

Python 数据分析

Cloudera 大数据培训基地

重庆翰海睿智大数据科技有限公司

可视化图表类库 Matplotlib

可视化图表类库 Matplotlib

可视化图表类库 Matplotlib

在本章中，您将了解到：

1. Matplotlib 类库快速绘图
2. Figure 和 Subplot 对象
3. Matplotlib 类库基本功能
4. Pandas 绘图函数
5. Matplotlib 类库绘图

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

Matplotlib 简介

Matplotlib 是基于 numpy 的一套 Python 工具包。这个包提供了丰富的数据绘图工具，主要用于绘制一些统计图形。

Matplotlib API 函数都位于 matplotlib.pyplot 模块中。

引入方式: `import matplotlib.pyplot as plt`

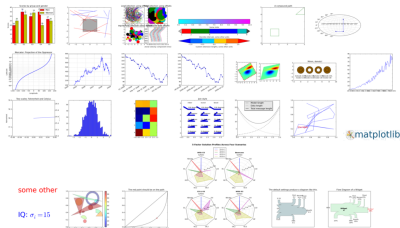


图 1: 'Matplotlib 图形集合'

绘制 2D 等高线图

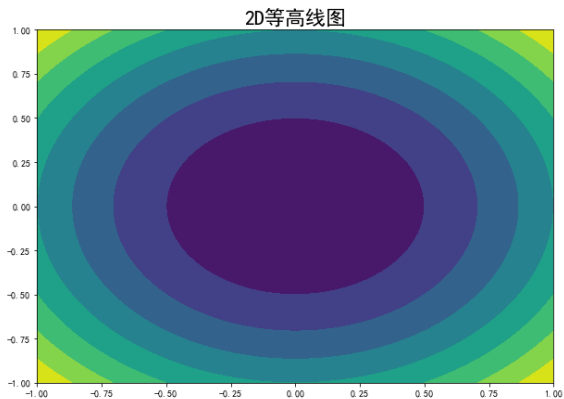


图 2: '2D 等高线图效果'

绘制 3D 曲面“瓦片”图

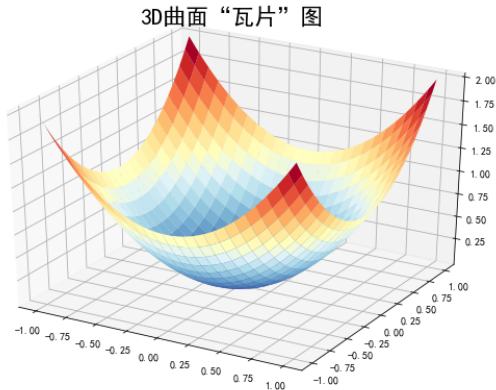


图 3: ‘3D 曲面“瓦片”图效果’

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

Figure 和 Subplot 的作用 (1)

在 Matplotlib 下，一个 Figure 对象可以包含多个子图 (Axes)，可以使用 subplot() 快速绘制。

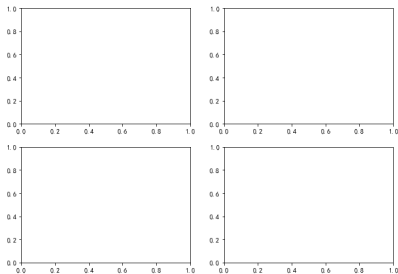


图 4: '绘制子图 1'

示例 1: 在第一个和最后一个进行绘制虚线和实线走势图

Figure 和 Subplot 的作用 (2)

根据特定布局创建 Figure 和 Subplot 是一件非常常见的任务，于是便出现了一个更为方便的方法 (`pyplot.subplots`)，它可以创建一个新的 Figure，并返回一个含有已创建的 Subplot 对象的 NumPy 数组。

`pyplot.subplots` 的选项

参数	说明
<code>nrows</code>	subplot 的行数
<code>ncols</code>	subplot 的列数
<code>sharex</code>	所有 subplot 应该使用相同的 X 轴刻度 (调节 <code>xlim</code> 将会影响所有 subplot)
<code>sharey</code>	所有 subplot 应该使用相同的 Y 轴刻度 (调节 <code>ylim</code> 将会影响所有 subplot)
<code>subplot_kw</code>	用于创建各 subplot 的关键字典
<code>**fig_kw</code>	创建 figure 时的其他关键字，如 <code>plt.subplots(2,2,figsize=(8,6))</code>

图 5: ‘subplots 选项’

调整多个 Subplot 之间的间距

默认情况下，Matplotlib 会在 subplot 外围自动产生一定的边距，并在 subplot 之间也会产生一定的间距。间距跟图像的高度和宽度有关，不管是手工还是编程调整图像大小间距也会自动调整。利用 Figure 的 `subplots_adjust` 方法可以修改间距。

语法：`plt.subplots_adjust(left=None, bottom=None, right=None, top=None, wspace=None, hspace=None)`

