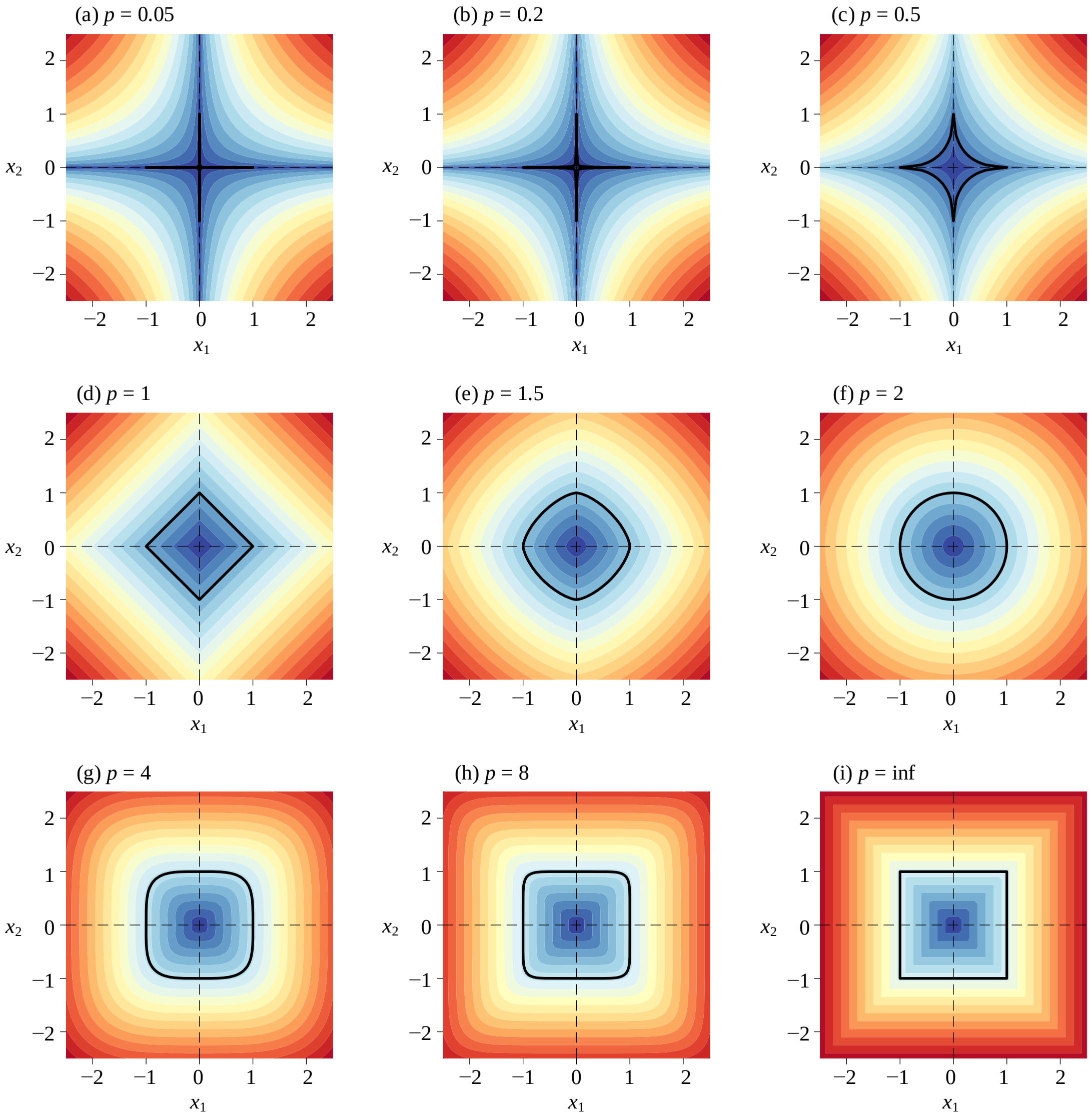


图 3. p 取不同正数时，L^p 范数等高线形状变化

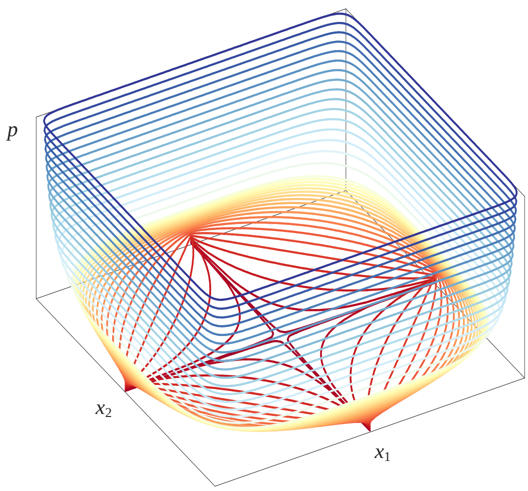


图 4. L^p 范数，随着 p 增大，等高线一层层包裹

凸凹性

$p > 1$ 时，L^p 范数等高线形状为凸 (convex)，比如图 5。$0 < p < 1$ 时，L^p 范数等高线形状为凹 (concave)，比如图 6。

