

25

Types of Distances

距离

两点连线、欧氏距离、闵氏距离、马氏距离



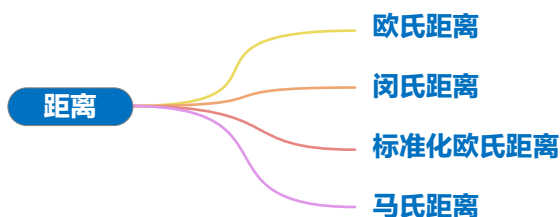
没有人会两次踏入同一条河流；江河川流不息，红尘物是人非。

No man ever steps in the same river twice. For it's not the same river and he's not the same man.

—— 赫拉克利特 (Heraclitus) | 古希腊哲学家 | 535 ~ 475 BC



- ◀ `numpy.diag()` 如果 A 为方阵, `numpy.diag(A)` 函数提取对角线元素, 以向量形式输入结果; 如果 a 为向量, `numpy.diag(a)` 函数将向量展开成方阵, 方阵对角线元素为 a 向量元素
- ◀ `numpy.linalg.inv()` 计算逆矩阵
- ◀ `numpy.linalg.norm()` 计算范数
- ◀ `scipy.spatial.distance.chebyshev()` 计算切比雪夫距离
- ◀ `scipy.spatial.distance.cityblock()` 计算城市街区距离
- ◀ `scipy.spatial.distance.euclidean()` 计算欧氏距离
- ◀ `scipy.spatial.distance.mahalanobis()` 计算马氏距离
- ◀ `scipy.spatial.distance.minkowski()` 计算闵氏距离
- ◀ `scipy.spatial.distance.seuclidean()` 计算标准化欧氏距离
- ◀ `sklearn.metrics.pairwise.euclidean_distances()` 计算成对欧氏距离矩阵
- ◀ `sklearn.metrics.pairwise_distances()` 计算成对距离矩阵



本 PDF 文件为作者草稿, 发布目的为方便读者在移动终端学习, 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

25.1 欧氏距离

在《编程不难》中，我们聊过有关距离度量相关 Python 函数。本章则和大家探讨各种距离的可视化方案。

欧氏距离，也叫**欧几里得距离** (Euclidean Distance)，是最常见的距离度量方法，它计算两个点之间的线段距离。如图 1 所示，在一维数轴上，任意一点 x 到原点的距离就是 x 的绝对值 $|x|$ 。

对于平面直角坐标系，欧氏距离可以通过使用勾股定理来计算。平面上任意一点 (x_1, x_2) 和原点 $(0, 0)$ 的距离为 $\sqrt{x_1^2 + x_2^2}$ 。如图 1 所示，欧氏距离的等距线为一系列同心圆。

如图 7 所示为平面上常见的两点连线的可视化方案。

BK_2_Ch25_01.ipynb 绘制图 7，请大家自行学习这个代码文件。

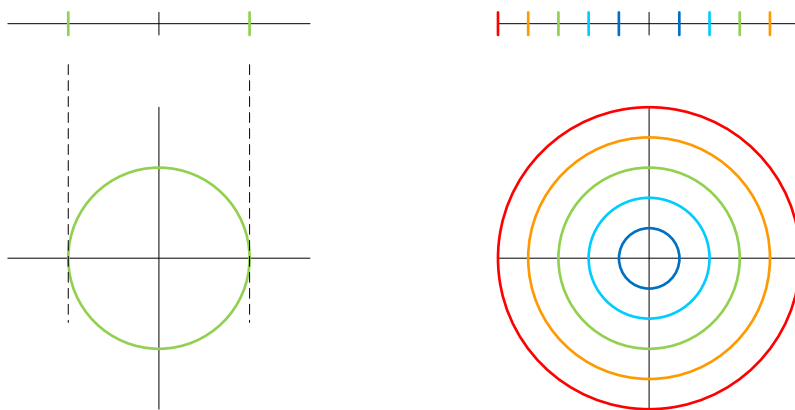


图 1. 数轴、平面直角坐标系上的欧氏距离等距线

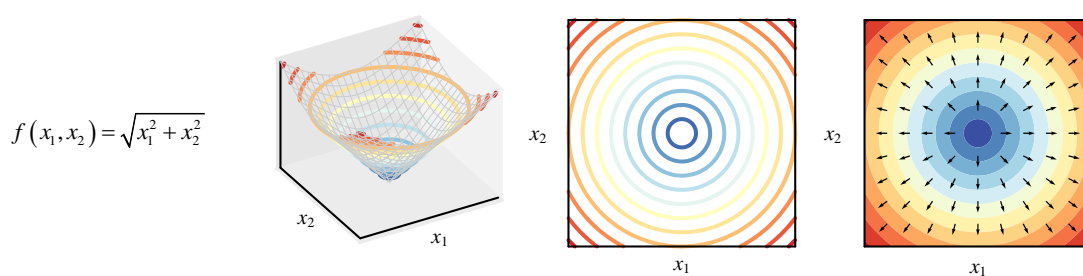



图 2. 平面上任意一点 (x_1, x_2) 和 $(0, 0)$ 之间的欧氏距离的几种可视化方案 |  BK_2_Ch25_02.ipynb

图 8 所示为三维空间中的两点连线。BK_2_Ch25_03.ipynb 绘制图 8，代码也很简单，请大家自行学习。

如图 3 所示，在三维直角坐标系中，任意一点 (x_1, x_2, x_3) 和 $(0, 0, 0)$ 之间的欧氏距离为 $\sqrt{x_1^2 + x_2^2 + x_3^2}$ ，其等距线为正球体。

