

基于 Python 爬虫和机器学习的西安市天气数据分析

葛宇辰 2018300314

(西北工业大学 航空学院)

1 源码说明

表 1 源代码文件介绍

源代码文件	功能	行数	备注
weather_monthly.py	爬取数据	64	爬取 2011 年-2020 年西安市每月平均天气数据 (数据集大小: 120)
weather_daily.py	爬取数据	85	爬取 2021 年 11 月 1 日-21 日西安市每日天气数据 (数据集大小: 20)
AQI_daily.py	爬取数据	119	爬取 2016 年-2020 年西安市每天空气质量指数及各污染物的数据 (数据集大小: 1786)
month_data.py	数据处理	39	处理爬虫得到的每月天气数据
day_data.py	数据处理	34	处理爬虫得到的每日天气数据
AQI_data.py	数据处理	19	处理爬虫得到的空气质量数据
month_plot.py	数据可视化	56	实现月平均气温、平均空气质量指数和每年 11 月气温的可视化
day_plot.py	数据可视化	56	实现每日最高温、最低温、风力等级、空气质量指数的可视化
AQI_plot.py	数据可视化	25	实现 AQI、AQI 排名、PM2.5、PM10、SO2、NO2、CO、O3 数据可视化
month_regression.py	数据挖掘	55	回归分析平均高温和平均低温并可视化
day_fit.py	数据挖掘	74	拟合最高温、最低温、风力等级、空气质量指数
AQI_regression.py	数据挖掘	56	采用随机森林模型回归分析 6 种污染物浓度与空气质量指数的关系
regression_model.py	机器学习	73	调用 sklearn 库中的 9 种回归分析模型并可视化

注：代码行数共计 755 行

2 源码逻辑

2.1 机器学习模型

通过 **regression_model.py** 将实现回归分析的计算和可视化的函数加以封装，回归模型包括决策树回归、线性回归、SVM 回归、KNN 回归、随机森林回归、Adaboost 回归、GBRT 回归、Bagging 回归、极端随机树回归共 9 种模型。

使用的库包括：

- (1) **matplotlib**: 数据可视化
- (2) **numpy**: 科学计算
- (3) **sklearn**: 机器学习

封装函数包括：

- (1) **regression_model_compute(model,X_train,X_test,y_train,y_test)**: 计算回归模型的结果
- (2) **regression_model_plot(i,prediction,score,y_test)**: 实现第 i 个回归模型实际值和预测值的可视化
- (3) **regression_model_compare(X_train,X_test,y_train,y_test)**: 比较 9 种回归模型

2.2 每月天气数据

2.2.1 数据爬取

通过 **weather_monthly.py** 实现对 2011 年-2020 年西安市每月平均天气数据的爬取，其中天气数据包括：月平均高温，月平均低温，月平均空气质量指数，共 120 个数据。其中月平均空气质量指数数据存在缺失，一直到 2018 年以后才有数据，结果保存在./result_data/ data_monthly.csv 文件中。

使用的库包括：

- (1) **requests**: 爬取网页
- (2) **BeautifulSoup**: 解析网页
- (3) **Pandas**: 处理数据

封装函数包括：

- (1) **format_date(year,month)**: 格式化日期
- (2) **all_date(years)**: 输出要分析的日期

(3) all_html(years): 保存所有日期所在网页的超链接

(4) get_html(url): 爬取网页

(5) get_data(years): 解析所有网页

运行截图如下：

```
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d;; cd 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine'; & 'C:\Users\GVC\Anaconda3\envs\weather\Scripts\python.exe' 'c:\Users\GVC\vscode\extensions\ms-python.python-2021.11.1422169775\pythonFiles\lib\python\debugpy\launcher' '2315' '--' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\weather_monthly.py'
analyzing:1/120
analyzing:2/120
analyzing:3/120
analyzing:4/120
analyzing:5/120
analyzing:6/120
analyzing:7/120
analyzing:8/120
analyzing:9/120
analyzing:10/120
analyzing:11/120
analyzing:12/120
analyzing:13/120
analyzing:14/120
analyzing:15/120
```

2.2.2 数据处理

通过 **month_data.py** 实现对爬取的每月平均天气数据进行处理，其中包括截取平均空气质量指数的缺失值、字符串保存的数据转化为整型数据，提取每年同一个月的数据。

使用的库包括：

(1) Pandas: 处理数据

封装函数包括：

(1) select_month(item,month): 提取每年某一个月的数据

2.2.3 数据可视化

通过 **month_plot.py** 实现对 **month_data.py** 处理后的每月平均天气数据进行可视化处理。

使用的库包括：

(1) matplotlib: 数据可视化

封装函数包括：

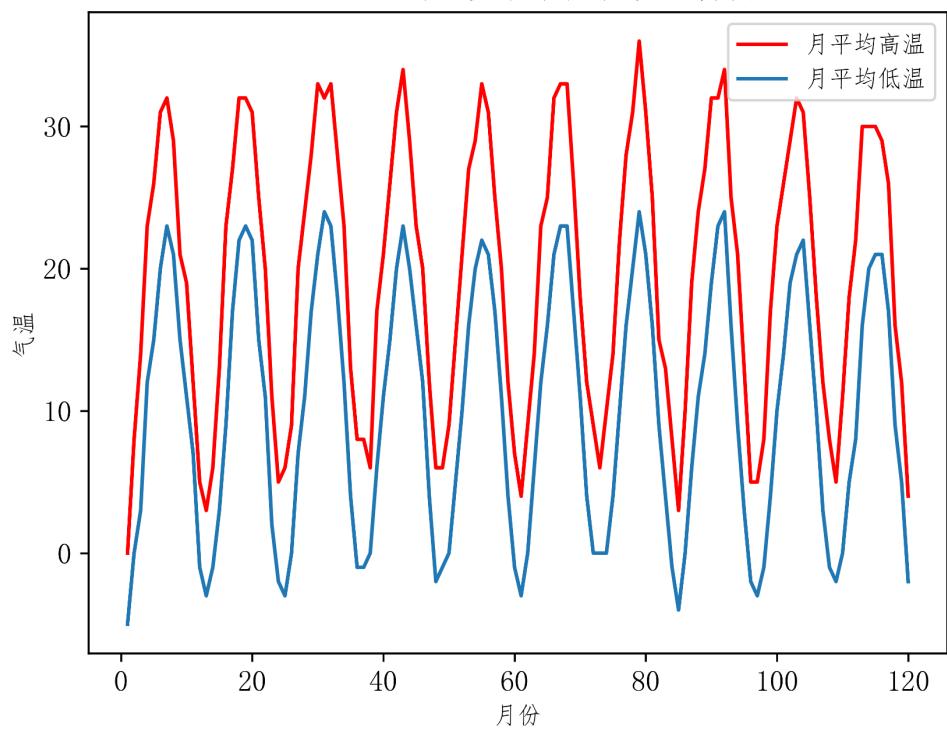
(1) plot_temperature(): 绘制气温走势图

(2) plot_AQI(): 绘制平均空气质量走势图

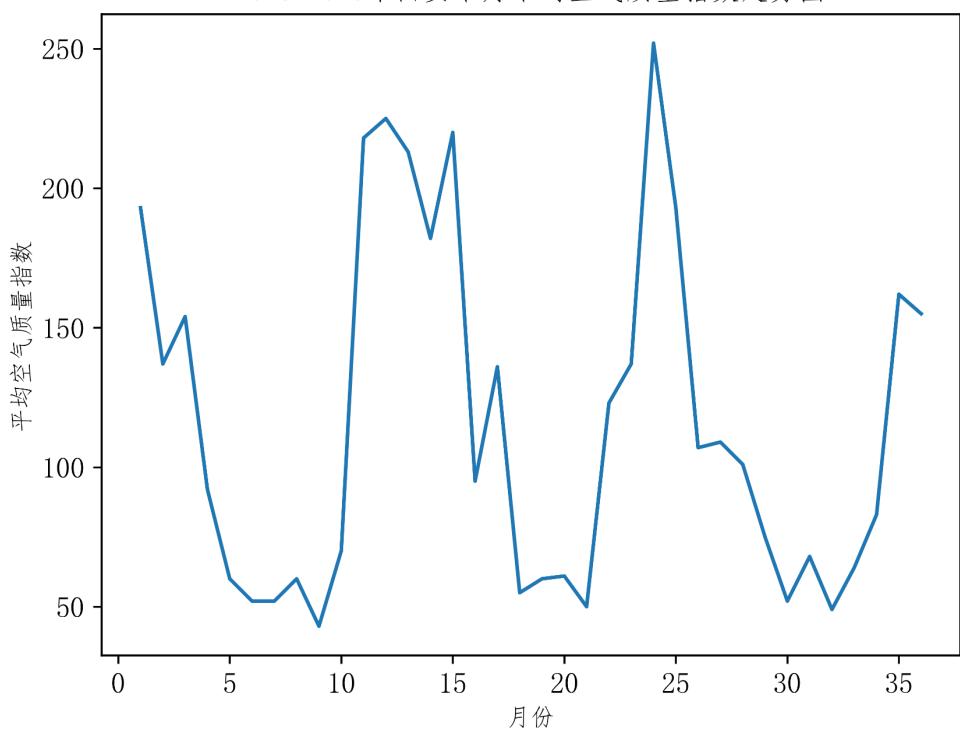
(3) plot_November_temperature(): 绘制 10 年间每年 11 月的气温走势图