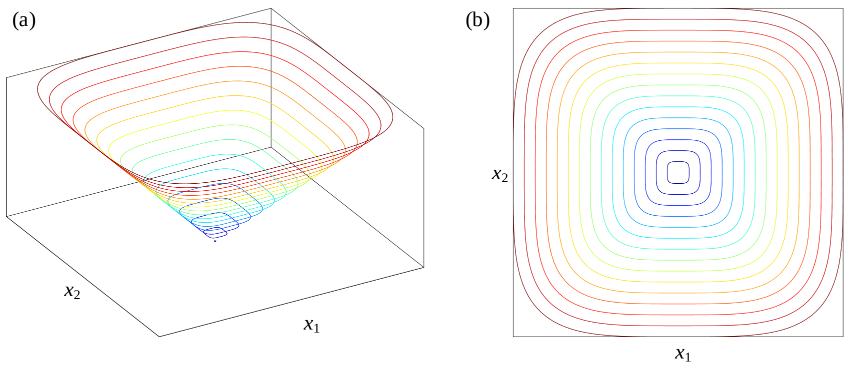


图 5. $p = 4$，L^p 范数等高线图像

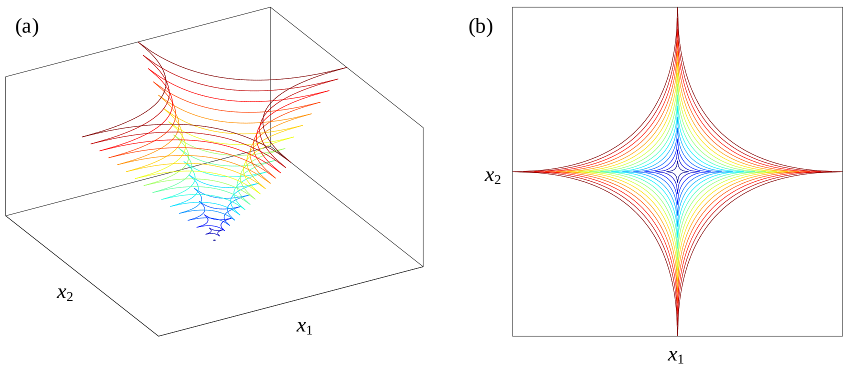


图 6. $p = 0.5$，L^p 范数等高线图像

p 为负数

虽然 L^p 范数 p 的取值一般为正数。有一些情况，p 可以取负数，图 7 所示为 p 取不同负数时，L^p 范数等高线形状变化。

图 8 所示为 $p = -2$，L^p 范数曲面空间和平面等高线。

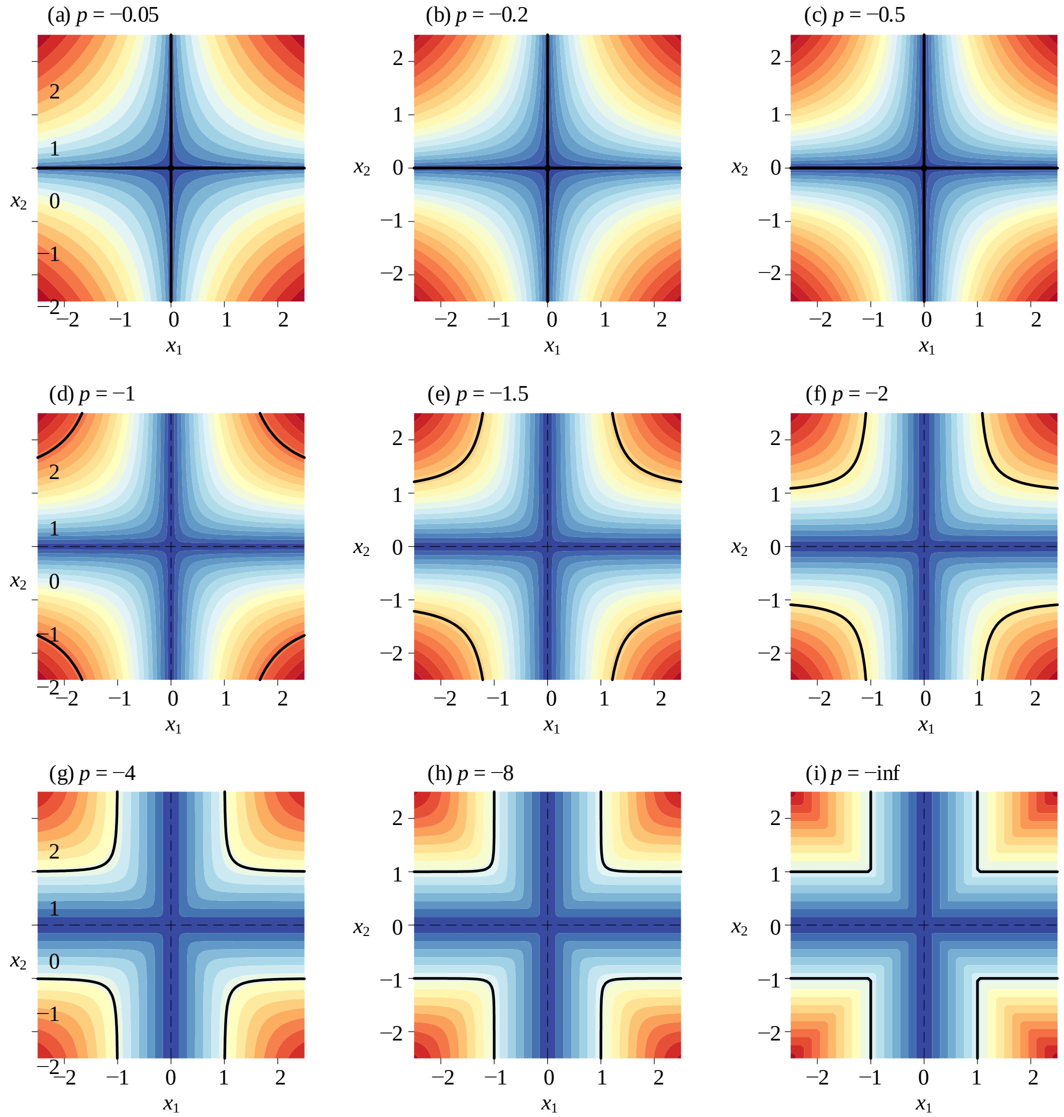


图 7. p 取不同负数时，L^p 范数等高线形状变化

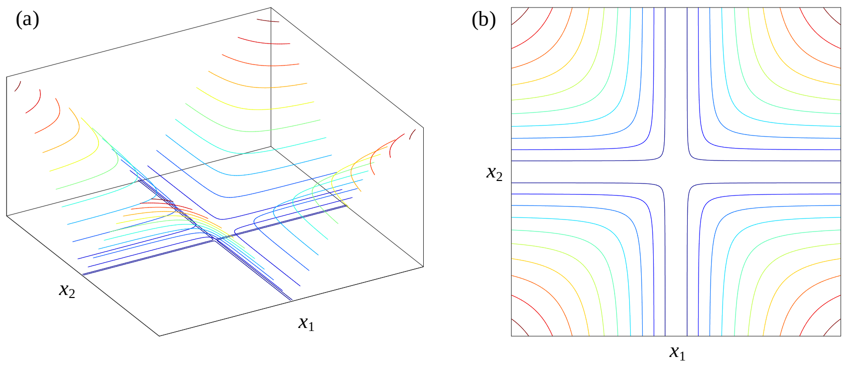


图 8. $p = -2$，L^p 范数等高线图像

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 $\boldsymbol{x}$ 的 L^1 范数定义为：

$$\|\boldsymbol{x}\|_1 = |x_1| + |x_2| + \cdots + |x_D| = \sum_{i=1}^{D} |x_i| \tag{17}$$

当 $D = 2$ 时，L^1 范数为：

$$\|\boldsymbol{x}\|_1 = |x_1| + |x_2| \tag{18}$$

(18) 中 L^1 范数等于 1 时，得到解析式：

$$|x_1| + |x_2| = 1 \tag{19}$$

下面，我分成几种情况展开 (19)，并绘制图像。

几何图形

观察 (19) 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，我们分四种情况考虑，得到如下展开式：

$$\begin{cases} x_1 + x_2 = 1 & 0 < x_1 < 1,\ \ 0 < x_2 < 1 \\ -x_1 + x_2 = 1 & -1 < x_1 < 0,\ \ 0 < x_2 < 1 \\ x_1 - x_2 = 1 & 0 < x_1 < 1,\ \ -1 < x_2 < 0 \\ -x_1 - x_2 = 1 & -1 < x_1 < 0,\ \ -1 < x_2 < 0 \end{cases} \tag{20}$$

