

17

Heatmap

热图

鸢尾花书中常用来可视化矩阵运算



每个孩子都是艺术家。问题在于他长大后如何保持艺术家的本质。

Every child is an artist. The problem is how to remain an artist once he grows up.

—— 毕加索 (Pablo Picasso) | 西班牙艺术家 | 1881 ~ 1973



- ◀ `numpy.linalg.cholesky()` Cholesky 分解
- ◀ `numpy.linalg.eig()` 特征值分解
- ◀ `numpy.linalg.svd()` 奇异值分解
- ◀ `numpy.zeros_like()` 用来生成和输入矩阵形状相同的零矩阵
- ◀ `seaborn.clustermap()` 绘制聚类热图
- ◀ `seaborn.heatmap()` 绘制热图
- ◀ `sklearn.datasets.load_iris()` 加载鸢尾花数据

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

17.1 热图

Seaborn 中的热图

热图 (heatmap) 是“鸢尾花书”中极为常见的可视化手段。特别是在展示数据、矩阵分解时，我们常用热图可视化矩阵。

虽然，matplotlib 中也有绘制热图的工具；但是，推荐大家使用 seaborn 中的 heatmap 函数。这个函数绘制热图更方便。

Seaborn 是一款基于 matplotlib 的数据可视化库，其中包括了各种绘图函数，其中之一就是 heatmap。使用 Seaborn 的 heatmap 函数可以让大家快速而方便地可视化矩阵数据，使得数据分析更加直观和易于理解。

heatmap 可以用于可视化二维数组。图 1 所示为用热图可视化鸢尾花四个量化特征数据。在 Jupyter notebook 中，大家可以看到我们用 cmap 控制色谱，用 xticklabels、yticklabels 分别控制横轴、纵轴标签，用 cbar_kws 设置色谱条位置，并用 vmin、vmax 控制色谱条起止位置。

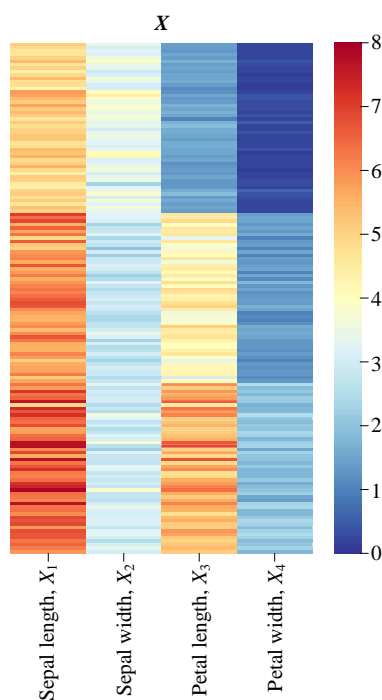


图 1. 热图可视化鸢尾花数据

Seaborn 中的 heatmap 函数还包括许多其他参数，用于自定义热图的外观和行为。例如，大家可以使用 annot 参数在热图中显示数值，使用 fmt 参数指定数字格式，使用 linewidths 参数调整单元格边框宽度等等。

聚类热图

Seaborn 中，`clustermap` 是一个用于绘制聚类热图的函数，其原理是将矩阵中的行和列进行聚类，并以聚类后的顺序重新排列矩阵的行和列。这样可以将具有相似特征的行和列放在一起，从而更容易地发现它们之间的相似性和差异性。图 2 所示为鸢尾花数据的聚类热图。

