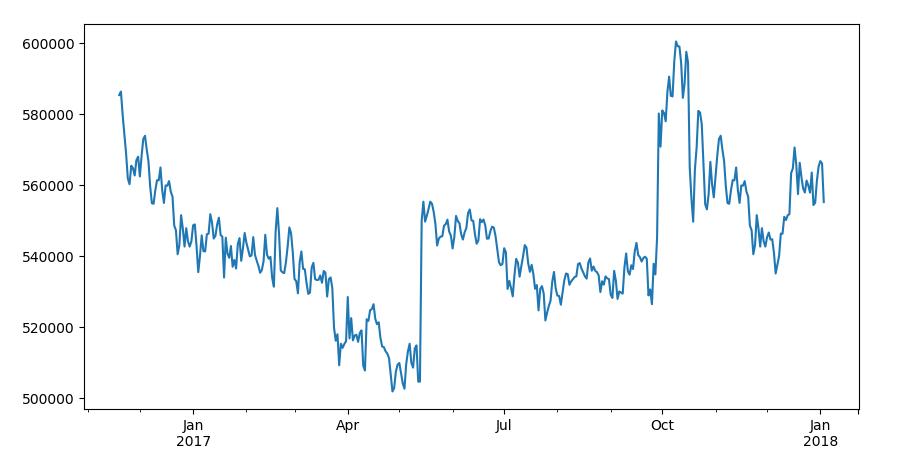
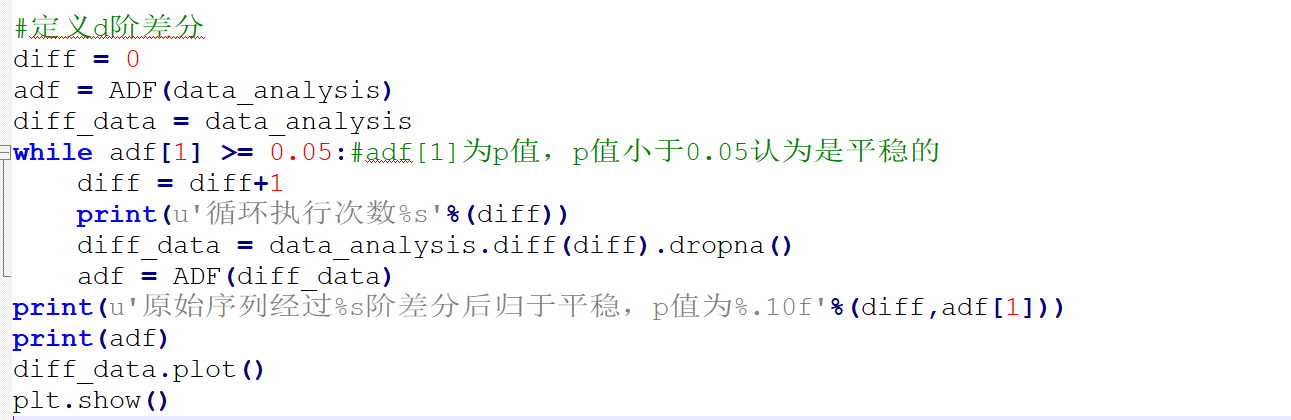
针对集群的实际剩余cpu（总量-实际使用量）分别使用算法Arima和Lstm完成对未来数据的预测。原始图形展现如下：



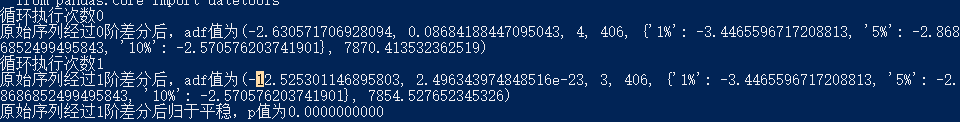
# Arima处理步骤：

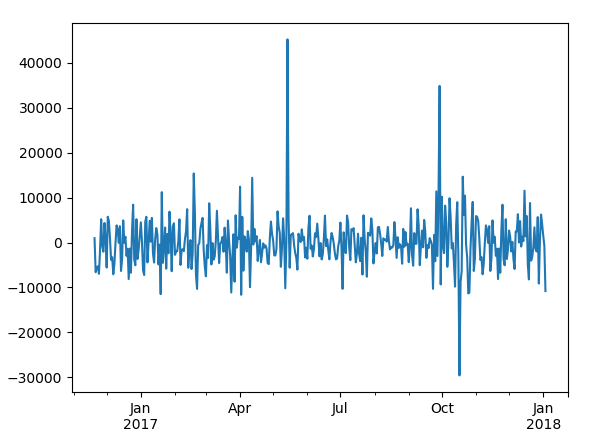
## 数据平稳性分析

对数据绘图，观测是否为平稳时间序列；对于非平稳时间序列要先进行d阶差分运算，化为平稳时间序列；判断标准（ADF检验可以得到单位根检验统计量对应的p值，若此值显著大于0.05，则该序列非平稳，否则认为是平稳数据），参考算法：



