

小提琴图

Guangyao Zhao

2023-02-09

Contents

Salinity intrusiton 项目的特征分布可视化。

```
1 import os
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 # toc: 全局设置
7 config = {"font.size": 16}
8 plt.rcParams.update(config)
9
10 # toc: 读取文件, 准备数据
11 dir_path = os.path.dirname(__file__) # directory path
12 file_path = os.path.join(dir_path, "Dataset_24.csv")
13 dataset = pd.read_csv(file_path, header=[0], index_col=[0]).iloc[:, :-2]
14 feature_ranking = [
15     "Water temperature",
16     "Water level 1",
17     "Tide",
18     "Temperature",
19     "Sea level pressure",
20     "Wind speed",
21     "Atmospheric pressure",
```

```

22     "Water level 2",
23     "Precipitation",
24 ] # 想要的特征排序
25 dataset = dataset[feature_ranking] # 特征排序后的数据
26
27
28 # toc: figure
29 subplots_num = len(feature_ranking) # 子图个数
30 num_wid = 3 # 横向子图个数
31 num_hei = int(subplots_num / num_wid) # 纵向子图个数
32 fig = plt.figure(figsize=(8 * num_wid, 6 * num_hei))
33 axes = fig.subplots(nrows=num_hei, ncols=num_wid)
34 for i in range(num_hei): # 按横向画图, 第 i 行
35     for j in range(num_wid): # 第 i 行的第 j 个子图
36         tmp = i * num_wid + j # 第 tmp 个子图
37         tmp_data = dataset[feature_ranking[tmp]] # 当前特征的数据
38         violin = axes[i, j].violinplot(
39             dataset=tmp_data,
40             vert=False,
41             showextrema=False,
42         )
43         for patch in violin[
44             "bodies"
45         ]: # A list of the PolyCollection instances containing the filled area of each violin http://m
46             patch.set_facecolor((232 / 255, 125 / 255, 115 / 255)) # 填充色
47             patch.set_edgecolor("black") # 边框颜色
48             patch.set_alpha(1) # 饱和度
49
50         quantile1, median, quantile3 = np.percentile(tmp_data, [25, 50, 75]) # 选择要画的分位数
51         IQR = quantile3 - quantile1
52         left_IQR, right_IQP = quantile1 - 1.5 * IQR, quantile3 + 1.5 * IQR # 置信区间
53         axes[i, j].scatter(median, 1.0, color="white", zorder=4) # 中位数
54         axes[i, j].hlines(y=1.0, xmin=quantile1, xmax=quantile3, lw=9, zorder=3) # 四分位
55         axes[i, j].hlines(y=1.0, xmin=left_IQR, xmax=right_IQP, lw=2, zorder=2)
56
57 # 图形设定

```

```
58     axes[i, j].set_xlabel(feature_ranking[tmp])
59     axes[i, j].tick_params(axis="x", which="major", direction="inout")
60     axes[i, j].tick_params(axis="y", which="both", left=False)
61     axes[i, j].set_yticklabels([]) # 隐藏标签
62
63 # toc: 保存
64 fig_path = os.path.join(
65     dir_path.split("Dataset")[0], "Figure/Distribution", "Violin.pdf"
66 )
67 fig.savefig(fig_path)
68
69 fig_path = os.path.join(dir_path, "Violin.pdf")
70 fig.savefig(fig_path)
```