“鸢尾花书”的《机器学习》将专门讲解各种聚类算法。

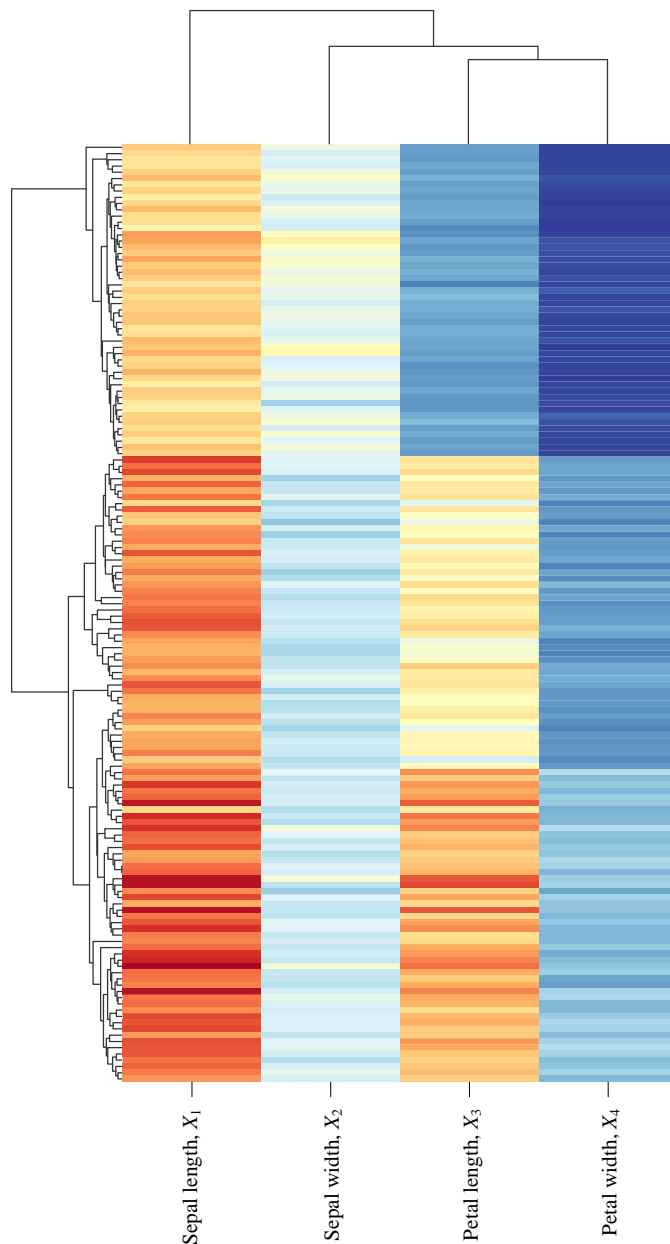


图 2. 热图可视化鸢尾花数据

17.2 矩阵运算

“鸢尾花书”中，大家会经常看到用一组热图可视化矩阵运算，特别是矩阵分解。图 5 所示为常见的几个矩阵运算。注意，后期制作时，热图的形状做了修改。

《矩阵力量》一册将从代数、数据、线性组合、优化、几何、统计等角度和大家讨论这些矩阵运算。此外，大家还会看到我们用热图可视化协方差矩阵、相关性系数矩阵，以及这些矩阵对应的线性代数运算。本节就不再展开讨论了。

17.3 伪彩色网格图

在 Matplotlib 中，`pcolormesh` 函数用于创建一个伪彩色网格图，类似热图。它可以用于绘制二维数据的色彩填充图，其中每个数据点的颜色根据其对应的数值进行映射。

在 `pcolormesh` 函数中，可以使用 `rasterized` 参数来控制是否将图形渲染为矢量图形或光栅图像。`rasterized` 参数是一个布尔值，用于指定是否将图形渲染为光栅图像。当设置为 `True` 时，图形将以光栅化的形式保存，这对于包含大量数据点或复杂图形时可以提高渲染性能和文件大小。默认情况下，`rasterized` 参数的值为 `False`，即图形以矢量格式渲染。

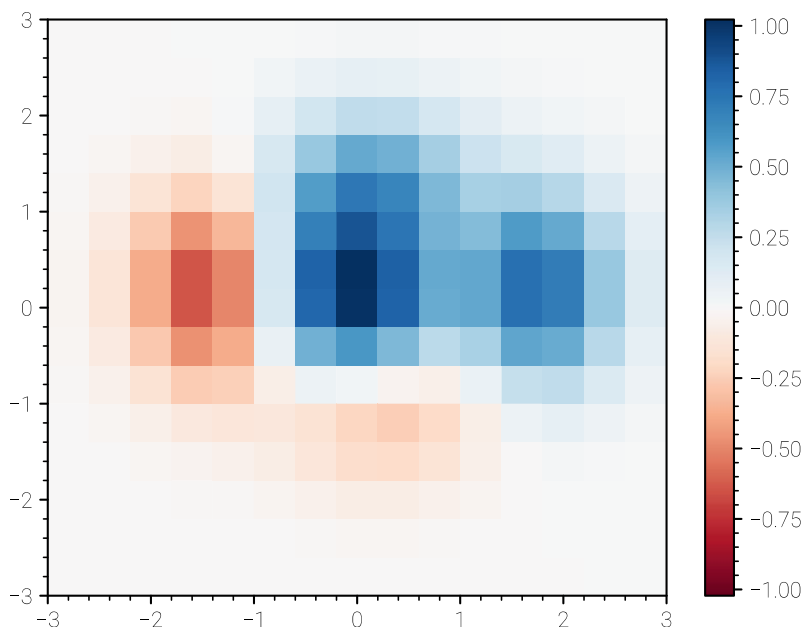


图 3. 利用 `pcolormesh` 绘制的伪彩色图

如图 6 所示，`pcolormesh` 函数还常用来绘制分类算法的决策边界。此外，`pcolormesh` 函数可以绘制网格，并用来可视化线性、非线性变换，具体如图 4、图 7 所示。