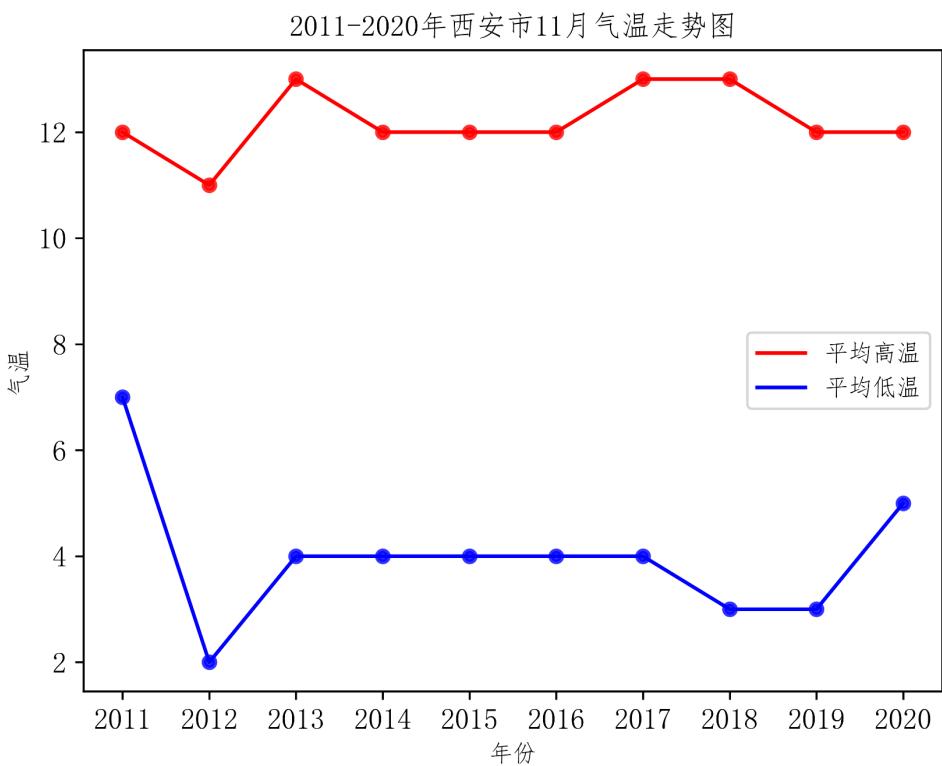
结果保存在./result_figure/monthly/plot 文件夹下，包括：

2011-2020年西安市月平均气温走势图



2018-2020年西安市月平均空气质量指数走势图





2.2.4 数据回归

通过 `month_regression.py` 实现对 120 个月的平均气温进行回归分析，调用了 `month_data.py` 中的数据和 `regression_model.py` 中的回归分析模型及函数。

使用的库包括：

- (1) `matplotlib`: 数据可视化
- (2) `numpy`: 科学计算
- (3) `Pandas`: 处理数据
- (4) `sklearn`: 机器学习

封装函数包括：

- (1) `load_data(amount, variate)`: 导入天气数据
- (2) `regression_weather(date, weather, name)`, `regression_model_compare(X_train, X_test, y_train, y_test)`: 回归天气数据。

计算结果如图所示：

以下是对月平均高温的回归:

训练集大小: 90
测试集大小: 30

model	score	Mean Absolute Error	Median Absolute Error
0	决策树回归	0.592877	5.100000
1	线性回归	-0.099619	8.173043
2	SVM回归	-0.319189	8.936334
3	KNN回归	0.396801	5.366667
4	随机森林回归	0.775228	3.530000
5	Adaboost回归	-0.152438	8.231005
6	GBRT回归	0.592237	5.103143
7	Bagging回归	0.700231	4.113333
8	极端随机树回归	0.527565	5.466667

回归模型性能最好的是:[4]随机森林回归

以下是对月平均低温的回归:

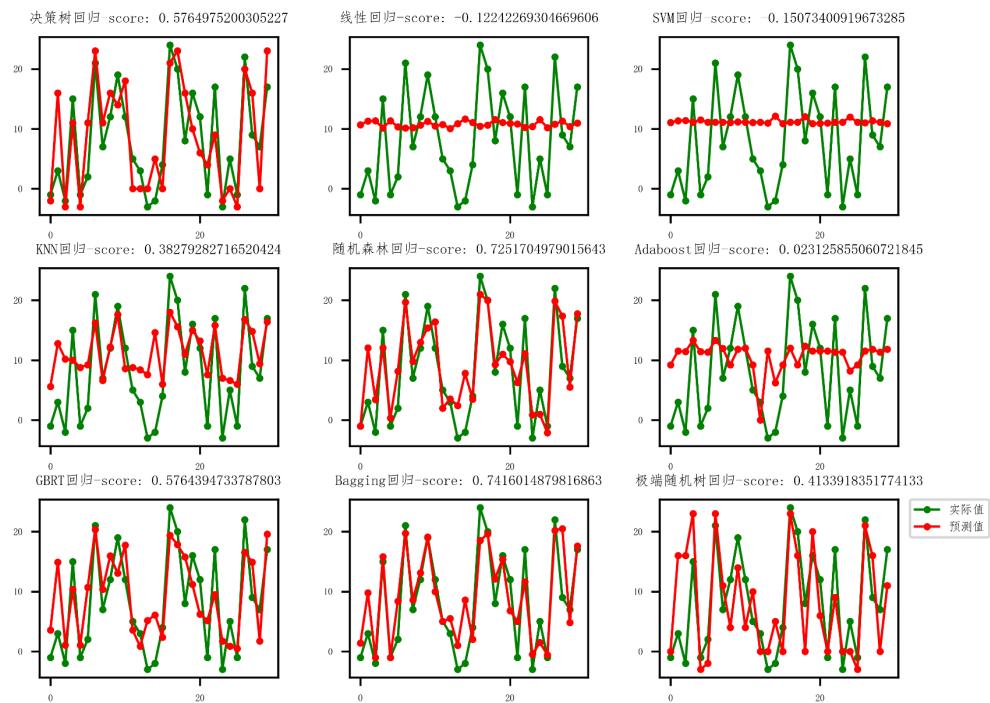
训练集大小: 90
测试集大小: 30

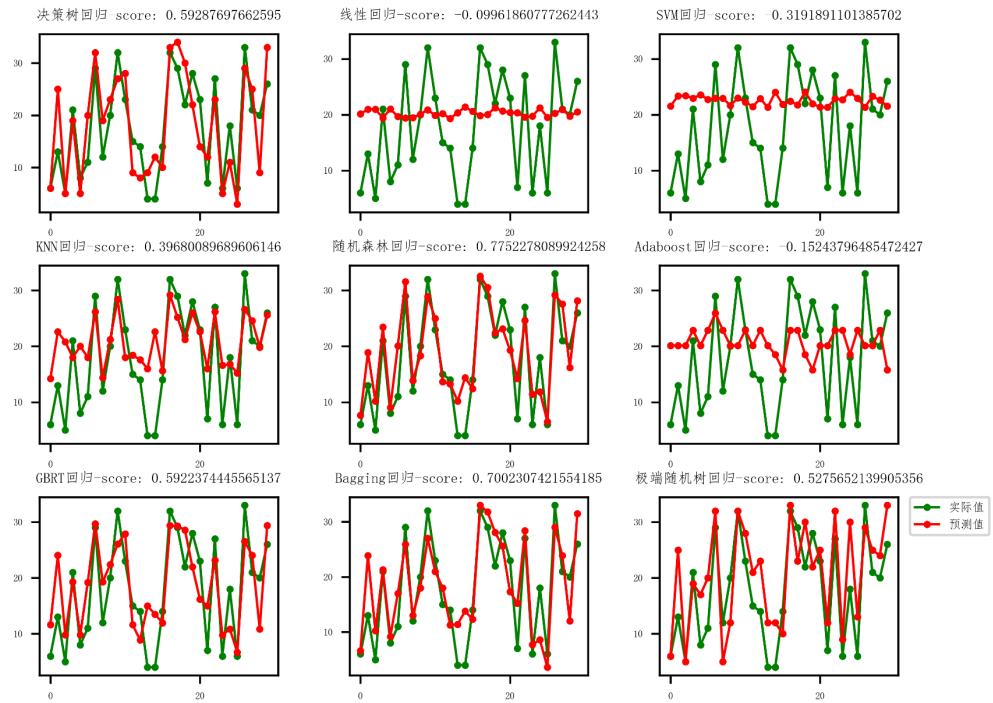
model	score	Mean Absolute Error	Median Absolute Error
0	决策树回归	0.576498	4.733333
1	线性回归	-0.122423	7.867326
2	SVM回归	-0.150734	7.957773
3	KNN回归	0.382793	5.240000
4	随机森林回归	0.725170	3.461667
5	Adaboost回归	0.023126	7.140112
6	GBRT回归	0.576439	4.812597
7	Bagging回归	0.741601	3.076667
8	极端随机树回归	0.413392	5.266667