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

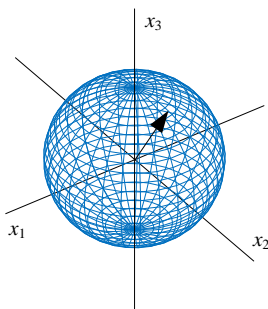


图 3. 三维直角坐标系上的欧氏距离等距线

图 9 所示为用“切豆腐”这种可视化方案展示三元欧氏距离。大家应该对“切豆腐”这种可视化方案并不陌生，本书后续还要用它展示各种三元函数。

图 10 所示为我们以欧氏距离为概念创作的两幅“生成艺术”。这幅图的概念十分简单，我们用颜色映射渲染欧氏距离远近。BK_2_Ch25_05.ipynb 绘制图 10，请大家自行学习这段代码。

25.2 其他距离度量

在机器学习中，距离不再仅仅是两点之间最短的线段。距离变成用于衡量两个对象或数据之间的相似性或差异性的概念。不同的距离度量方法可以基于不同的度量标准和算法来计算。以下是几种常见的距离度量方法。

如图 4 所示，在平面上，城市街区距离，也叫**曼哈顿距离**（Manhattan Distance）或 L^1 距离，通过沿着坐标轴的垂直和水平线段的长度之和来测量两个点之间。

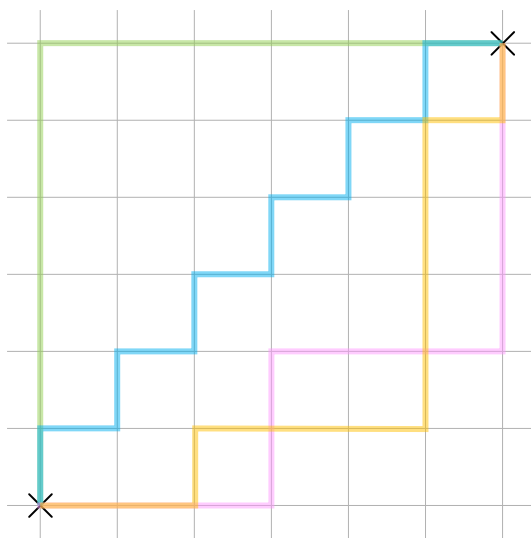


图 4. 城市街区距离

L^p 范数是向量的一种度量方式，用来衡量向量中各个元素的大小。在 L^p 范数中， p 是一个实数，并且 p 大于等于 1。



《矩阵力量》第 3 章将专门讲解向量 L^p 范数。

L^p 距离，也叫**闵氏距离** (Minkowski Distance)，是使用 L^p 范数来度量两个向量之间的距离。图 11 所示为利用填充等高线可视化 L^p 距离度量。请大家格外关注图 11 等高线的形状。

图 12 则采用三维等高线织成的网面来呈现三维空间中的 L^p 距离“等距面”。

当 $p = 1$ 时，得到的是城市街区距离；当 $p = 2$ 时，得到的是欧氏距离；当 p 趋近于无穷大时，得到的是**切比雪夫距离** (Chebyshev distance)。

图 5 所示为用 Streamlit 创作的展示平面 L^p 范数的 App。图 6 中 Streamlit App 可视化三维 L^p 范数。

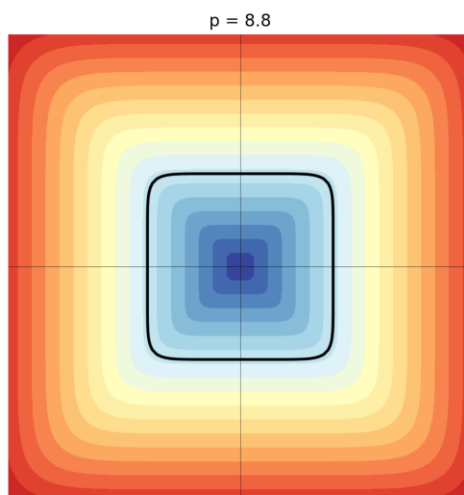
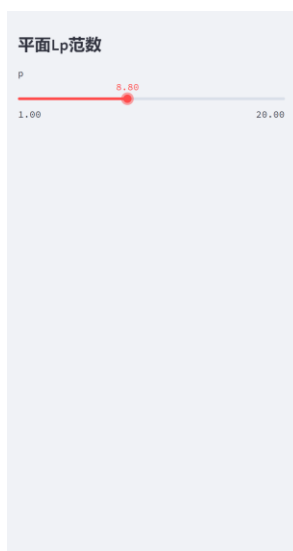