根据 (20) 定义的四个一次函数解析式，可以得到图 9 所示图形。

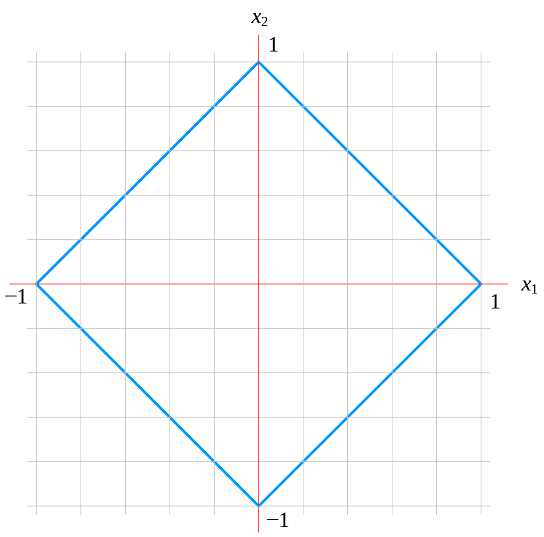


图 9. $|x_1| + |x_2| = 1$ 解析式图像

图 10 所示为如下函数的等高线图像：

$$f(x_1, x_2) = |x_1| + |x_2| \tag{21}$$

图 10 (b) 中每一条等高线上的点距离原点有相同的 L^1 范数。

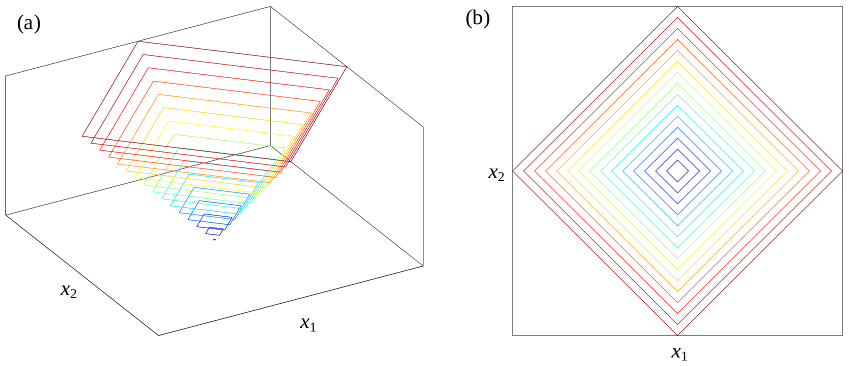


图 10. $p = 1$ 时，L^p 范数等高线图像

L^1 范数是我们将在本系列丛书《机器学习》一册中专门介绍的**城市街区距离** (city block distance)，也称**曼哈顿距离** (Manhattan distance)。

如图 11 所示，一个街区布局方正的城市，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 L1 正则项中起到重要作用。

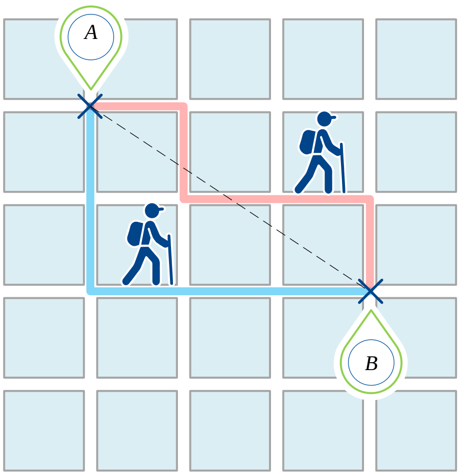


图 11. 城市街区距离

3.4 L^2 范数：正圆

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

$$\|\boldsymbol{x}\|_2 = \left(x_1^2 + x_2^2 + \cdots + x_D^2\right)^{1/2} = \left(\sum_{i=1}^{D} |x_i|^2\right)^{1/2} \tag{22}$$