回归模型性能最好的是:[7]Bagging回归

```
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> █
```

可视化结果保存在./result_figure/monthly/regression 文件夹下，包括:





2.3 每日天气数据

2.3.1 数据爬取

通过 `weather_daily.py` 实现对 2021 年 11 月 1 日-21 日西安市每日天气数据的爬取，其中包括最高温、最低温、风力等级、空气质量指数，共 20 个数据，结果保存在 `./result_data/ data_daily.csv` 文件中。

使用的库包括：

- (1) `requests`: 爬取网页
- (2) `BeautifulSoup`: 解析网页
- (3) `Pandas`: 处理数据

封装函数包括：

- (1) `get_html(url)`: 爬取网页
- (2) `get_data(html)`: 解析所有网页

运行截图如下：

```

PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d:; cd 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine'; & 'C:\Users\GYC\AppData\Local\Programs\Python\Python310\python.exe' 'c:\Users\GYC\.vscode\extensions\ms-python.python-2021.11.142216975\pythonfiles\lib\python\debugpy\launcher' '2398' '-l' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\weather_daily.py'
status code: 200
[[{"日期": "最高温", "最低温", "天气", "风力向风", "空气质量指数"}, ["2021-11-01 周一", "15°", "8°", "多云~阴", "东风2级", "75 良"], ["2021-11-02 周二", "18°", "10°", "雾~阴", "西南风2级", "70 良"], ["2021-11-03 周三", "18°", "9°", "阴", "东风2级", "73 良"], ["2021-11-04 周四", "18°", "9°", "多云", "东北风2级", "103 轻度"], ["2021-11-05 周五", "18°", "6°", "多云~小雨", "东北风2级", "113 轻度"], ["2021-11-06 周六", "18°", "10°", "小雨", "西风4级", "125 轻度"], ["2021-11-07 周日", "8°", "-3°", "多云~晴", "西风3级", "44 优"], ["2021-11-08 周一", "11°", "8°", "多云~晴", "南风4级", "59 良"], ["2021-11-09 周二", "14°", "2°", "多云~晴", "西南风2级", "55 良"], ["2021-11-10 周三", "15°", "8°", "多云~晴", "西南风2级", "67 良"], ["2021-11-11 周四", "19°", "2°", "晴", "西南风2级", "54 良"], ["2021-11-12 周五", "14°", "2°", "多云", "西南风2级", "69 良"], ["2021-11-13 周六", "17°", "3°", "晴", "西南风2级", "76 良"], ["2021-11-14 周日", "12°", "4°", "多云~阴", "北风1级", "115 轻度"], ["2021-11-15 周一", "18°", "2°", "多云~晴", "西南风1级", "98 良"], ["2021-11-16 周二", "16°", "3°", "多云", "东北风1级", "124 轻度"], ["2021-11-17 周三", "16°", "8°", "多云~阴", "东北风1级", "103 轻度"], ["2021-11-18 周四", "11°", "4°", "阴~小雨", "西风2级", "118 轻度"], ["2021-11-19 周五", "12°", "-3°", "多云~晴", "西北风3级", "66 良"], ["2021-11-20 周六", "8°", "-2°", "晴", "西风2级", "118 轻度"], ["2021-11-21 周日", "13°", "8°", "晴", "西南风1级", "89 良"]], [{"日期": "最高温", "最低温", "天气", "风力向风", "空气质量指数"}]

```

2.3.2 数据处理

通过 `day_data.py` 实现对爬取的每月平均天气数据进行处理，其中包括格式化数据、字符串保存的数据转化为整型数据。

使用的库包括：

(1) Pandas： 处理数据

2.3.3 数据可视化

通过 `day_plot.py` 实现对 `day_data.py` 处理后的每日天气数据进行可视化处理。

使用的库包括：

(1) matplotlib： 数据可视化

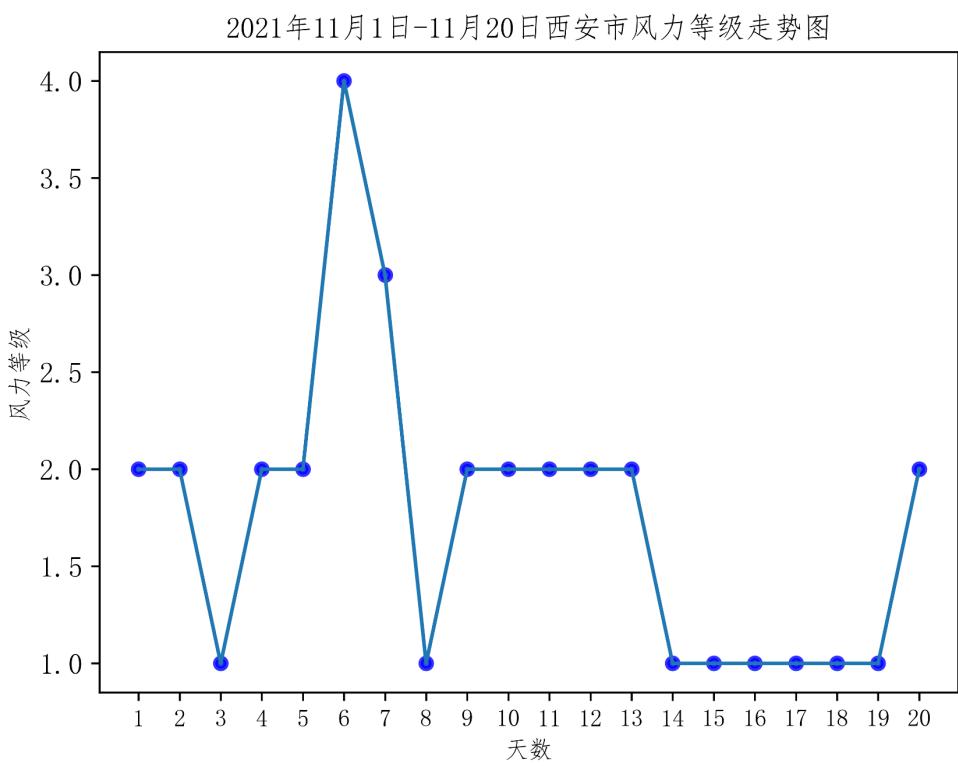
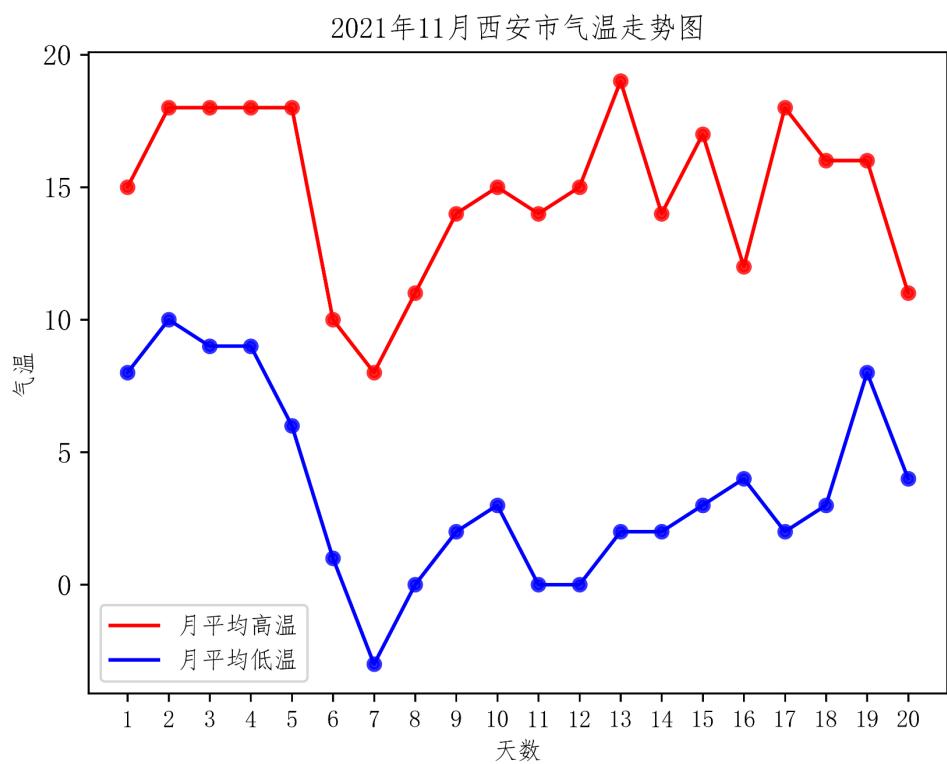
封装函数包括：