图 5. 展示平面 L^p 范数的 App, Streamlit 搭建 |  Streamlit_2D_Lp_Norm.py

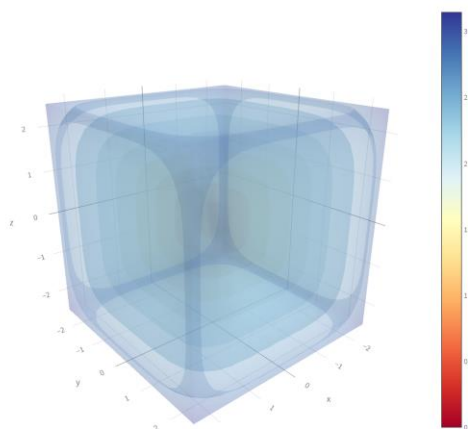
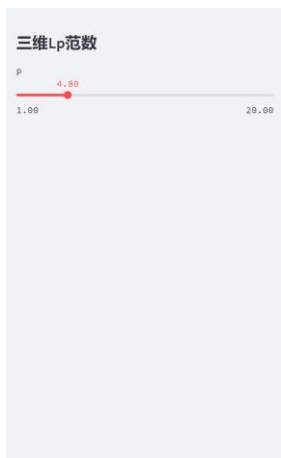



图 6. 展示三维 L^p 范数的 App, Streamlit 搭建 |  Streamlit_3D_Lp_Norm.py

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

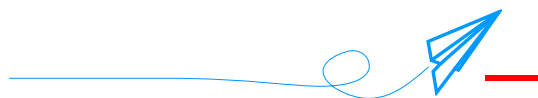
标准化**欧氏距离** (Standardized Euclidean Distance) 是对欧氏距离进行标准化的一种方法。在计算标准化欧氏距离时，对每个维度的值进行标准化处理，然后再计算欧氏距离。

读到这里，大家应该对马氏距离并不陌生了。简单来说，**马氏距离** (Mahalanobis Distance) 是一种考虑特征之间相关性的距离度量方法。它使用协方差矩阵来衡量特征之间的相关性，从而在计算距离时考虑到了特征之间的相关性。

图 13 所示为基于鸢尾花样本数据的六种距离度量的等高线。观察这幅图，大家可以看到各种距离度量的起点都是数据质心。除了散点图、等高线，这组图还用到了本书前文介绍的“网格”这种可视化方案。



