

气象数据分析_数据可视化&线性回归

笔记本: python

创建时间: 2020/4/17 9:31

更新时间: 2020/4/17 11:25

作者: 来碗芋泥

URL: <https://notebook.shiyanlou.com/land-eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJ...>

库:

matplotlib 库: 对数据进行图表化处理

scikit-learn 库: SVM 库: 对数据进行回归分析

数据分析: 把原始数据转化为信息, 再把信息转化为知识

问题: 海洋对一个地区的气候有何影响

第一步, 选(能代表整个平原地区的) 10 个城市作为参照组。其中 5 个城市在距海 100 公里范围内, 其余 5 个距海 100~400 公里

第二步, 计算城市间距离: [TheTimeNow 网站](http://www.thetimenow.com)

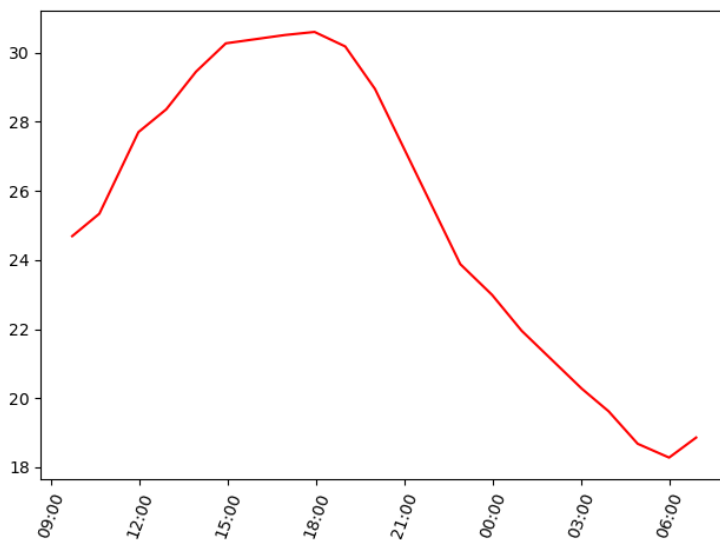
第三步, 获取数据源: <http://openweathermap.org/>

实验:

温度数据分析

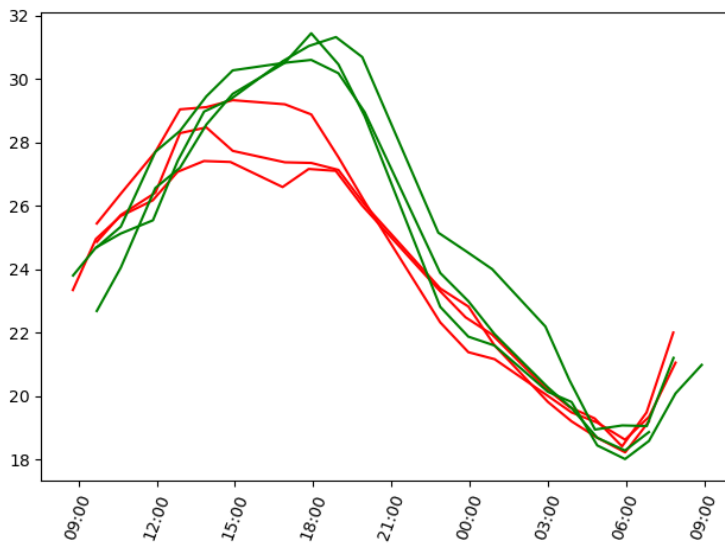
```
weather_milano.py
```

以米兰为例, 分析一天中气温的变化趋势



```
weather_3jin3yuan.py
```

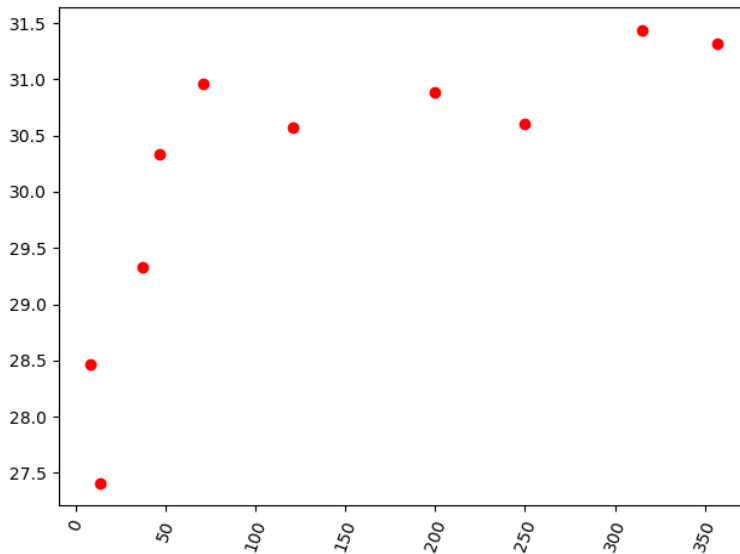
三个离海最近以及三个离海最远的城市, 分析一天中气温的变化趋势



如图所示。离海最近的三个城市的最高气温比离海最远的三个城市低不少，而最低气温看起来差别较小。

```
weather_10max.py
```

```
# 收集 10 个城市的最高温和最低温，用线性图表示气温最值点和离海远近之间的关系
# 最高温变化趋势与离海远近之间的关系
```

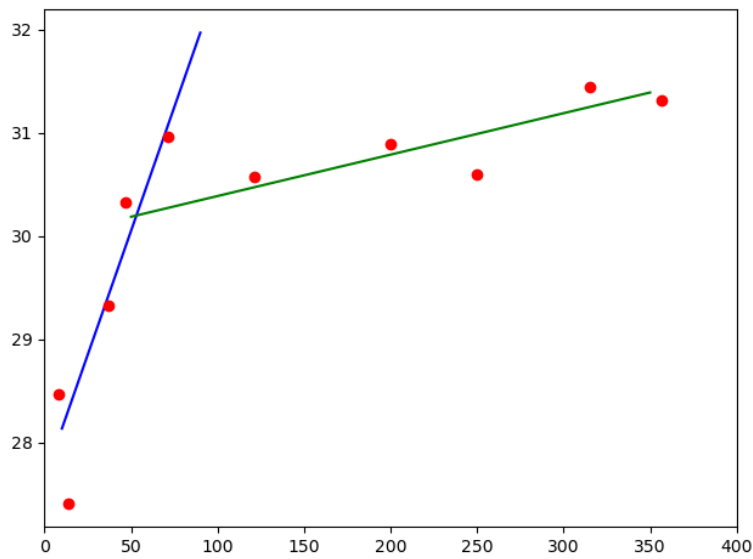


如图所示。可以证实，海洋对气象数据具有一定程度的影响这个假设是正确的（至少这一天如此）

海洋的影响衰减得很快，离海 60 ~ 70 公里开外，气温就已攀升到高位

```
weather_regression.py
```

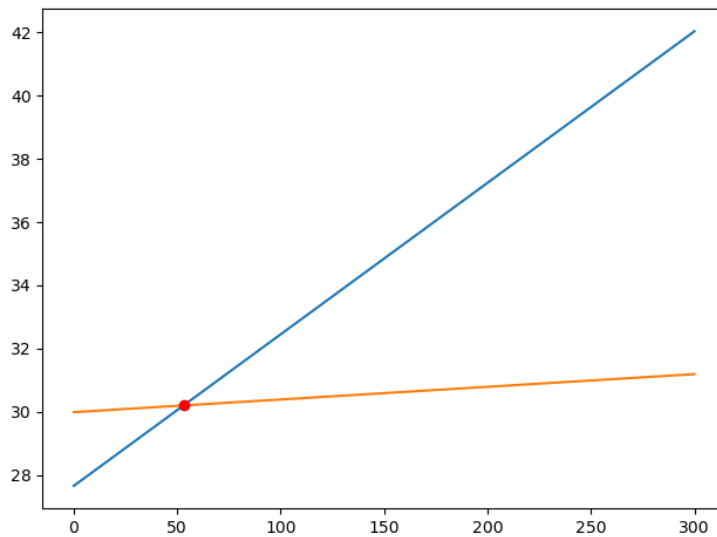
```
# 用线性回归算法得到两条直线，分别表示两种不同的气温趋势
# 我们可以使用 scikit-learn 库的 SVR 方法
```