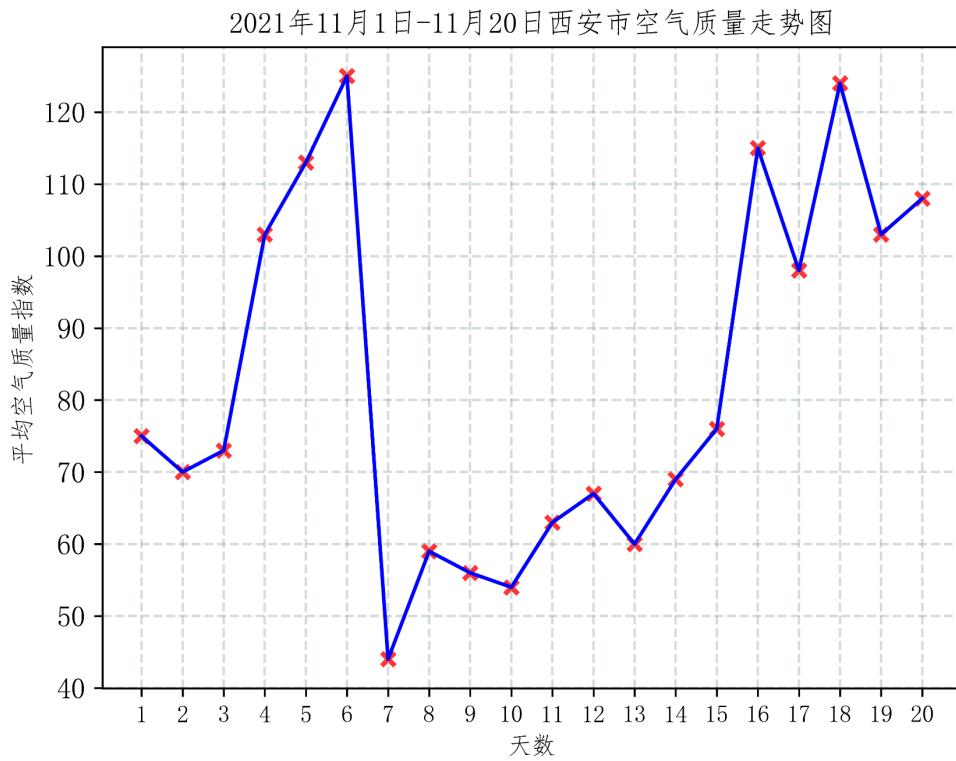
(1) `plot_temperature()`: 气温数据可视化

(2) `plot_wind_power()`: 风力等级数据可视化

(3) `plot_AQI()`: 空气质量指数可视化

结果保存在`./result_figure/daily/plot`文件夹下，包括：





2.3.4 数据拟合

通过 `day_fit.py` 实现对 20 天的天气数据进行拟合，调用了 `day_data.py` 中的数据，采用多项式拟合，最高次数从 2 变化到 13，取标准误差（均方根误差）最小的拟合结果。

使用的库包括：

- (1) `matplotlib`: 数据可视化
- (2) `numpy`: 科学计算
- (3) `sklearn`: 机器学习

封装函数包括：

- (1) `stdError_func(y_test, weather)`: 计算标准误差（均方根误差）
- (2) `R2_func(y_test, weather)`: 计算 R 方
- (3) `load_data(amount, variate)`: 导入天气数据
- (4) `daily_weather_fit(amount, weather, weather_name, i)`: 拟合天气数据。

计算结果如图所示：

```

以下是对最高温的多项式拟合：
degree=2: strError=3.06, R2=0.69, clf.score=0.69
degree=3: strError=2.73, R2=0.28, clf.score=0.28
degree=4: strError=2.73, R2=0.28, clf.score=0.28
degree=5: strError=2.26, R2=0.51, clf.score=0.51
degree=6: strError=2.19, R2=0.53, clf.score=0.53
degree=7: strError=2.03, R2=0.68, clf.score=0.68
degree=8: strError=1.79, R2=0.69, clf.score=0.69
degree=9: strError=1.71, R2=0.72, clf.score=0.72
degree=10: strError=1.71, R2=0.72, clf.score=0.72
degree=11: strError=1.79, R2=0.69, clf.score=0.69
degree=12: strError=4.18, R2=-0.69, clf.score=-0.69
degree=13: strError=5.78, R2=-2.16, clf.score=-2.16
拟合最优结果：9次多项式拟合

```

```

以下是对最低温的多项式拟合：
degree=2: strError=3.04, R2=0.34, clf.score=0.34
degree=3: strError=2.45, R2=0.57, clf.score=0.57
degree=4: strError=2.33, R2=0.61, clf.score=0.61
degree=5: strError=2.06, R2=0.70, clf.score=0.70
degree=6: strError=2.06, R2=0.70, clf.score=0.70
degree=7: strError=1.78, R2=0.77, clf.score=0.77
degree=8: strError=1.68, R2=0.82, clf.score=0.82
degree=9: strError=1.56, R2=0.83, clf.score=0.83
degree=10: strError=1.55, R2=0.83, clf.score=0.83
degree=11: strError=1.59, R2=0.82, clf.score=0.82
degree=12: strError=2.82, R2=0.43, clf.score=0.43
degree=13: strError=3.67, R2=0.04, clf.score=0.04
拟合最优结果：10次多项式拟合

```

```

以下是对风力的多项式拟合：
degree=2: strError=0.75, R2=0.08, clf.score=0.08
degree=3: strError=0.69, R2=0.21, clf.score=0.21
degree=4: strError=0.68, R2=0.23, clf.score=0.23
degree=5: strError=0.59, R2=0.42, clf.score=0.42
degree=6: strError=0.58, R2=0.44, clf.score=0.44
degree=7: strError=0.52, R2=0.55, clf.score=0.55
degree=8: strError=0.52, R2=0.56, clf.score=0.56
degree=9: strError=0.48, R2=0.62, clf.score=0.62
degree=10: strError=0.45, R2=0.67, clf.score=0.67
degree=11: strError=0.49, R2=0.68, clf.score=0.68
degree=12: strError=0.67, R2=0.26, clf.score=0.26
degree=13: strError=0.66, R2=0.27, clf.score=0.27
拟合最优结果：10次多项式拟合

```

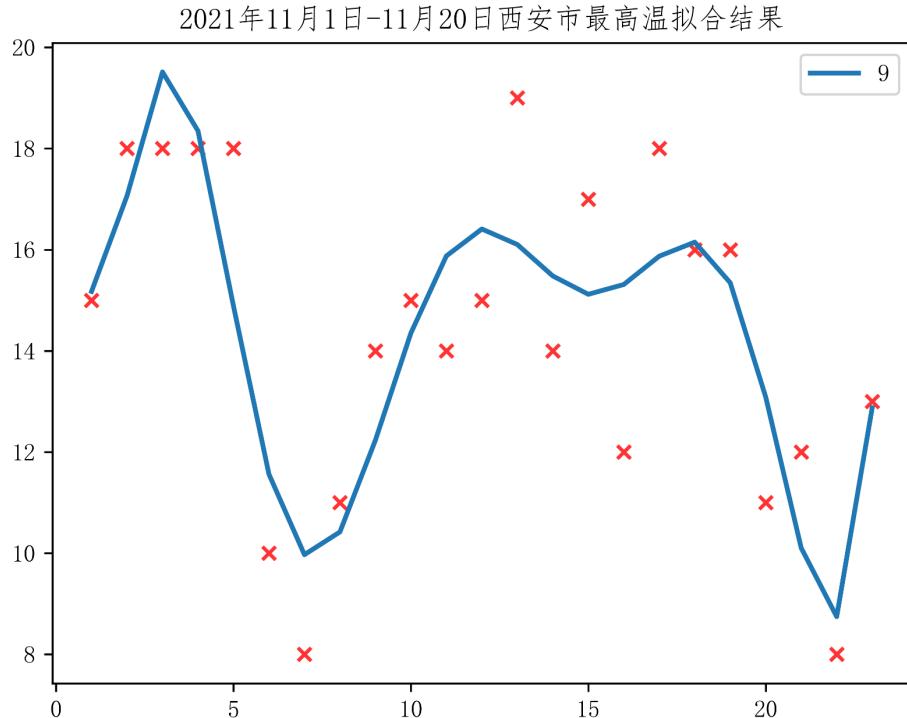
```

以下是对空气质量指数的多项式拟合：
degree=2: strError=22.58, R2=0.17, clf.score=0.17
degree=3: strError=22.27, R2=0.18, clf.score=0.18

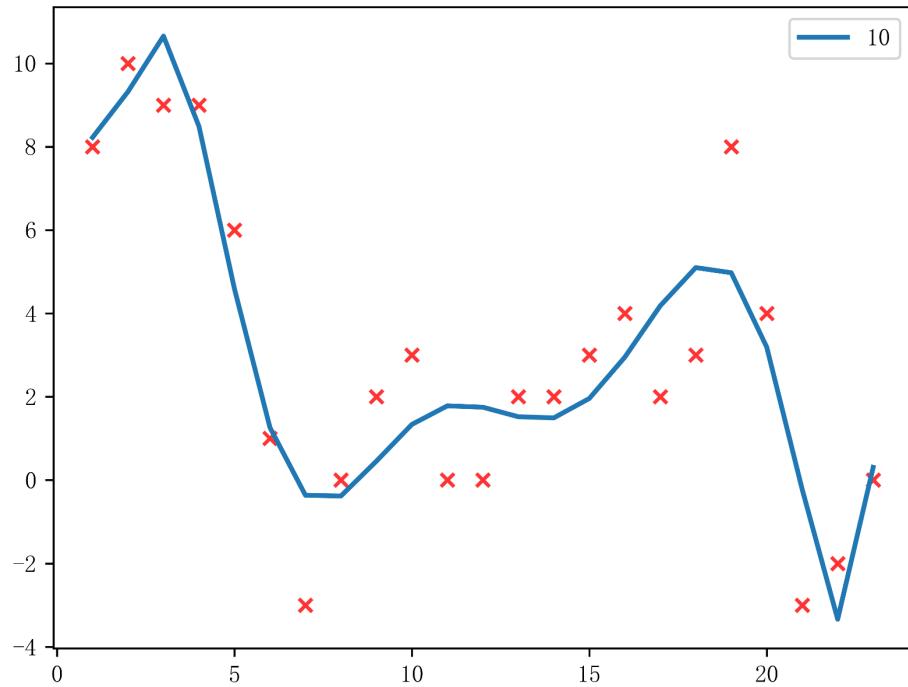
```