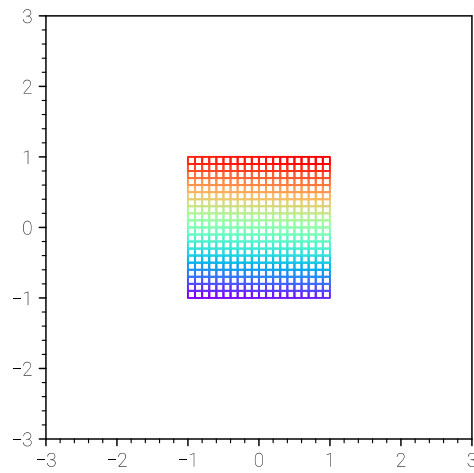


图 4. 利用 pcolormesh 绘制方正网格

pcolor 函数也是 matplotlib 库中的函数，用于绘制伪彩色图，效果和 pcolormesh 类似。与 pcolor 相比，pcolormesh 在效率上更高，特别适用于绘制大型数据集。

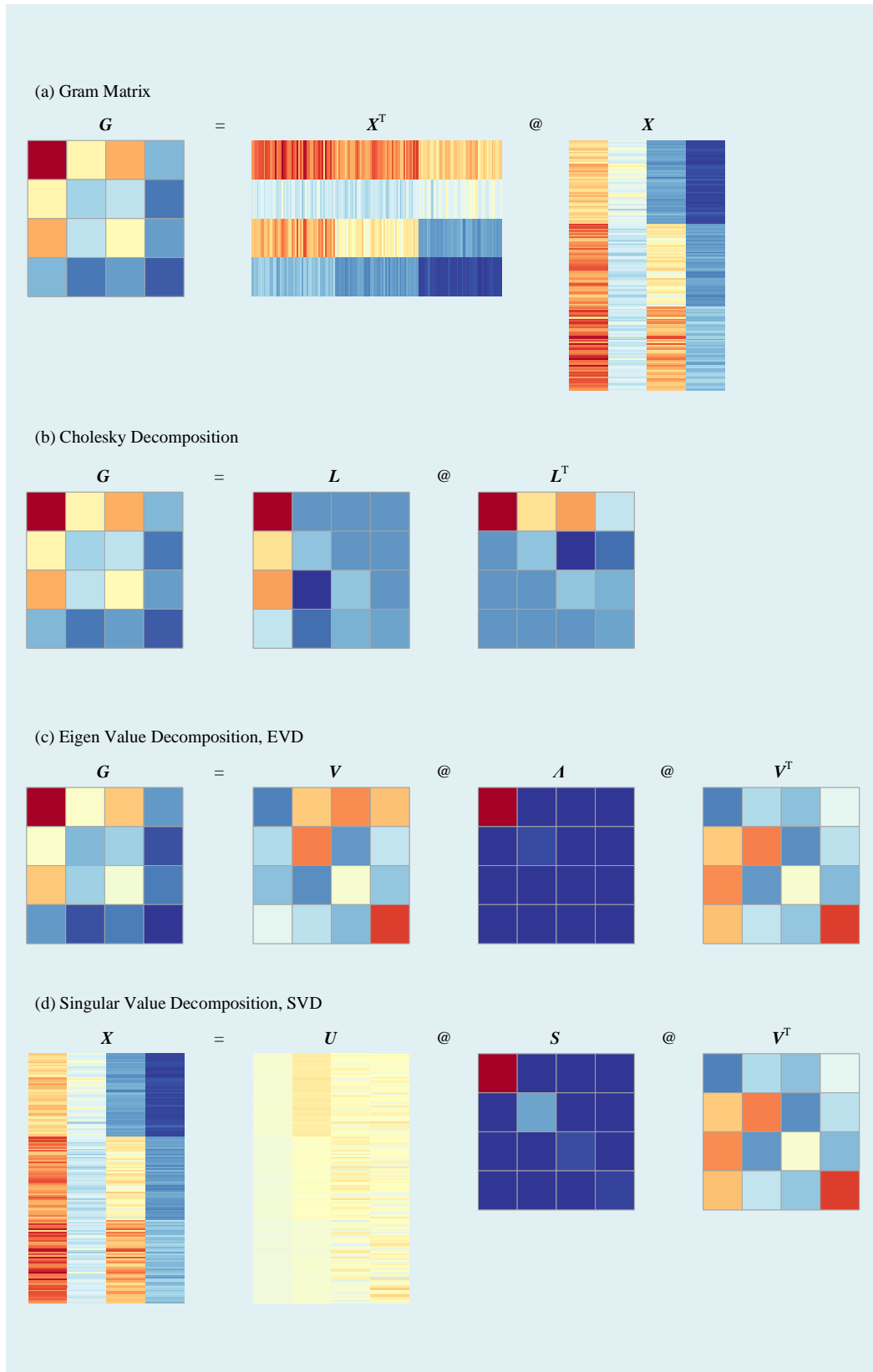


图 5. 用热图可视化矩阵运算

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

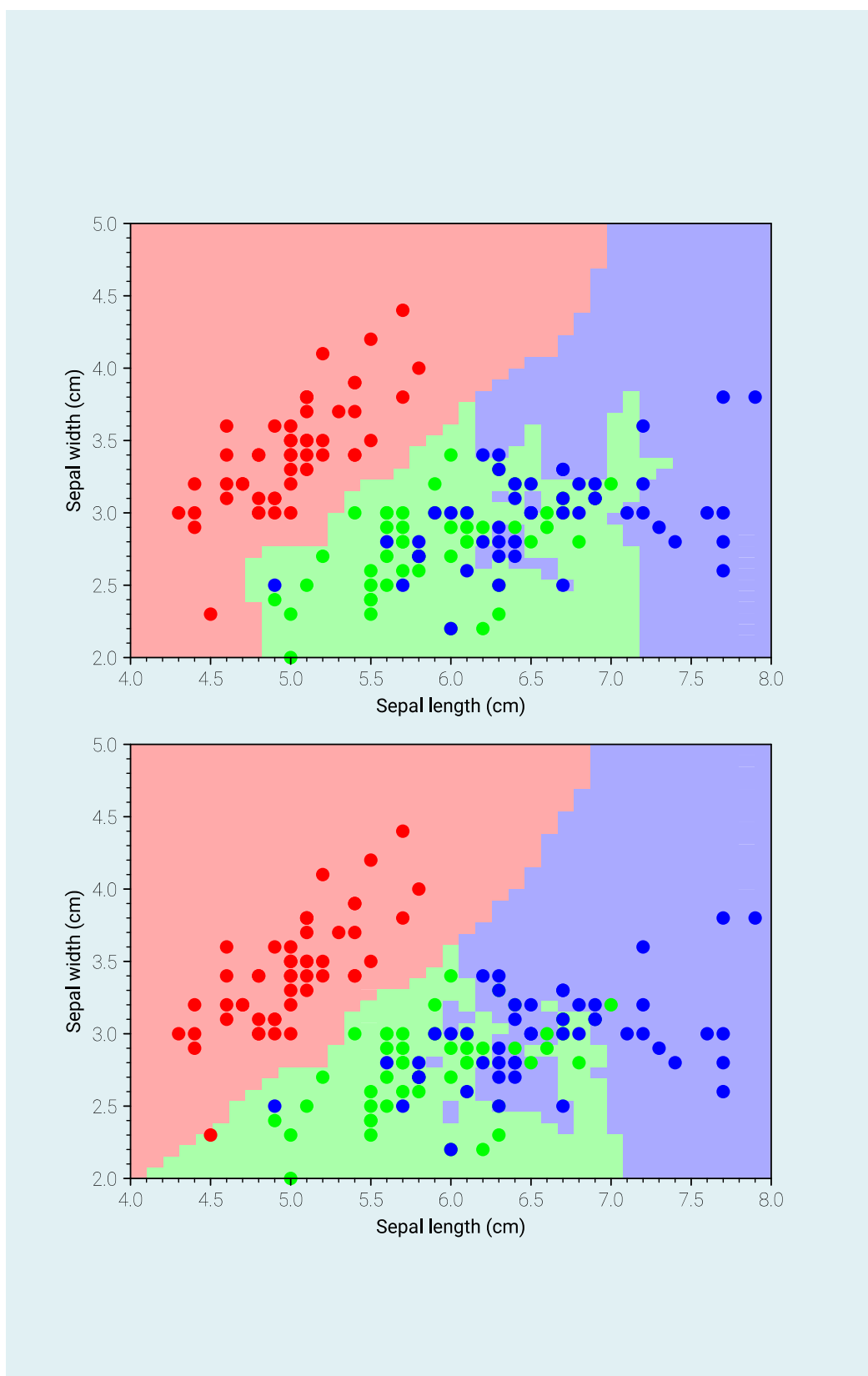


图 6. 用 pcolormesh 函数绘制分类决策边界