特别建议大家回顾本书第 13 章介绍的马氏距离计算，以及多元高斯分布 PDF 计算过程。



本章可视化的对象是距离。我们用到了散点、线图、等高线、填充等高线、网格面等等可视化方案向大家距离不仅仅是两点一线。

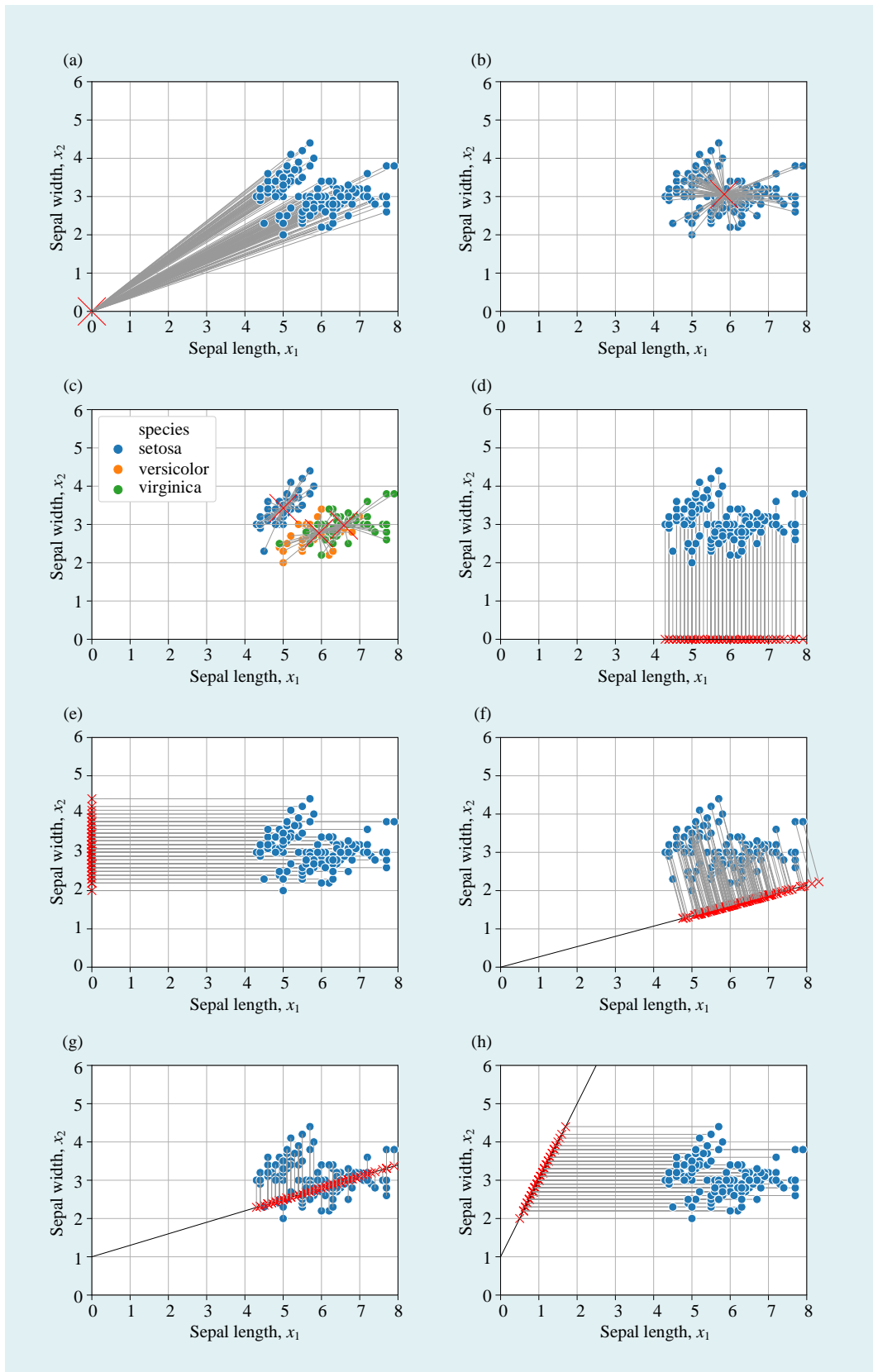


图 7. 平面两点距离的几种情况 | BK_2_Ch25_01.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