特别地，当特征数 $D = 2$ 时，向量 $\boldsymbol{x}$ 的 L^2 范数为：

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2} \tag{23}$$

从距离度量角度，L^2 范数为欧几里得距离。

(23) 中 L^2 范数等于 1 时，对应如下平面图形解析式：

$$x_1^2 + x_2^2 = 1 \tag{24}$$

图 12 所示为 (24) 图像。图 13 所示为 L^2 三维和二维等高线图像。

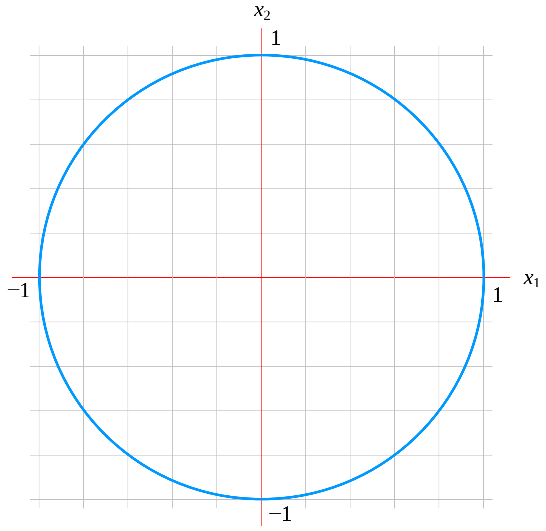


图 12. $x_1^2 + x_2^2 = 1$ 解析式图像

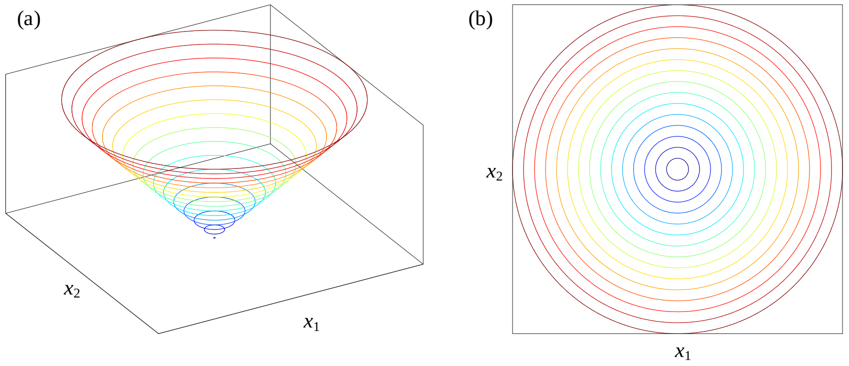


图 13. $p = 2$，L^p 范数等高线图像

另外，实践中也经常使用 L^2 范数的平方，即，

$$\|\boldsymbol{x}\|_2^2 = x_1^2 + x_2^2 \tag{25}$$

再次强调，范数、向量内积、矩阵乘法关系：

$$\|\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x} \cdot \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \tag{26}$$

图 14 所示为 L^2 范数平方的平面和三维等高线图像。

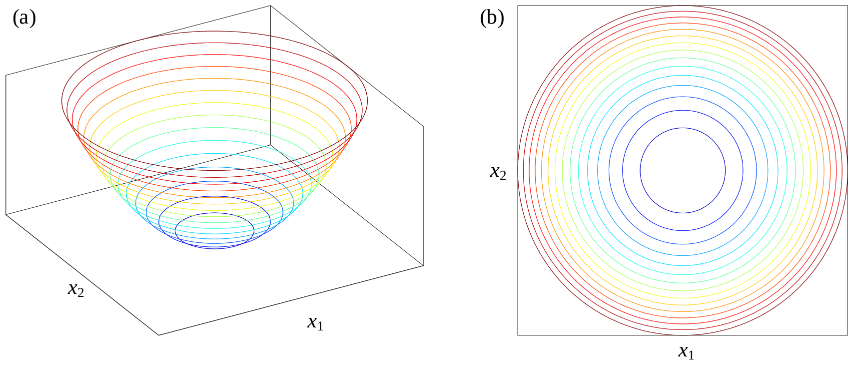


图 14. $p = 2$，L^p 范数平方等高线图

图 15 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。

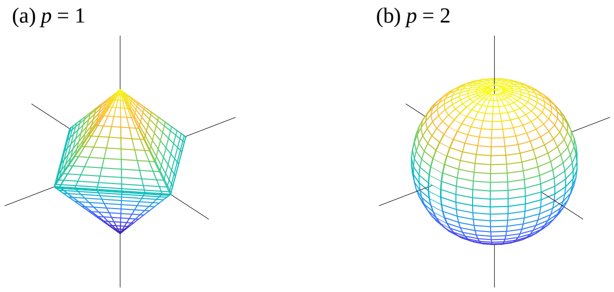


图 15. $p = 1$、2，$D = 3$ 时，L^p 范数对应的几何体

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 16 所示，正则化曲面是 L^1 和 L^2 曲面按不同比例叠加。注意，比例参数可以根据不同应用场景调整。图 16 中正则化曲面既有 L^1 的“尖点”，也有 L^2 的凸曲面。