如图所示。离海 60 公里以内，气温上升速度很快，从 28 度陡升至 31 度，随后增速渐趋缓和（如果还继续增长的话），更长的距离才会有小幅上升。这两种趋势可分别用两条直线来表示，直线的表达式为： $y=ax+b$ 。其中 a 为斜率， b 为截距。

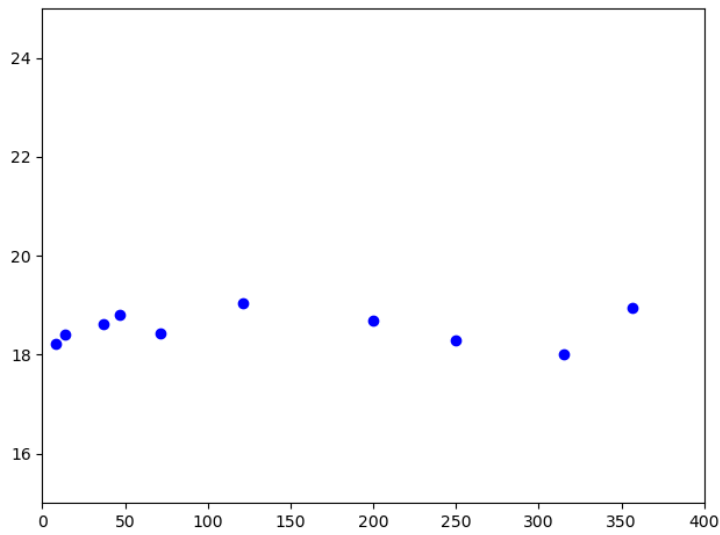
```
weather_regression_jiaodian.py
```

找到由线性回归所得到的两条直线的交点
将这两条直线的交点作为受海洋影响和不受海洋影响的区域的分界点，或者至少是海洋影响较弱的分界点



交点坐标: $[x,y]=[53,30]$
可以说海洋对气温产生影响的平均距离（该天的情况）为 53 公里。
同理分析最低气温。

```
weather_10min.py
```