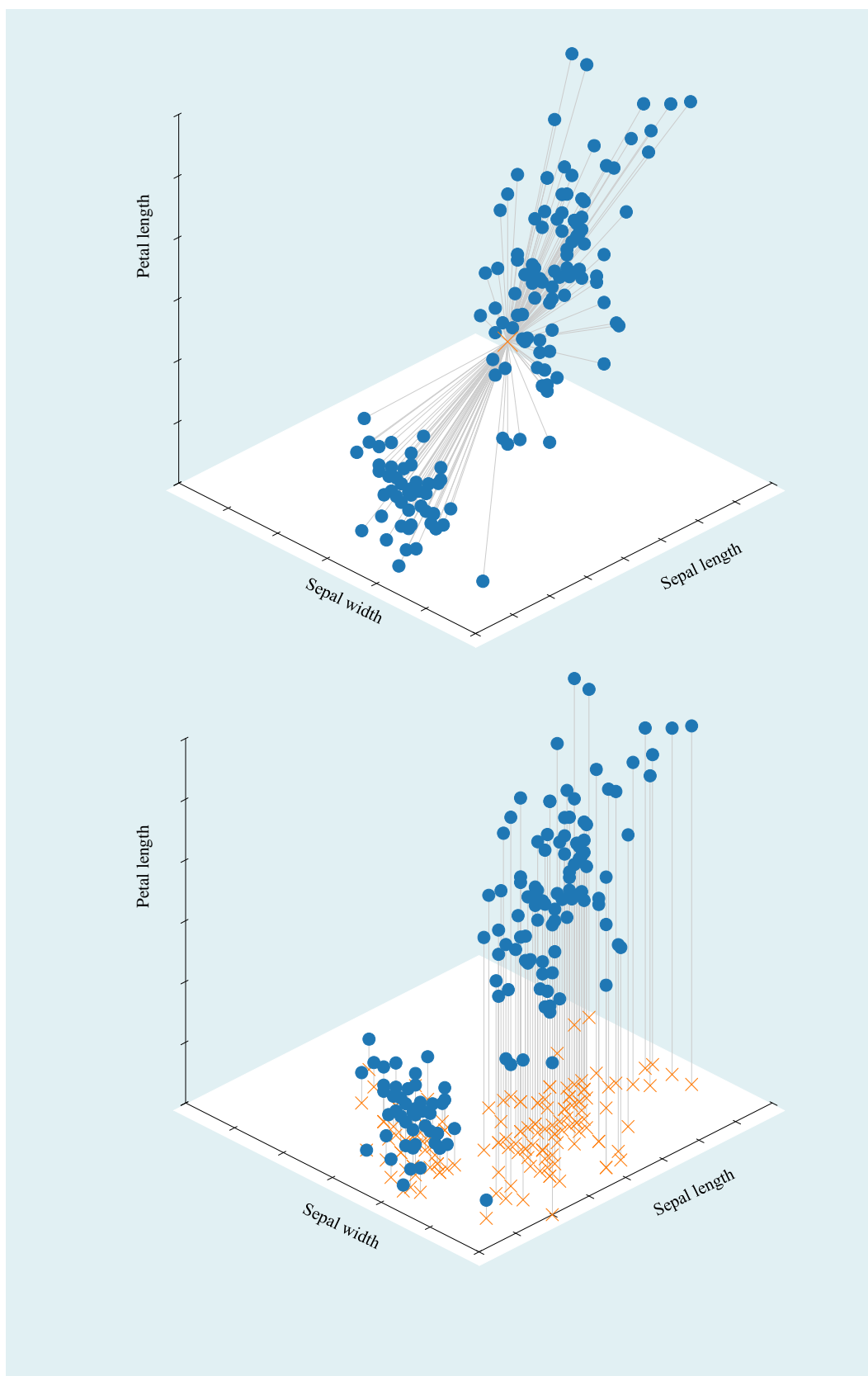


图 8. 三维空间距离 | BK_2_Ch25_03.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

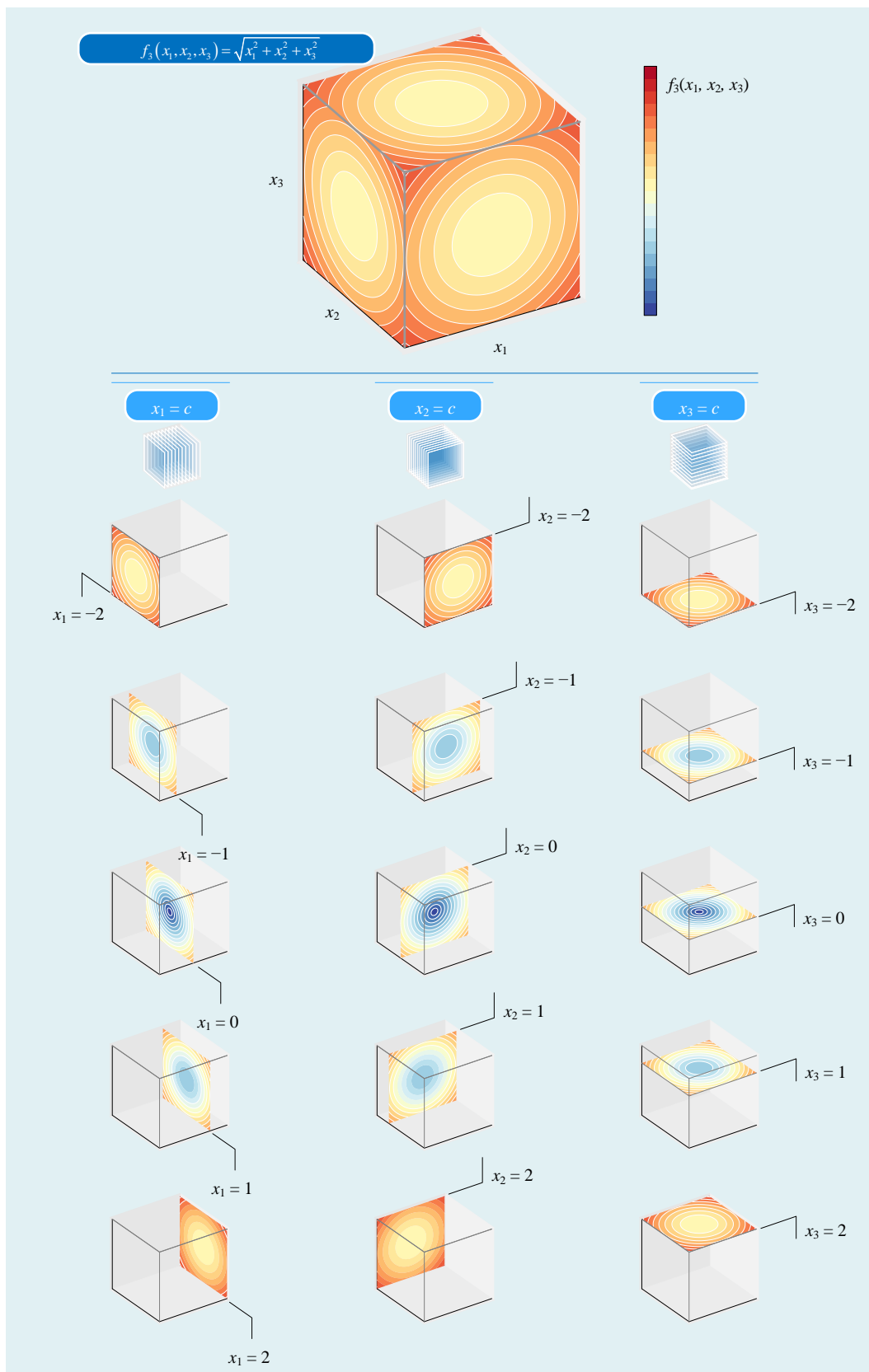


图 9. 三维直角坐标系中，任意一点 (x_1, x_2, x_3) 和 $(0, 0, 0)$ 之间的欧氏距离 | BK_2_Ch25_04.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