示例 3：间距的调整

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

颜色、线型和标记

Matplotlib 的 plot 函数接受一组 X 和 Y 的坐标，还可以接受一个表示颜色字符串或 RGB 值、线型的字符串缩写和线型标记 (强调数据点)。

示例 4：简单颜色和线型设置

示例 5：更为明确的颜色和线型设置

示例 6：简单标记设置

示例 7：更为明确的标记设置

示例 8：线型强调点连接

图例、刻度和轴标签

对于大多数图表而言，有时候我们需要装饰图表项。如图表的图例、图表的范围、刻度位置、轴标签等。

示例 9：为图表添加图例

示例 10：设置图表 X 和 Y 轴的范围

示例 11：设置图表的刻度位置

示例 12：设置图表 X 和 Y 轴的轴标签

设置图表标题和刻度、刻度标签、轴标签的其它方式

在制作图表过程中，我们有时候设置其标题。除了前面讲的刻度范围、标签和轴标签的方式外，还有其它方式也可以设置。

示例 13：设置图表标题及其它刻度和轴标签

示例 14：将刻度设置为其它值

除了标准图表对象外，有时候希望在图表上绘制一些自定义的注解。

`annotate()` 方法提供辅助函数，使标注变得容易。在标注中，有两个要考虑的点：由参数 `xy` 表示的标注位置和 `xytext` 的文本位置。这两个参数都是 `(x, y)` 元组。

示例 15： 绘制简单的注解

示例 16： 绘制世界金融危机重要日期注解

将图表保存为图片

利用 `plt.savefig` 可以将当期图表保存到文件。该方法相当于 Figure 对象的实例方法 `savefig`。

参数	说明
<code>fname</code>	含有文件路径的字符串或Python的文件型对象。图像格式由文件扩展名推断得出，例如， <code>.pdf</code> 推断出PDF， <code>.png</code> 推断出PNG
<code>dpi</code>	图像分辨率（每英寸点数），默认为100
<code>facecolor</code> 、 <code>edgecolor</code>	图像的背景色，默认为“w”（白色）
<code>format</code>	显式设置文件格式（“png”、“pdf”、“svg”、“ps”、“eps”……）
<code>bbox_inches</code>	图表需要保存的部分。如果设置为“tight”，则将尝试剪除图表周围的空白部分

图 6: ‘savefig 的选项’

示例 17: 保存世界金融危机的图表

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

为什么使用 Pandas 绘图函数

Matplotlib 实际上是一种比较低级的工具。要组装一张图表，需要用到它的各种基础组件才行：数据展示（线形图、柱状图、盒形图、散布图、等值线图）、图例、标题、刻度标签以及其他注解信息。

那么在 Pandas 中，有行标签、列标签以及分组信息（可能有）。这也就是说，要制作一张完整的图表、原本需要大堆的 Matplotlib 代码，现在只需要一两条简洁的语句就可以了。Pandas 有许多能够利用 DataFrame 对象数据组织特点来创建标准图表的高级绘图方法（这些函数的数量还在不断增加）。

Series 和 DataFrame 对象的绘图方法 (1)

Series 和 DataFrame 都有一个用于生成各类图表的 plot 方法。默认情况下，它们所生成的是线型图。

参数	说明
label	用于图例的标签
ax	要在其上绘制matplotlib subplot对象。如果没有设置，则使用当前matplotlib subplot
style	将要传给matplotlib的风格字符串（如'ko--'）
alpha	图表的填充不透明度（0到1之间）

图 7: 'Series.plot 方法的参数'

Series 和 DataFrame 对象的绘图方法 (2)

参数	说明
kind	可以是'line'、'bar'、'barh'、'kde'
logy	在Y轴上使用对数标尺
use_index	将对象的索引用作刻度标签
rot	旋转刻度标签（0到360）
xticks	用作X轴刻度的值
yticks	用作Y轴刻度的值
xlim	X轴的界限（例如[0, 10]）
ylim	Y轴的界限
grid	显示轴网格线（默认打开）

图 8: 'Series.plot 方法的参数'

Series 和 DataFrame 对象的绘图方法 (3)

参数	说明
subplots	将各个DataFrame列绘制到单独的subplot中
sharex	如果subplots=True, 则共用同一个X轴, 包括刻度和界限
sharey	如果subplots=True, 则共用同一个Y轴
figsize	表示图像大小的元组
title	表示图像标题的字符串
legend	添加一个subplot图例 (默认为True)
sort_columns	以字母表顺序绘制各列, 默认使用当前列顺序

图 9: 'Series.plot 方法的参数'

Series 和 DataFrame 对象的绘图方法 (4)

示例 18: Series 对象绘制简单的线形图

示例 19: DataFrame 对象绘制简单的线形图

示例 20: Series 对象绘制简单的垂直和水平柱状图

示例 21: DataFrame 对象绘制简单的垂直柱状图

Series 和 DataFrame 对象的绘图方法 (5)

示例 22: 餐厅聚餐数据堆积柱状图

示例 23: 餐厅小费与消费总金额占比数据直方图

示例 24: 餐厅小费与消费总金额占比数据密度图

示例 25: 直方图和密度图组成双峰分布图

示例 26: 绘制散布图

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

使用 Matplotlib 进行作图

示例 27: 绘制误差条形图

示例 28: 绘制饼图

示例 29: 绘制等高线图

示例 30: 绘制 3D 直方图

可视化图表类库 Matplotlib

Matplotlib 类库快速绘图

Figure 和 Subplot

Matplotlib 类库基本功能

Pandas 绘图函数

Matplotlib 类库绘图

结论

基本要点

- Matplotlib 类库快速绘图
- 绘制 2D 等高线图、绘制 3D 曲面“瓦片”图
- Figure 和 Subplot 对象
- Figure 和 Subplot 的作用
- 调整多个 Subplot 之间的间距
- Matplotlib 类库基本功能
- 颜色、线型和标记、图例、刻度和轴标签
- 设置图表标题和刻度、刻度标签、轴标签的其它方式
- 绘制注解、将图表保存为图片
- Pandas 绘图函数
- Series 和 DataFrame 对象的绘图方法 (线形图、垂直和水平柱状图、堆积柱状图、直方图、密度图、散布图)
- Matplotlib 类库绘图