行 74, 列 86 空格: 4 UTF-8 CRLF

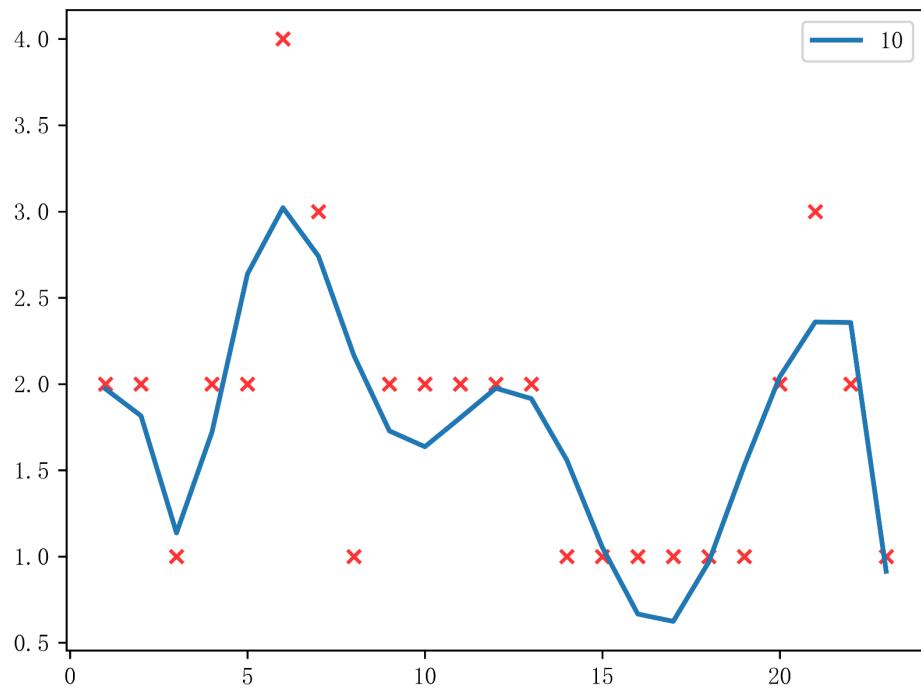
可视化结果保存在./result_figure/daily/fit 文件夹下，包括：

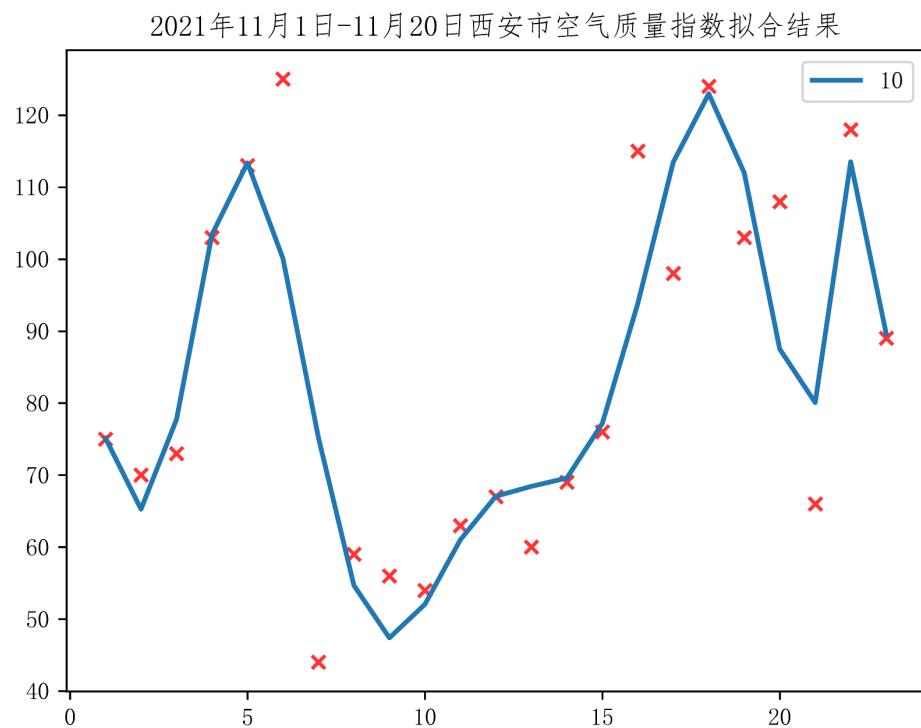


2021年11月1日-11月20日西安市最低温拟合结果



2021年11月1日-11月20日西安市风力拟合结果





2.4 每日空气质量

2.4.1 数据爬取

通过 `AQI_daily.py` 实现对 2016 年-2020 年西安市 5 年间每天的空气质量指数、空气质量排名、6 种污染物（PM2.5、PM10、SO₂、NO₂、CO、O₃）浓度数据的爬取，共 1786 个数据。结果保存在 `./result_data/AQI_daily.csv` 文件中。

使用的库包括：

- (1) `requests`: 爬取网页
- (2) `BeautifulSoup`: 解析网页
- (3) `Pandas`: 处理数据

封装函数包括：

- (1) `format_date_monthly(year,month)`: 格式化日期 (年月)，用于实现网页翻页
- (2) `format_date_daily(year,month,day)`: 格式化日期 (年月日)：用于结果保存
- (3) `all_date_monthly(years)`: 输出要分析的日期 (年月)，用于实现网页翻

页

(4) `all_html(years)`: 保存所有日期所在网页的超链接

(5) `get_html(url)`: 爬取网页

(6) `get_data(years)`: 解析所有网页

运行截图如下:

```
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d;; cd "d:\OPTIMUS\APP_usr\Code\Python与机器学习\mine"; & 'C:\Users\GYC\AppData\Local\Programs\Python\Python310\python.exe' 'c:\Users\GYC\vscode\extensions\ms-python.python-2021.11.1422169775\pythonFiles\lib\python\debugpy\launcher' '2158' '--' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\AQI_daily.py'
analyzing 1/60 month
analyzing 1/31 day:20160101
analyzing 2/31 day:20160102
analyzing 3/31 day:20160103
analyzing 4/31 day:20160104
analyzing 5/31 day:20160105
analyzing 6/31 day:20160106
analyzing 7/31 day:20160107
analyzing 8/31 day:20160108
analyzing 9/31 day:20160109
analyzing 10/31 day:20160110
analyzing 11/31 day:20160111
analyzing 12/31 day:20160112
analyzing 13/31 day:20160113
analyzing 14/31 day:20160114
analyzing 15/31 day:20160115
analyzing 16/31 day:20160116
analyzing 17/31 day:20160117
analyzing 18/31 day:20160118
analyzing 19/31 day:20160119
analyzing 20/31 day:20160120
```

2.4.2 数据处理

通过 `AQI_data.py` 实现对爬取的每日空气质量数据进行处理，其中包括将 `dataframe` 类型的数据转化为 `list` 类型。

使用的库包括:

(1) Pandas: 处理数据

2.4.3 数据可视化

通过 `AQI_plot.py` 实现对 `AQI_data.py` 处理后的每日天气数据进行可视化处理。

使用的库包括:

(1) matplotlib: 数据可视化

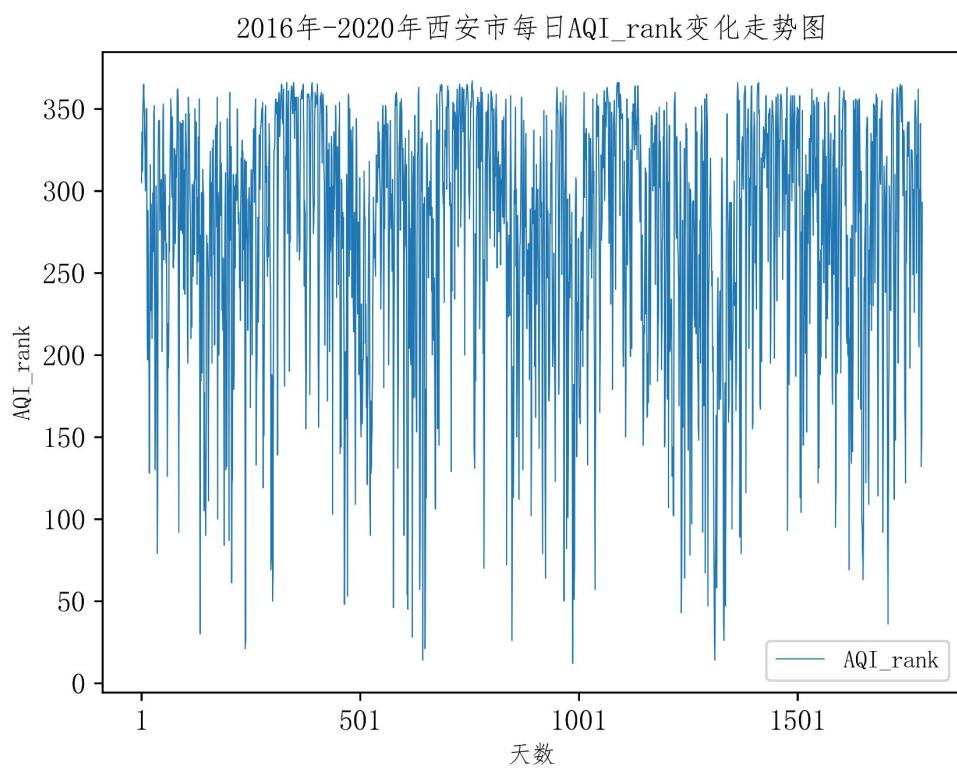
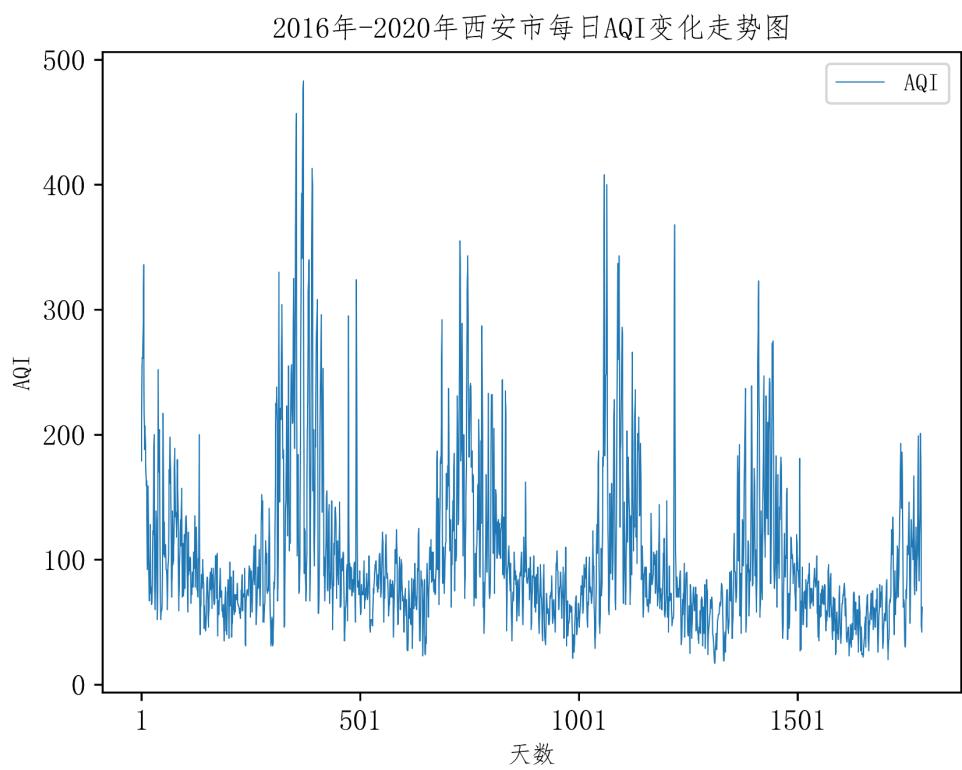
封装函数包括:

(1) `plot_temperature()`: 气温数据可视化

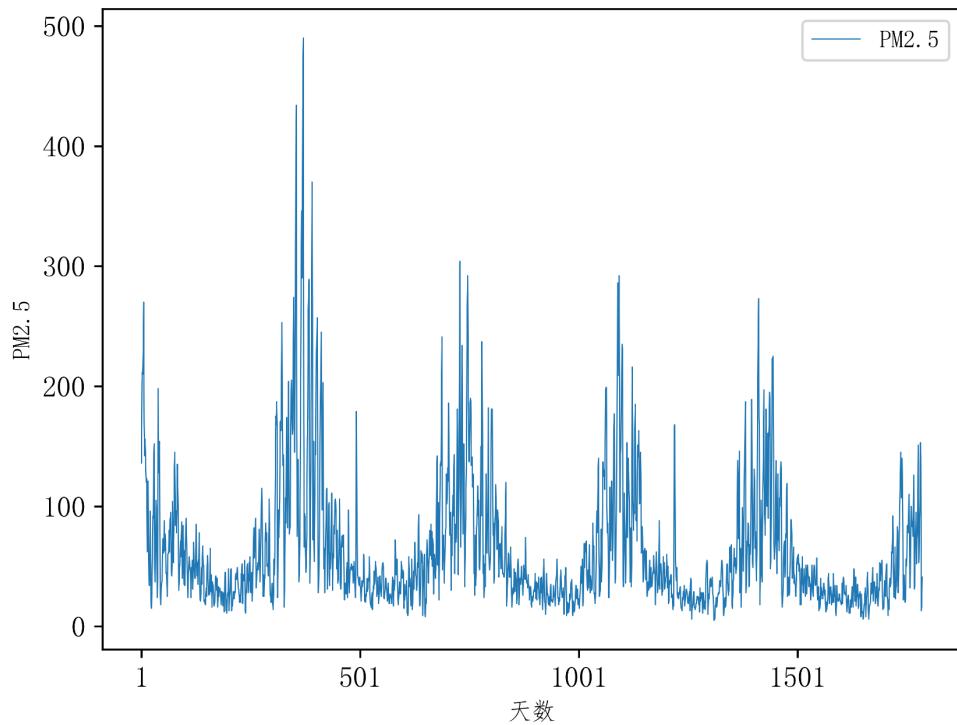
(2) `plot_wind_power()`: 风力等级数据可视化

(3) `plot_AQI()`: 空气质量指数可视化

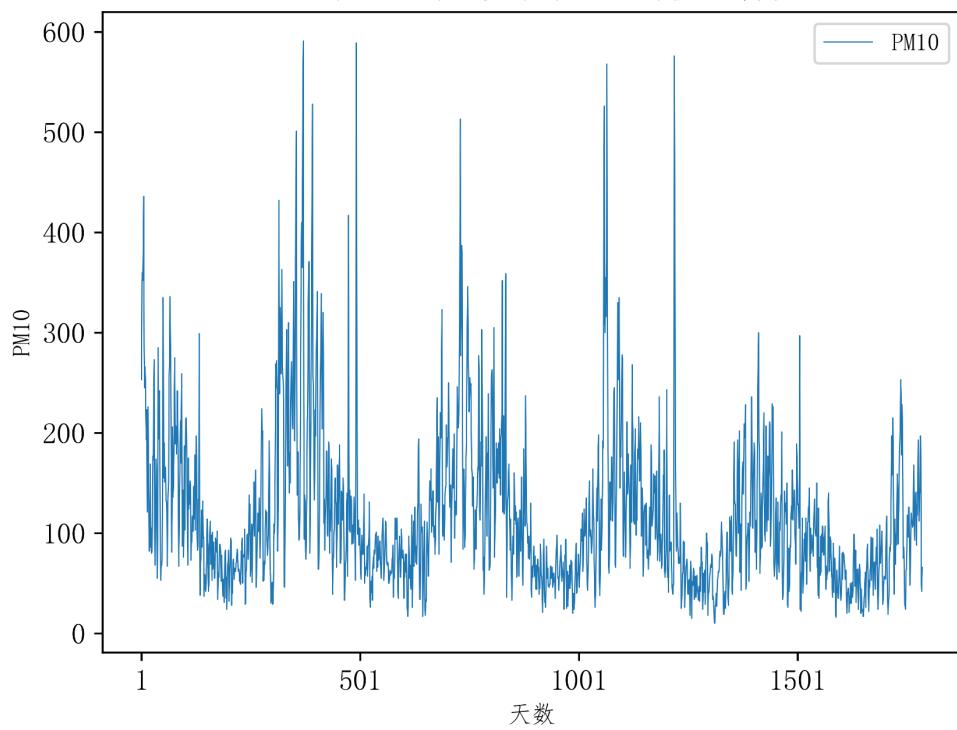
结果保存在./result_figure/AQI/plot 文件夹下，包括:



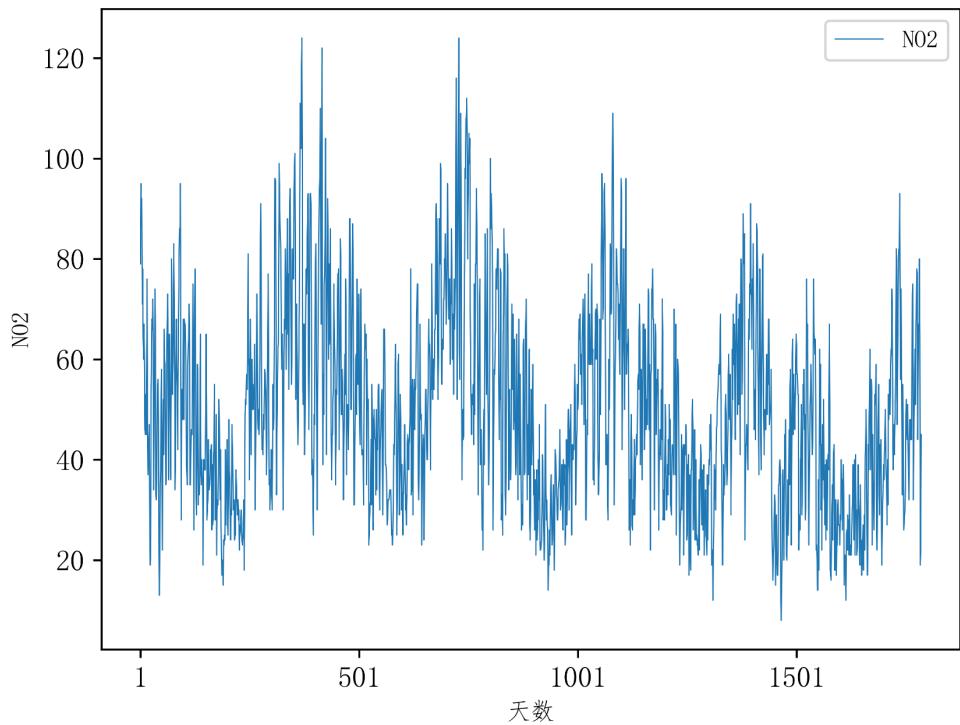
2016年-2020年西安市每日PM2.5变化走势图



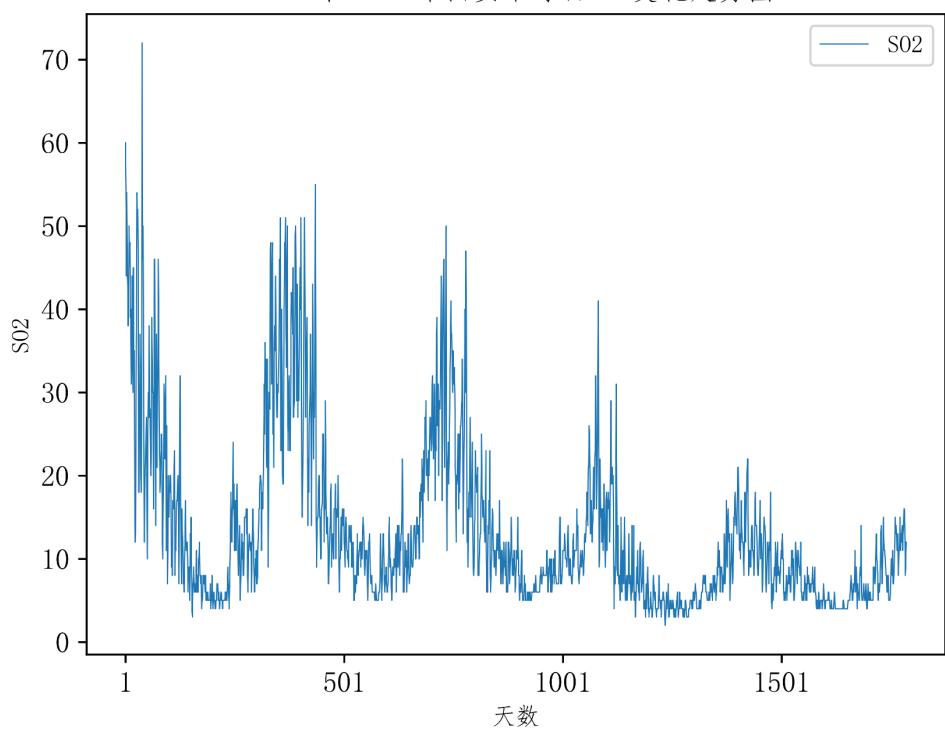
2016年-2020年西安市每日PM10变化走势图



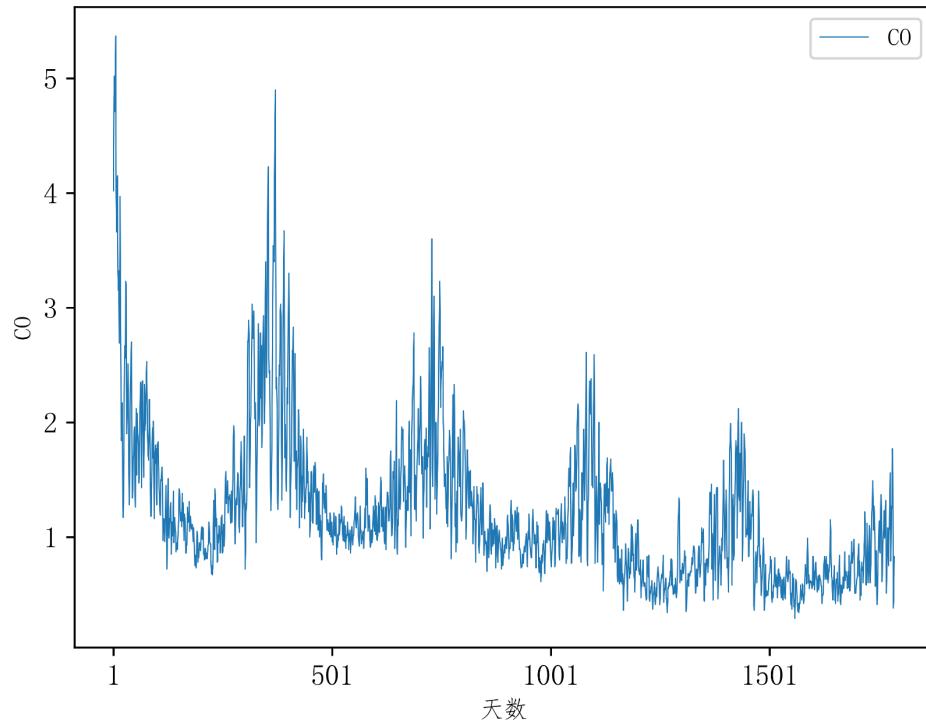
2016年-2020年西安市每日NO₂变化走势图



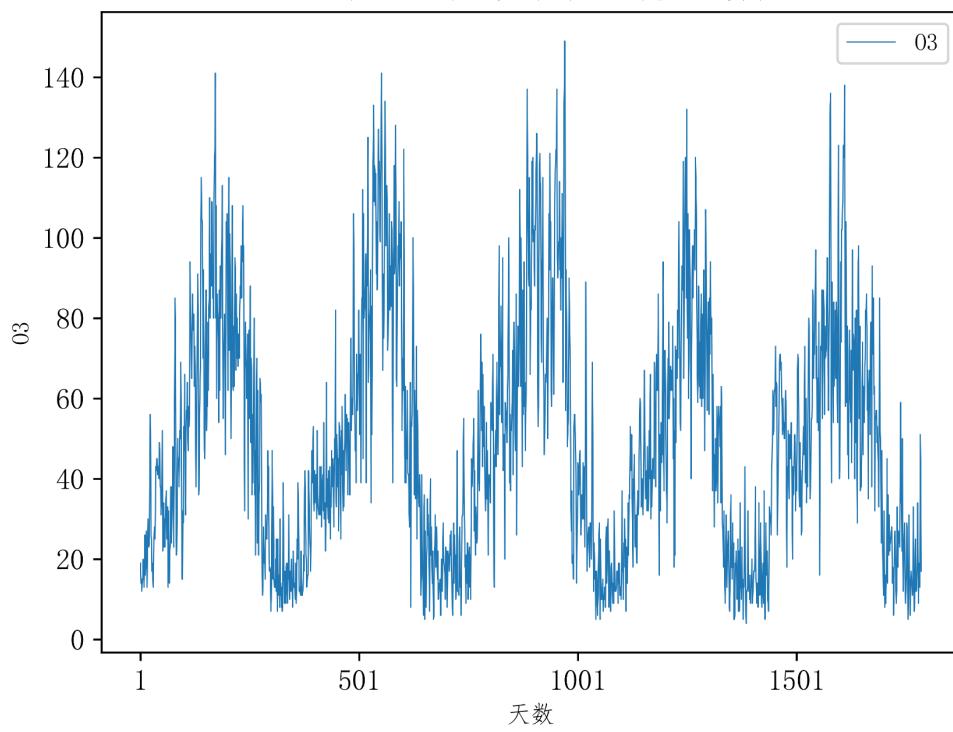
2016年-2020年西安市每日SO₂变化走势图



2016年-2020年西安市每日CO变化走势图



2016年-2020年西安市每日O₃变化走势图



2.4.4 数据回归

通过 **AQI_regression.py** 实现对 1786 天的空气质量数据进行回归分析，调用了 **AQI_data.py** 中的数据，空气质量指数数据作为因变量，6 种污染物浓度数据作为自变量，回归模型采用随机森林回归，决策树个数取 100。

