例 5.11 Python 程序代码

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

plt.rcParams['font.sans-serif'] = 'SimHei'
%matplotlib inline

df = pd.read_csv('./RFM 分析 1.csv')
df.info()
df['月度数据'] = pd.to_datetime(df['月度数据'])
# 查看是否修改成功
df.info()
df.describe()
df[df['月度数据'] < 0]
df = df[df['月度数据'] > 0]
# 重新看月度数据的最小值
df.describe()
df.index = df['月度数据']
df.head()
df_Month = df.resample('M').sum()
plt.figure(figsize=(18, 7), dpi=128)
df_Month['月度数据'].plot()
from statsmodels.tsa.arima_model import ARMA
from datetime import datetime
from itertools import product

# 设置 p 阶，q 阶范围
# product p,q 的所有组合
# 设置最好的 aic 为无穷大
# 对范围内的 p,q 阶进行模型训练，得到最优模型
ps = range(0, 6)
qs = range(0, 6)
parameters = product(ps, qs)
parameters_list = list(parameters)

best_aic = float('inf')
results = []
for param in parameters_list:
    try:
        model = ARMA(df_Month['月度数据'], order=(param[0], param[1])).fit()
    except ValueError:
        print("参数错误：", param)
        continue
```

```
        aic = model.aic
        if aic < best_aic:
            best_model = model
            best_aic = model.aic
            best_param = param
        results.append([param, model.aic])
    results_table = pd.DataFrame(results)
    results_table.columns = ['parameters', 'aic']
    print("最优模型", best_model.summary())
    # 先增加后几个月的数据，进行合并
    date_list = [datetime(2014, 8, 31), datetime(2014, 9, 30), datetime(2014, 10, 31),
                 datetime(2014, 11, 30), datetime(2014, 12, 31),
                 datetime(2015, 1, 31)]
    df_Month = df_Month[['月度数据']]
    future = pd.DataFrame(index=date_list, columns= df_Month.columns)
    df_Month = pd.concat([df_Month, future])
    df_Month
    df_Month['forecast'] = best_model.predict(start=0, end=58)
    plt.figure(figsize=(20, 7))
    df_Month['月度数据'].plot(label='实际月度数据')
    df_Month['forecast'].plot(color='r', ls='--', label='预测月度数据')
    plt.legend()
    plt.show()
```