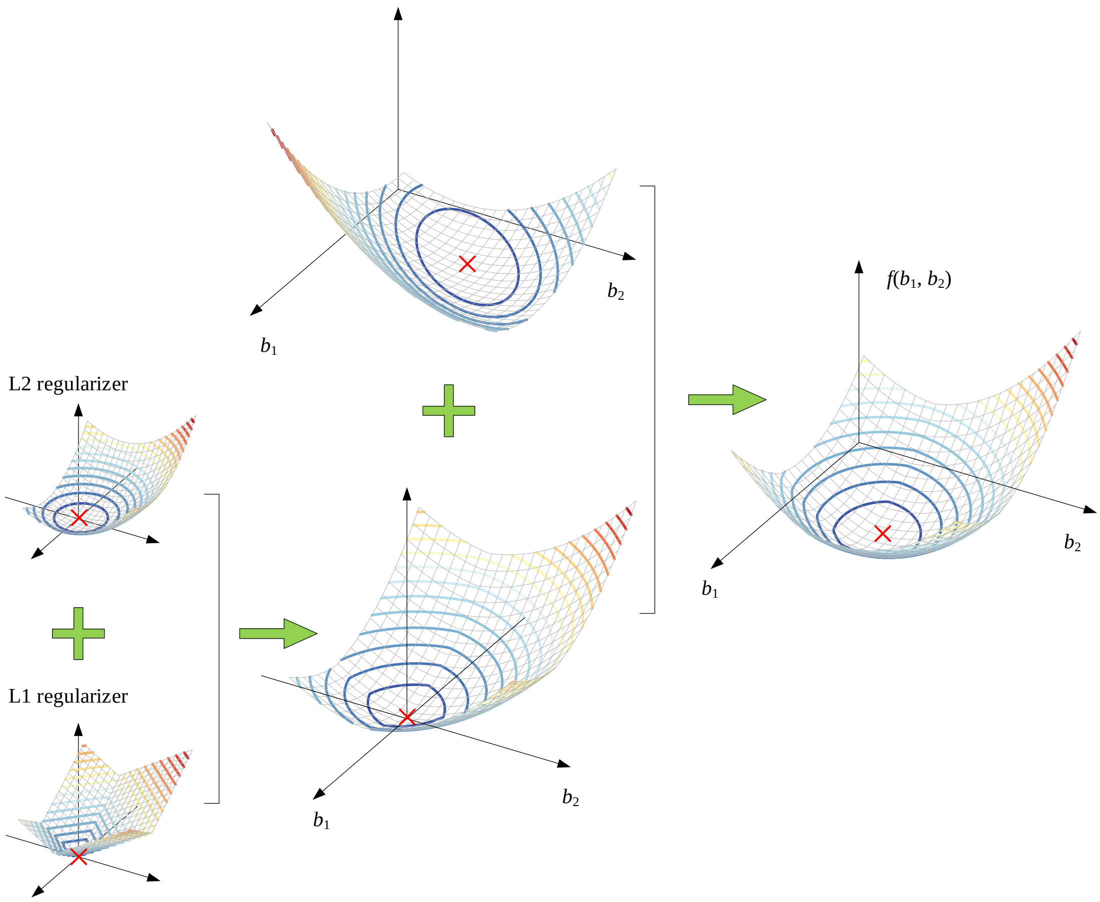


图 16. 弹性网络回归参数曲面

高斯函数、高斯核函数

本系列丛书《数学要素》一册介绍过高斯函数 (Gaussian function)。一元高斯函数的基本形式为：

$$f(x) = \exp(-\gamma x^2) \tag{27}$$

二元高斯函数的基本形式为：

$$f\left(x_1, x_2\right) = \exp\left(-\gamma\left(x_1^2 + x_2^2\right)\right) \tag{28}$$

图 17 所示 γ 对二元高斯核函数形状影响。γ 越大坡面越陡峭。

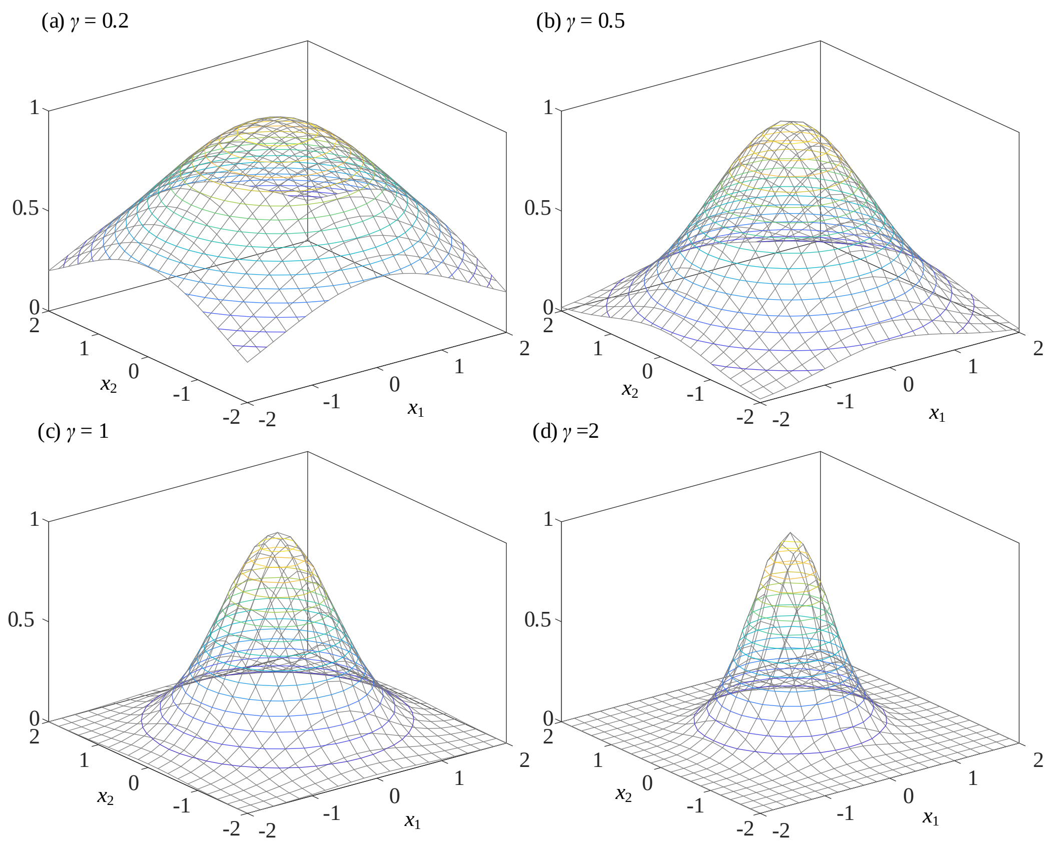


图 17. 高斯核曲面随 γ 变化

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：

$$\kappa_{\text{RBF}}\left(\boldsymbol{x}, \boldsymbol{q}\right) = \exp\left(-\gamma\left\|\boldsymbol{x} - \boldsymbol{q}\right\|_2^2\right) = \exp\left(-\gamma\left\|\boldsymbol{x} - \boldsymbol{q}\right\|^2\right) \tag{29}$$

