

数据可视化



纪慧诚

CFA FRM RFP

金程教育资深培训讲师

CONTENTS

PROFESSIONAL · LEADING · VALUE-CREATING

▶ PART 1

Pandas 自带图形绘制

▶ PART 2

Matplotlib 应用

▶ PART 3

Seaborn 应用

专业来自101%的投入!



➤ Matplotlib

- 用于创建出版质量图表的绘图工具库
- 目的是为Python构建一个Matlab式的绘图接口
- `import matplotlib.pyplot as plt`
- pyploy模块包含了常用的matplotlib API函数



➤ 颜色、标记、线型

- `ax.plot(x, y, 'r--')`
- 等价于 `ax.plot(x, y, linestyle='--', color='r')`

颜色

- b: blue
- g: green
- r: red
- c: cyan
- m: magenta
- y: yellow
- k: black
- w: white

标记

marker	description
"."	point
"."	pixel
"o"	circle
"v"	triangle_down
"^"	triangle_up
"<"	triangle_left

线型

linestyle	description
'-' or 'solid'	solid line
'--' or 'dashed'	dashed line
'-.' or 'dashdot'	dash-dotted line
':' or 'dotted'	dotted line
'None'	draw nothing
' '	draw nothing
' '	draw nothing



➤ **Plt常用功能：**

- `Plt.plot()`：绘制图形；
- `Plt.show()`：图形显示；
- `Plt.figure()`：生成图形对象；
- `Plt.title()`：确定图片标题；
- `Plt.xlabel()`：确定图片x轴；
- `Plt.legend()`：显示图例；
- `Plt.scatter()`：绘制散点图；
- `Plt.grid(True)`：出现网格；
- `Plt.subplot()`：绘制子图；
- `Plt.bar()`：绘制柱状图；



- 你可以用子图来将图样 (plot) 放在均匀的坐标网格中。用 `subplot` 函数的时候，你需要指明网格的行列数量，以及你希望将图样放在哪一个网格区域中。

`subplot(2,1,1)`

`subplot(2,1,2)`

`subplot(1,2,1)`

`subplot(1,2,2)`



➤ 类似的：

`subplot(2,2,1)`

`subplot(2,2,2)`

`subplot(2,2,3)`

`subplot(2,2,4)`

CONTENTS

PROFESSIONAL · LEADING · VALUE-CREATING

▶ PART 1

Pandas 自带图形绘制

▶ PART 2

Matplotlib 应用

▶ PART 3

Seaborn 应用

专业来自101%的投入!



➤ 什么是Seaborn

- Python中的一个制图工具库，可以制作出吸引人的、信息量大的统计图
- 在Matplotlib上构建，支持numpy和pandas的数据结构可视化，甚至是scipy和statsmodels的统计模型可视化

➤ 特点

- 多个内置主题及颜色主题
- 可视化单一变量、二维变量用于比较数据集中各变量的分布情况
- 可视化线性回归模型中的独立变量及不独立变量



➤ 特点(续)

- 可视化矩阵数据，通过聚类算法探究矩阵间的结构
- 可视化时间序列数据及不确定性的展示
- 可在分割区域制图，用于复杂的可视化

➤ 安装

- Conda install seaborn
- pip install seaborn



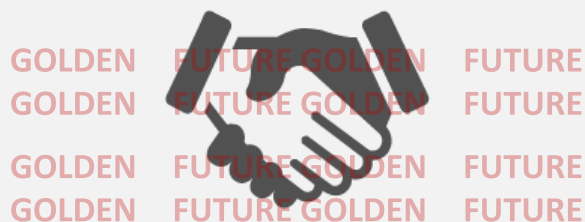
➤ 数据集分布可视化

- 单变量分布 `sns.distplot()`
 - ✓ 直方图 `sns.distplot(kde=False)`
 - ✓ 核密度估计 `sns.distplot(hist=False)` 或 `sns.kdeplot()`
 - ✓ 拟合参数分布 `sns.distplot(kde=False, fit=)`
- 双变量分布
 - ✓ 散布图 `sns.jointplot()`
 - ✓ 二维直方图 `Hexbinsns.jointplot(kind='hex')`
 - ✓ 核密度估计 `sns.jointplot(kind='kde')`
- 数据集中变量间关系可视化 `sns.pairplot()`



金程教育
GOLDEN FUTURE

Thank you!



专业来自101%的投入!