(26条消息) 用Python为直方图绘制拟合正态分布曲线的两种方法_jiangjiane-CSDN博客_python 正态分布拟合

2-3 minutes

直方图是用于展示数据的分组分布状态的一种图形，用矩形的宽度和高度表示频数分布，通过直方图，用户可以很直观的看出数据分布的形状、中心位置以及数据的离散程度等。

在python中一般采用matplotlib库的hist来绘制直方图，至于如何给直方图添加拟合曲线（密度函数曲线），一般来说有以下两种方法。

方法一：采用matplotlib中的mlab模块

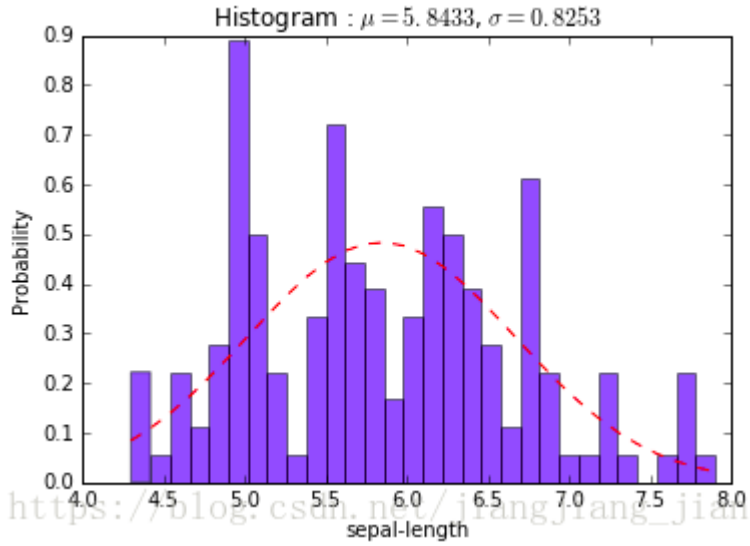
mlab模块是Python中强大的3D作图工具，立体感效果极佳。在这里使用mlab可以跳出直方图二维平面图形的限制，在此基础上再添加一条曲线。在这里，我们以鸢尾花iris中的数据为例，来举例说明。

```
1.
2. import matplotlib.mlab as mlab
3. import matplotlib.pyplot as plt
4.
5.
6. url = "https://archive.ics.uci.edu/ml/machine-learning-
   databases/iris/iris.data"
7. names = ['sepal-length', 'sepal-width','petal-length', 'petal-width',
   'class']
8. dataset = pandas.read_csv(url, names=names)
9.
10.
11. print(dataset.describe())
12.
13.
```

```
14.
15.
```

以上为通过python导入鸢尾花iris数据，然后提取第一列的sepal-length变量为研究对象，计算出其均值、标准差，接下来就绘制带拟合曲线的直方图。

```
1.
2. n, bins, patches = plt.hist(x, num_bins,normed=1, facecolor='blue',
   alpha=0.5)
3.
4. y = mlab.normpdf(bins, mu, sigma)
5.
6. plt.xlabel('sepal-length')
7. plt.ylabel('Probability')
8. plt.title(r'Histogram : $\mu=5.8433$,$\sigma=0.8253$')
9. plt.subplots_adjust(left=0.15)
10.
```

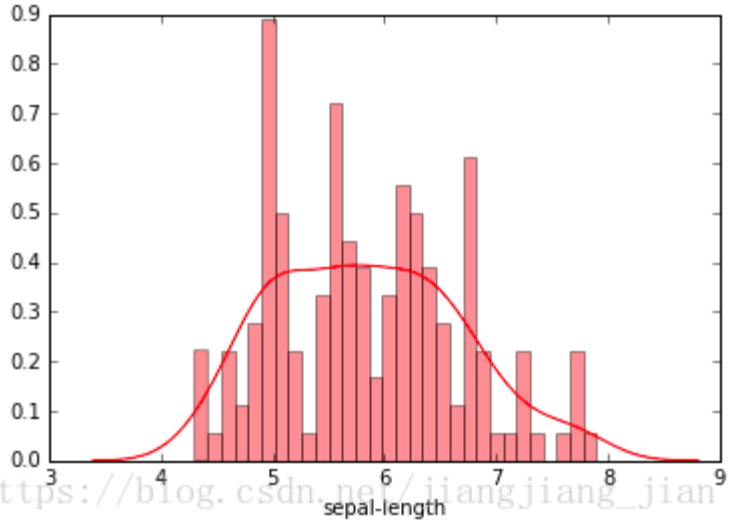


以上命令主要采用mlab.normpdf基于直方图的柱子数量、均值、方差来拟合曲线，然后再用plot画出来，这种方法的一个缺点就是画出的正态分布拟合曲线（红色虚线）并不一定能很好反映数据的分布情况，如上图所示。