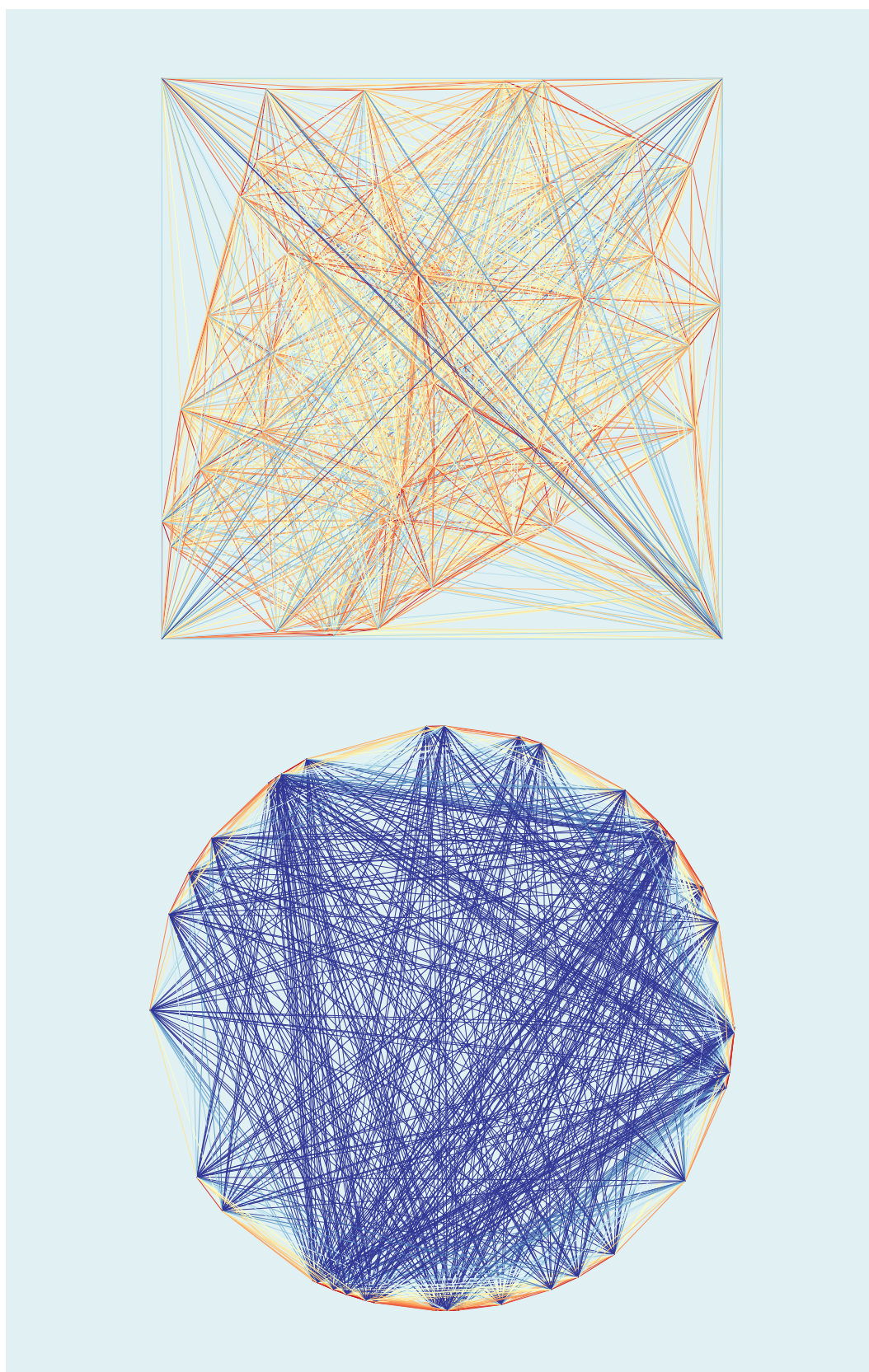



图 10. 根据欧氏距离远近渲染两点连线 |  BK_2_Ch25_05.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