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

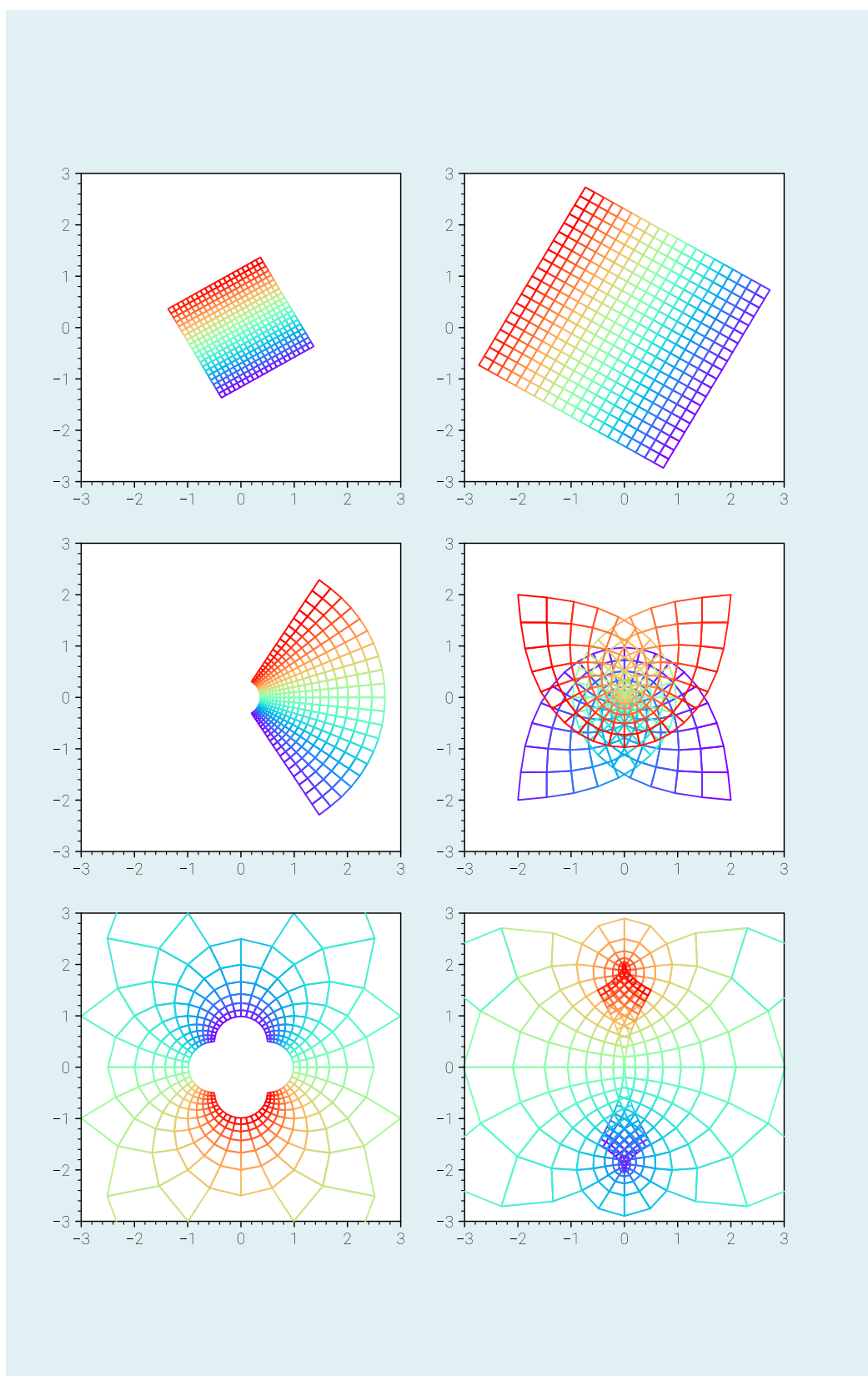


图 7. 用 pcolormesh 函数可视化线性、非线性变换

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