方法二：采用seaborn库中的distplot绘制

Seaborn其实是在matplotlib的基础上进行了更高级的API封装，从而使得作图更加容易，在大多数情况下使用seaborn就能做出很具有吸引力的图，而使用matplotlib就能制作具有更多特色的图。应该把Seaborn视为matplotlib的补充，而不是替代物。

```
1.
2.
3. sns.distplot(x,color="r",bins=30,kde=True)
4.
```



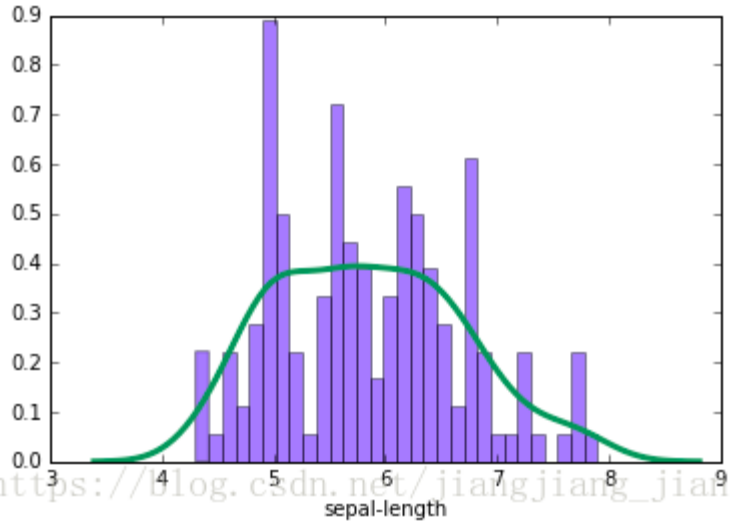
在这里主要使用sns.distplot（增强版dist），柱子数量bins也设置为30，kde=True表示是否显示拟合曲线，如果为False则只出现直方图。

在这里注意一下它与前边mlab.normpdf方法不同的是，拟合曲线不是正态的，而是更好地拟合了数据的分布情况，如上图，因此比mlab.normpdf更为准确。

进一步设置sns.distplot，可以采用kde_kws（拟合曲线的设置）、hist_kws（直方柱子的设置），可以得到：

```
1.
2.
3.
4. mpl.rc("figure", figsize=(6,4))
```

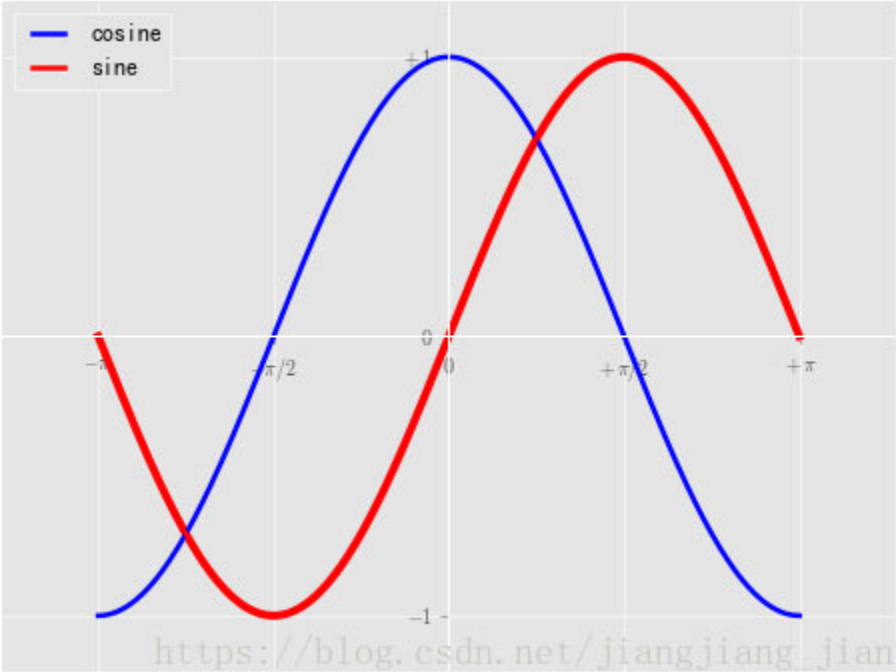
```
5. sns.distplot(x,bins=30,kde_kws={"color":"seagreen", "lw":3 }, hist_kws={
   "color": "b" })
6.
```



其中，lw为曲线粗细程度。

可使用plt.legend()添加图例

```
1. plot(X, C, color="blue", linewidth=2.5, linestyle="-", label="cosine")
2. plot(X, S, color="red",  linewidth=2.5, linestyle="-", label="sine")
3.
4.
```


cosine
sine
+1
0
−1
+π/2
+π