其中，$\gamma > 0$。高斯核函数也叫径向基核函数 (radial basis function kernel 或 RBF kernel)。

(29) 中 $\left\|\boldsymbol{x} - \boldsymbol{q}\right\|_2^2$ 是 L^2 范数平方，即 $\boldsymbol{x}$ 和 $\boldsymbol{q}$ 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 $\left\|\boldsymbol{x} - \boldsymbol{q}\right\|_2^2$ 变成亲近度。也就是说，距离平方值 $\left\|\boldsymbol{x} - \boldsymbol{q}\right\|_2^2$ 越大，径向基函数越小，代表 $\boldsymbol{x}$ 和 $\boldsymbol{q}$ 越疏远。相反，距离平方值 $\left\|\boldsymbol{x} - \boldsymbol{q}\right\|_2^2$ 越小，径向基函数越大，代表 $\boldsymbol{x}$ 和 $\boldsymbol{q}$ 越靠近。

(29) 也可以写成：

$$\kappa_{\text{RBF}}(\boldsymbol{x},\boldsymbol{q}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{q}\|^2}{2\sigma^2}\right) \tag{30}$$

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：

$$\|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2 \geq \|\boldsymbol{u}+\boldsymbol{v}\|_2 \tag{31}$$

比如下例：

$$\boldsymbol{u} = \begin{bmatrix} 4 & 3 \end{bmatrix}^{\text{T}}, \quad \boldsymbol{v} = \begin{bmatrix} -2 & 4 \end{bmatrix}^{\text{T}} \tag{32}$$

向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 两者之和为：

$$\boldsymbol{u}+\boldsymbol{v} = \begin{bmatrix} 4 & 3 \end{bmatrix}^{\text{T}} + \begin{bmatrix} -2 & 4 \end{bmatrix}^{\text{T}} = \begin{bmatrix} 2 & 7 \end{bmatrix}^{\text{T}} \tag{33}$$

图 18 所示为向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 以及 $\boldsymbol{u}+\boldsymbol{v}$ 在平面上的关系。

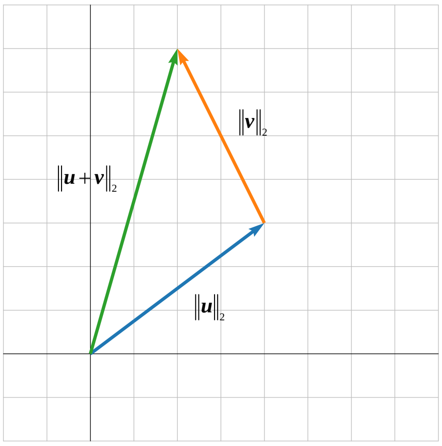


图 18. 向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 以及两者之和

$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的 L^2 范数分别为：

$$\|\boldsymbol{u}\|_2 = \sqrt{4^2+3^2} = 5, \quad \|\boldsymbol{v}\|_2 = \sqrt{(-2)^2+4^2} = \sqrt{20} \approx 4.4721 \tag{34}$$

$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的 L^2 范数和为：

$$\|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2 \approx 9.4721 \tag{35}$$

$\boldsymbol{u}+\boldsymbol{v}$ 的 L^2 范数为：

$$\|\boldsymbol{u}+\boldsymbol{v}\|_2 = \sqrt{2^2+7^2} = \sqrt{54} \approx 7.2801 \tag{36}$$

显然，(31) 成立。

Bk4_Ch3_02.py 绘制图 18 图 15。

3.5 L^∞范数：正方形

向量 $\boldsymbol{x}$ 的 L^∞范数的定义为：

$$\|\boldsymbol{x}\|_\infty = \max\left(|x_1|,|x_2|,\cdots,|x_D|\right) \tag{37}$$

上式就是我们将在丛书《机器学习》一册讲解的**切比雪夫距离** (Chebyshev distance)。

当特征数 $D = 2$ 时，向量 $\boldsymbol{x}$ 的 L^∞范数为：

$$\|\boldsymbol{x}\|_\infty = \max\left(|x_1|,|x_2|\right) \tag{38}$$

当 L^∞范数为 1 时，可以得到如下平面图形解析式：

$$\max\left\{|x_1|,\ \ |x_2|\right\} = 1 \tag{39}$$

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 (39) 解析式对应的图像。

几何图形

观察 (39) 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：

$$\begin{cases} |x_1| = 1 & |x_1| > |x_2| \\ |x_2| = 1 & |x_2| > |x_1| \end{cases} \tag{40}$$