使用的库包括：

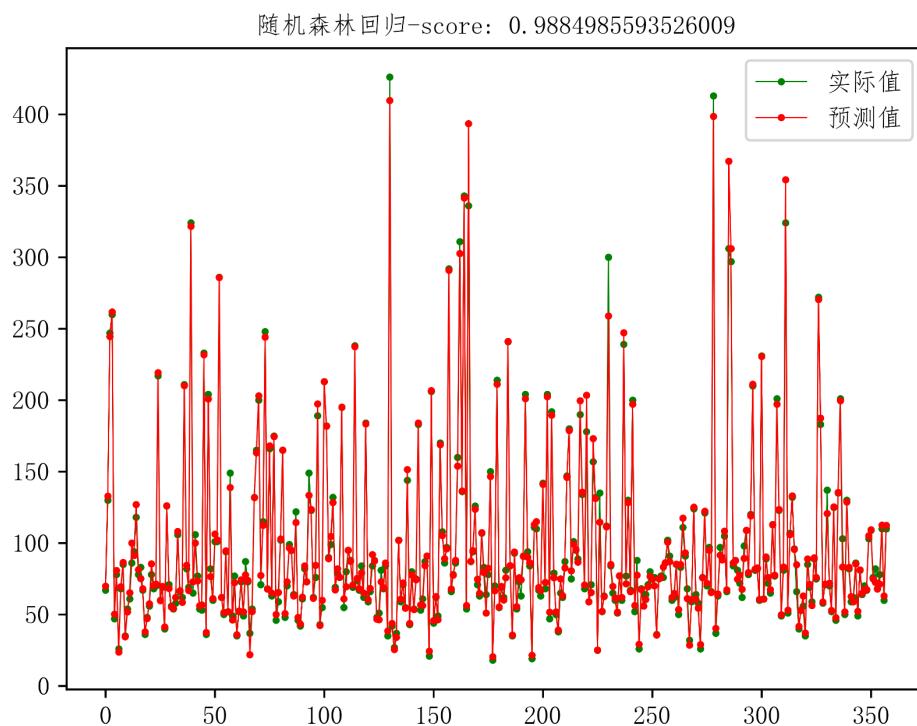
- (1) **matplotlib**: 数据可视化
- (2) **numpy**: 科学计算
- (3) **Pandas**: 处理数据
- (4) **sklearn**: 机器学习

计算结果如图所示：

```
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d;; cd 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine'; & 'C:\Users\GYC\AppData\Local\Programs\Python\Python310\python.exe' 'c:/Users/GYC/.vscode/extensions/ms-python.python-2021.11.1422169775/pythonFiles/lib/python\debugpy\launcher' '6051' '--' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\AQI_regression.py'
训练集大小: 1428
测试集大小: 358
R方: 0.9892    平均绝对误差: 3.4349    中值绝对误差: 1.8792
```

R 方值在 0.9892，非常接近于 1，说明训练得到的模型性能较好。

可视化结果保存在./result_figure/AQI/regression 文件夹下，包括：



2.4.5 模型应用

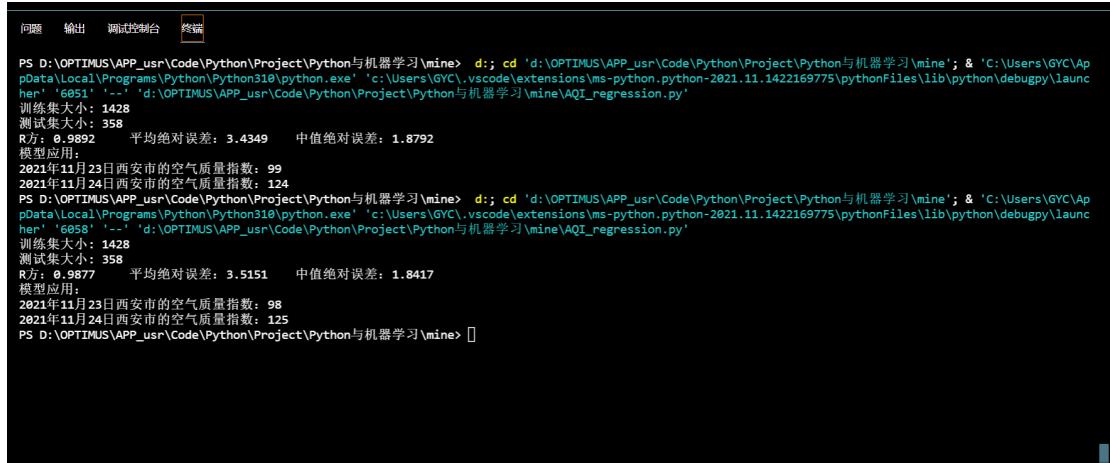
将 2021 年 11 月 23 日和 24 日西安市的空气质量数据应用到机器学习训练得到的模型之中，即当 6 种空气污染物 PM2.5、PM10、SO₂、NO₂、CO、O₃ 分别为：

表 2 模型应用采用的数据

成分	2021 年 11 月 23 日	2021 年 11 月 24 日
PM2.5	61	57
PM10	146	193
SO ₂	12	9
NO ₂	52	47
CO	0	0
O ₃	12	5
AQI	98	121



预测结果：



```
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d;; cd 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine'; & 'C:\Users\GYC\AppData\Local\Programs\Python\Python310\python.exe' 'c:\Users\GYC\.vscode\extensions\ms-python.python-2021.11.1422169775\pythonFiles\lib\python\debugpy\launcher' '6051' '-' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\AQI_regression.py'
训练集大小: 1428
测试集大小: 358
R方: 0.9892 平均绝对误差: 3.4349 中值绝对误差: 1.8792
模型应用:
2021年11月23日西安市的空气质量指数: 99
2021年11月24日西安市的空气质量指数: 124
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> d;; cd 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine'; & 'C:\Users\GYC\AppData\Local\Programs\Python\Python310\python.exe' 'c:\Users\GYC\.vscode\extensions\ms-python.python-2021.11.1422169775\pythonFiles\lib\python\debugpy\launcher' '6058' '-' 'd:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine\AQI_regression.py'
训练集大小: 1428
测试集大小: 358
R方: 0.9877 平均绝对误差: 3.5151 中值绝对误差: 1.8417
模型应用:
2021年11月23日西安市的空气质量指数: 98
2021年11月24日西安市的空气质量指数: 125
PS D:\OPTIMUS\APP_usr\Code\Python\Project\Python与机器学习\mine> 
```

由上图可知，由于训练集和测试集的划分是随机的，每次运行会出现不同的划分，每次运行机器学习训练得到的模型也存在差别。由图中的最后一次训练结果可知 23 日西安市空气质量指数预测值为 98，实际值为 98；24 日西安市空气质量指数预测值为 125，实际值为 121，预测精度较为理想。

3 结果说明

表 3 结果文件汇总

结果	类型	所在文件
2021 年 11 月 1 日-21 日西安市每日天气数据	数据	data_daily.csv
2011 年-2020 年西安市每月平均天气数据	数据	data_monthly.csv
2016 年-2020 年西安市每日空气质量数据	数据	AQI_daily.csv
2011-2020 年西安市 11 月气温走势图	图片	与结果同名
2011-2020 年西安市月平均气温走势图	图片	与结果同名
2018-2020 年西安市月平均空气质量指数走势图	图片	与结果同名
9 种回归模型对比_月平均低温	图片	与结果同名
9 种回归模型对比_月平均高温	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市风力等级走势图	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市空气质量走势图	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市气温走势图	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市风力拟合结果	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市空气质量指数拟合结果	图片	与结果同名

2021 年 11 月 1 日-11 月 20 日西安市最低温拟合结果	图片	与结果同名
2021 年 11 月 1 日-11 月 20 日西安市最高温拟合结果	图片	与结果同名
2016 年-2020 年西安市每日 AQI_rank 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 AQI 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 CO 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 NO2 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 O3 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 PM2.5 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 PM10 变化走势图	图片	与结果同名
2016 年-2020 年西安市每日 SO2 变化走势图	图片	与结果同名
随机森林回归结果	图片	与结果同名

4 总结

经过本学期《Python 与机器学习》的学习，我对 Python 基础语法、Python 科学计算与绘图、爬虫基础知识、机器学习基础知识形成了基本的认识和理解，其中从之前使用的集成开发环境 Pycharm 在本次课程过渡到 Visual Studio Code 代码编辑器，让我可以在同一个代码编辑器下完成多种编程语言的编辑和运行，比如在学习本课程期间，由于毕设需要我在 Visual Studio Code 上学习了 Fortran 语法，在完成大作业时我又用这一代码编辑器学习了 HTML 和 CSS 基础语法……这些对我的未来研究生阶段来说当然仍只是一个起点，仍有许多不足之处。比如(1)本次大作业中的爬虫过程可能存在访问不稳定导致网页数据爬取失败；(2)机器学习过程中对模型原理的理论学习还比较缺乏，只是简单的应用而无法做到训练模型的调参与优化；(3)由于神经网络运行环境的搭建还未完成，因此没有将神经网络应用到机器学习之中，或许模型表现会更好；(4)大作业中没有使用到 Python 类等等。

最后辛苦钟老师和王老师本学期对于 Python 和机器学习方面知识的教授和指导，在这里向两位老师表示衷心的感谢！

参考文献

- [1] 刘锋,李春燕,谭祥勇,王鹏飞.基于机器学习在空气质量指数中的应用[J].重庆工商大学学报(自然科学版),2017,34(03):82-87.