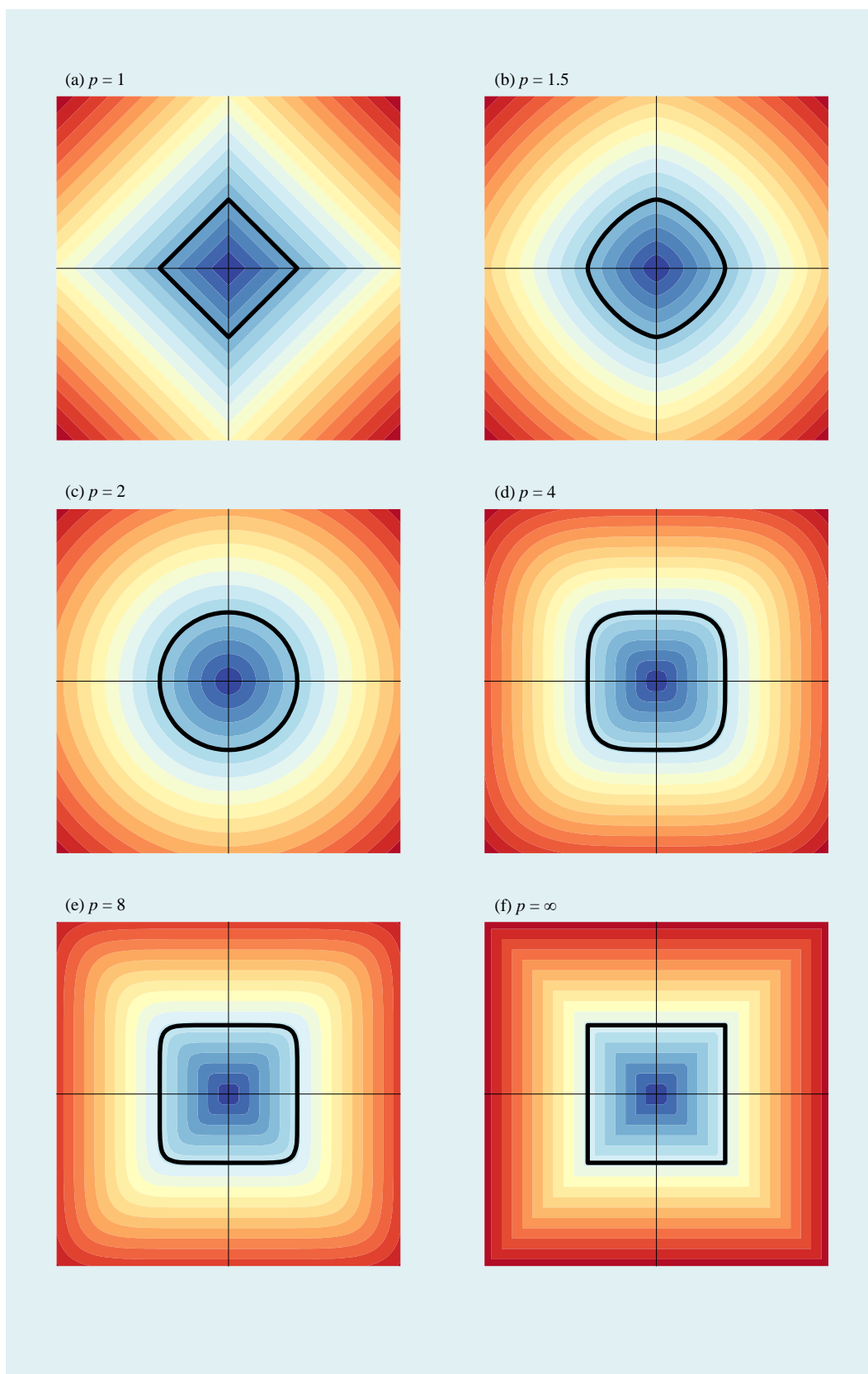


图 11. 向量范数, 二维等高线 |  BK_2_Ch25_06.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