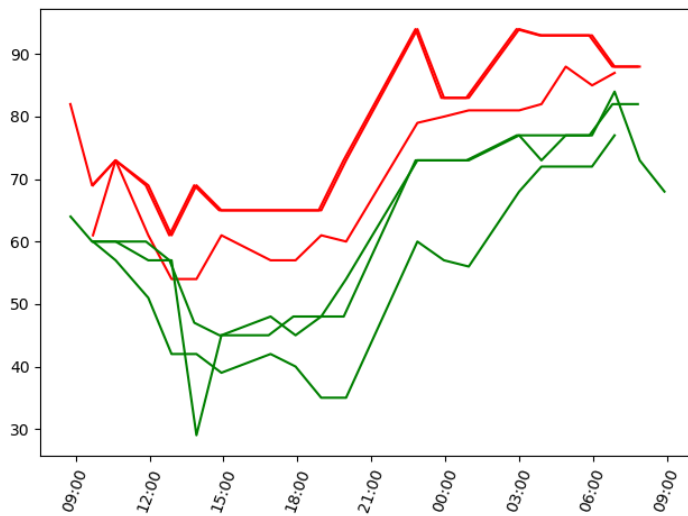


最低温几乎与海洋无关

湿度数据分析

wet.py

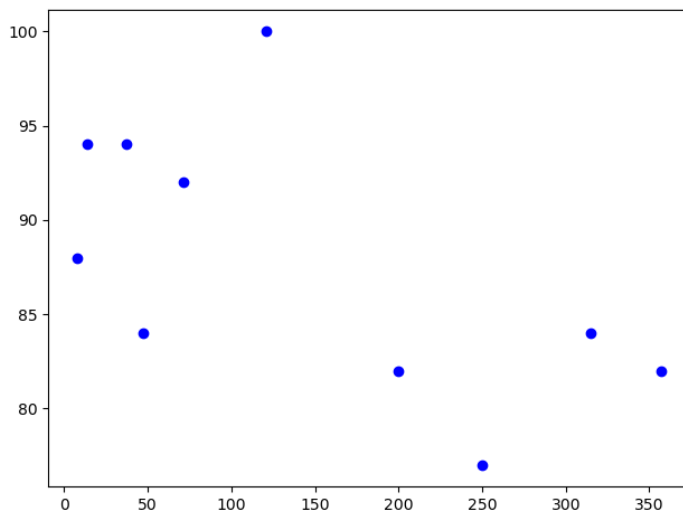
当天三个近海城市和三个内陆城市的湿度趋势



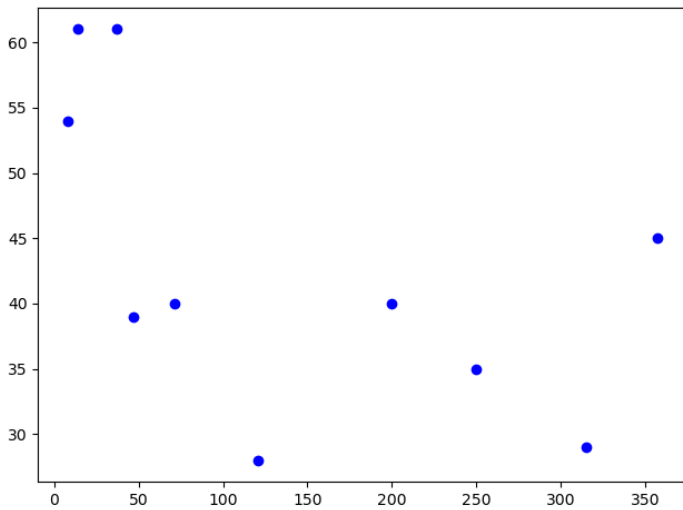
好像近海城市的湿度要大于内陆城市，全天湿度差距在 20% 左右
再来看一下湿度的极值和离海远近之间的关系，是否跟我们的第一印象相符

wetji.py

湿度的极值和离海远近之间的关系
最大湿度



最小湿度



如图所示。近海城市无论是最大还是最小湿度都要高于内陆城市。
然而,还不能说湿度和距离之间存在线性关系或者其他能用曲线表示的关系。
采集的数据点数量 (10) 太少, 不足以描述这类趋势。

风向频率玫瑰图

```
wind.py
```

极区图: 创建一个直方图, 也就是将 360 度分为八个面元, 每个面元为 45 度, 把所有的数据点分到这八个面元中

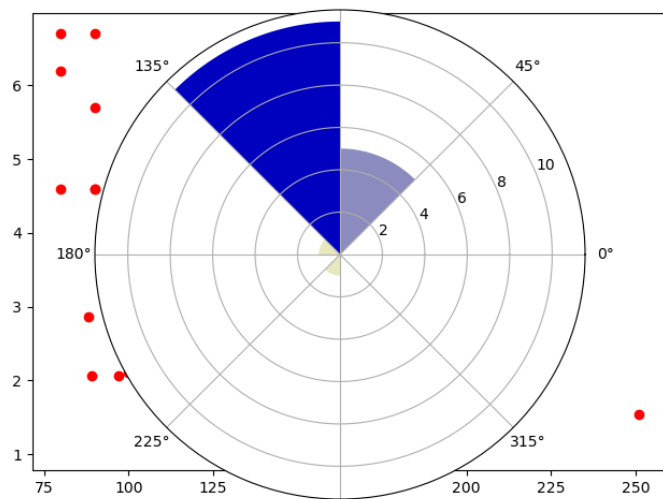
绘制极区图:

showRoseWind(values,city_name,max_value)函数:

values 数组, 指的是想为其作图的数据, 也就是这里的 hist 数组;

city_name 为字符串类型, 指定图表标题所用的城市名称;

max_value 为整型, 指定最大的蓝色值



以得知风向在极坐标系中的分布方式。
该图表示这一天大部分时间风都吹向西南和正西方向。

计算风速均值的分布情况：

RoseWind Speed(df city)函数：