为了进一步展开 (40)，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：

$$x_1^2 - x_2^2 > 0 \tag{41}$$

当把大于号 > 换成等号 = 时，得到下式：

$$x_1^2 - x_2^2 = 0 \tag{42}$$

可以很容易发现，(42) 为蜕化双曲线，如图 19 所示蓝色线。(41) 所示的不等式区域对应的是图 19 所示阴影区域。

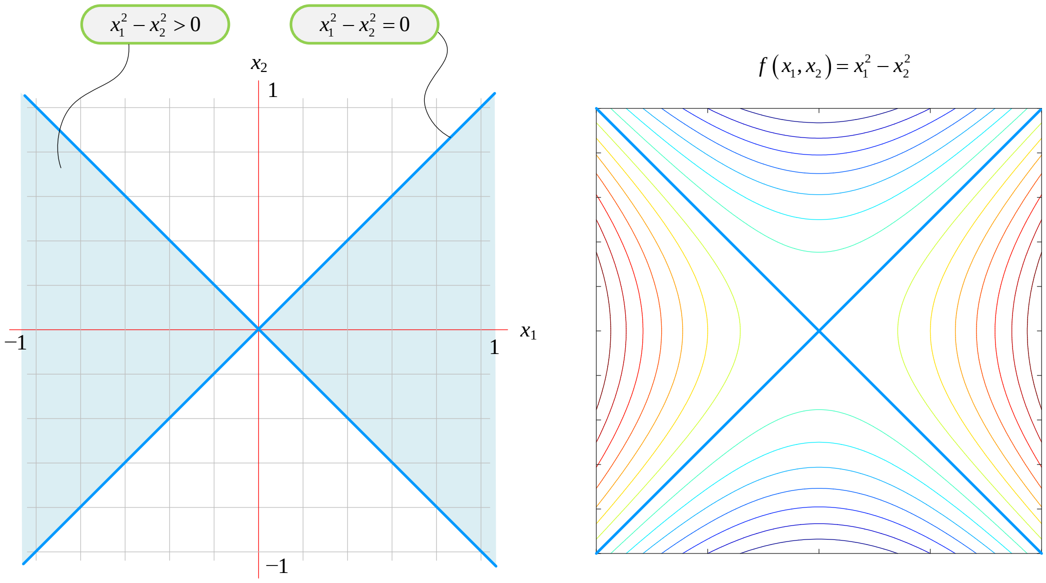


图 19. 蜕化双曲线及不等式区域

根据以上区域划分，改写 (40) 得到：

$$\begin{cases} x_1 = \pm 1 & x_1^2 - x_2^2 > 0 \\ x_2 = \pm 1 & x_1^2 - x_2^2 < 0 \end{cases} \tag{43}$$

由于 x_1 和 x_2 的取值范围均为 [−1, 1]，所以在图 19 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 (39) 对应的图像，具体如图 20 所示。

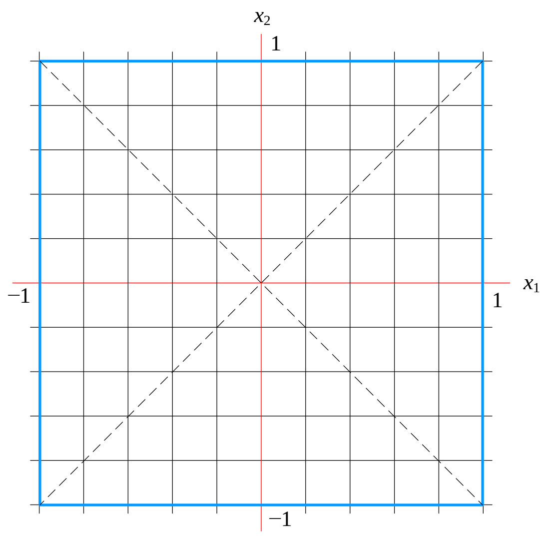


图 20. $\max\{|x_1|, |x_2|\}=1$ 解析式图像

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。

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。

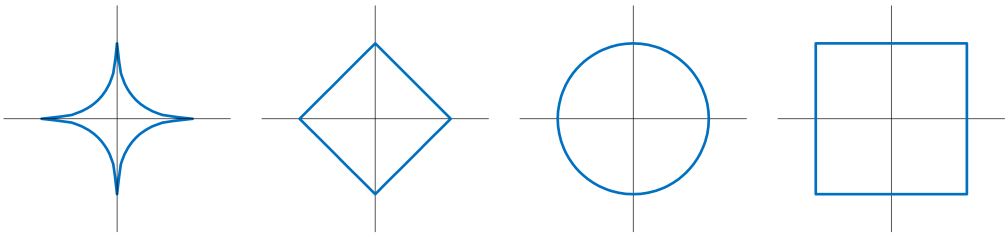


图 21. 总结本章重要内容的四副图