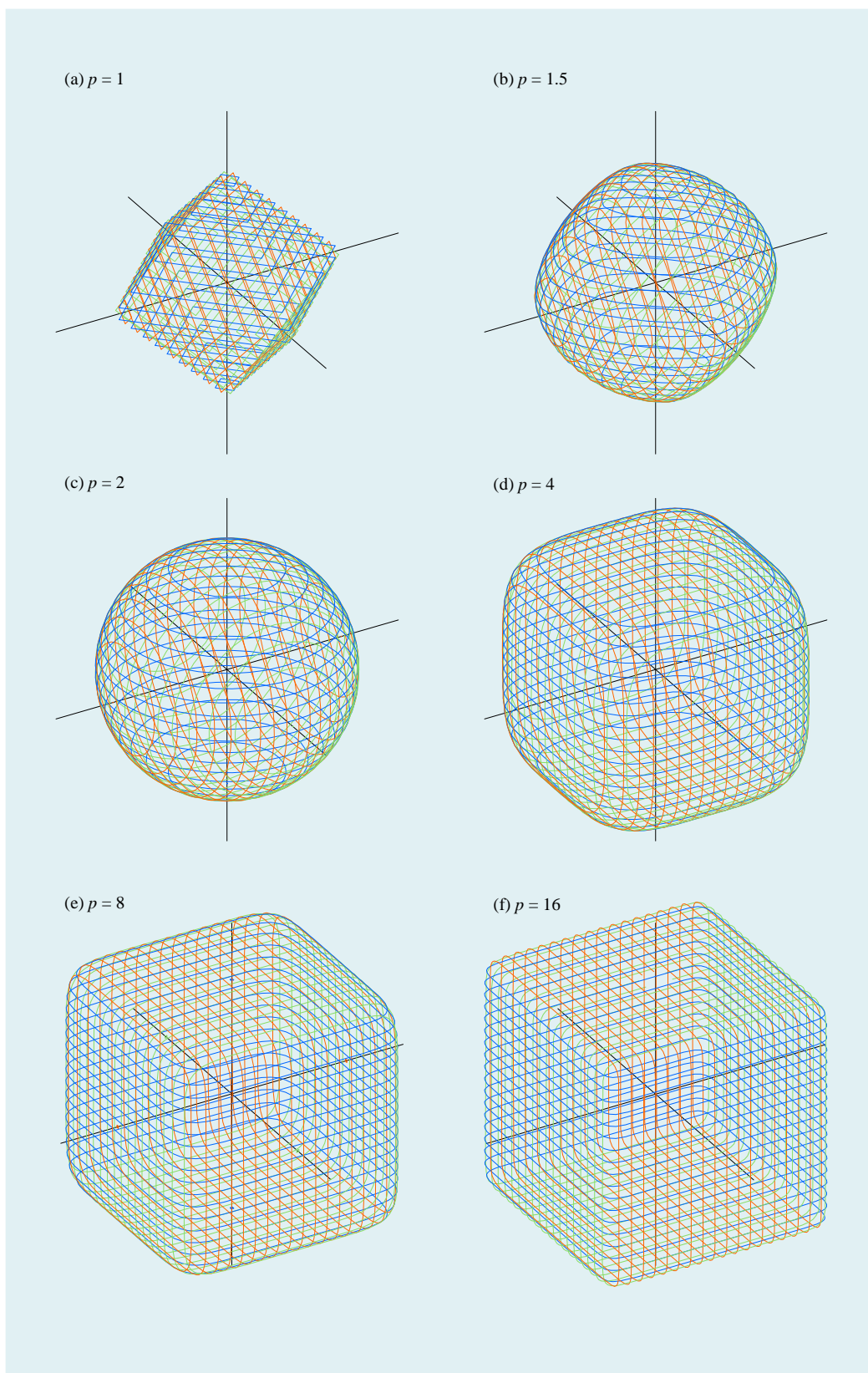


图 12. 向量范数, 三维几何体 |  BK_2_Ch25_07.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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

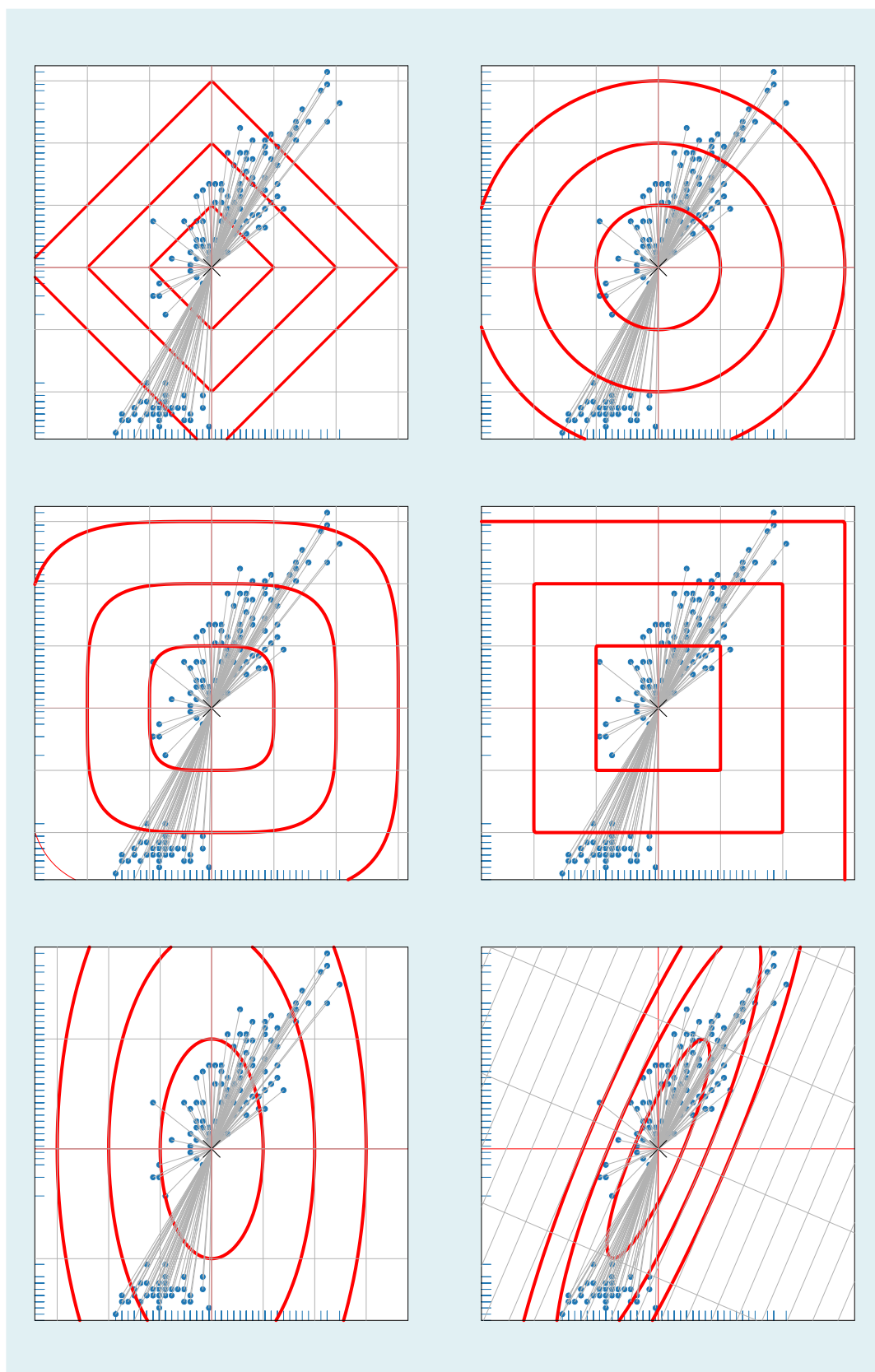



图 13. 基于鸢尾花数据的各种距离度量等高线 |  BK_2_Ch25_08.ipynb

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，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