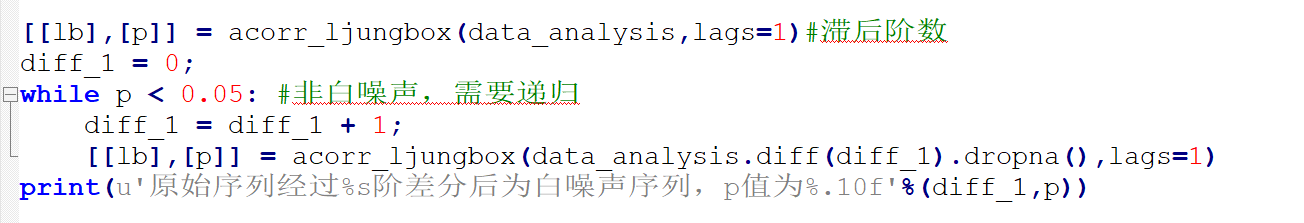
输出信息如下：





## 白噪声处理

为了验证平稳序列中有用信息是否已被提取完，需要对序列进行白噪声检验，如果序列是白噪声的说明序列中的有用信息已经被提取完毕了，剩下的全是随机扰动，无法进行预测和使用。判断标准（检验时间序列是否为白噪声序列，一般如果统计量的P值小于0.05时，则可以拒绝原假设，认为该序列为非白噪声序列，否则为白噪声序列），需要循环处理平稳性和白噪声处理，保障序列的平稳性和白噪声序列。（该步骤可以在模型识别之后处理，来完成模型的校验）

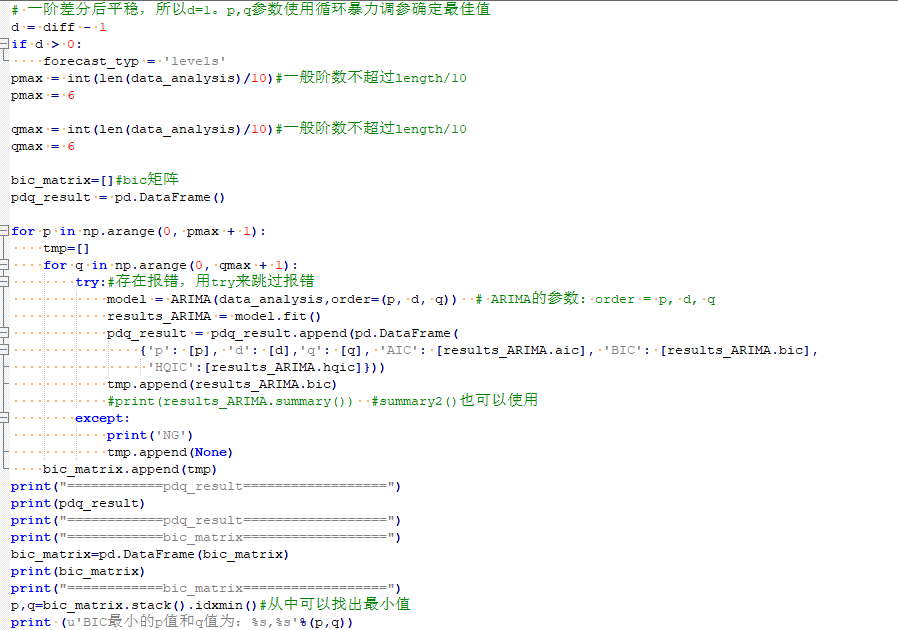


## 模型识别

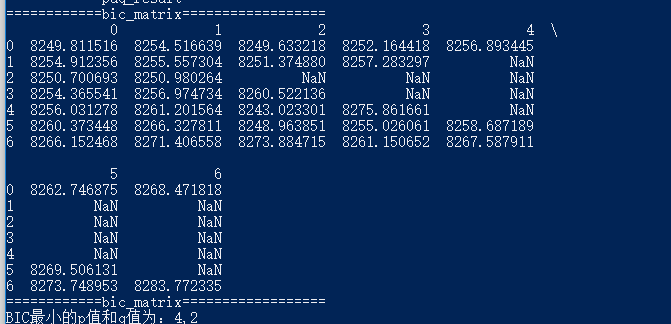
采用极大似然比方法进行模型的参数估计，估计各个参数值，然后针对各个不同模型，采用BIC信息准则（贝叶斯信息准则）对模型进行定阶，确定p,q参数，从而选择最优模型。

确定最佳p、d、q值,d来源于之前的阶数diff字段

p和q的值可以通过对自相关系数图和偏自相关系数图人为识别来确定，确定方法是根据自相关系数图和偏自相关系数图的拖尾和截尾的性质来确定的，同样我们也可以通过BIC矩阵中的最小BIC信息量的位置来决定.



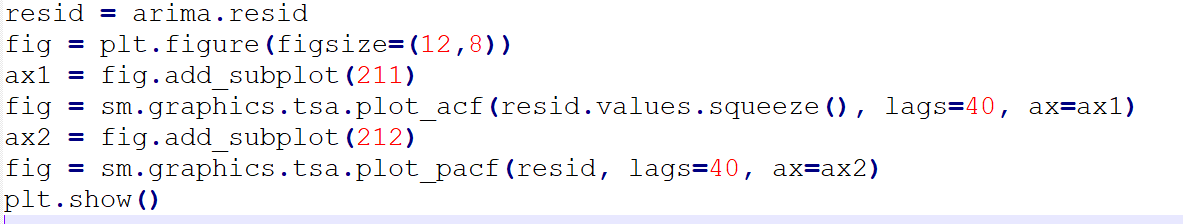
输出信息如下：

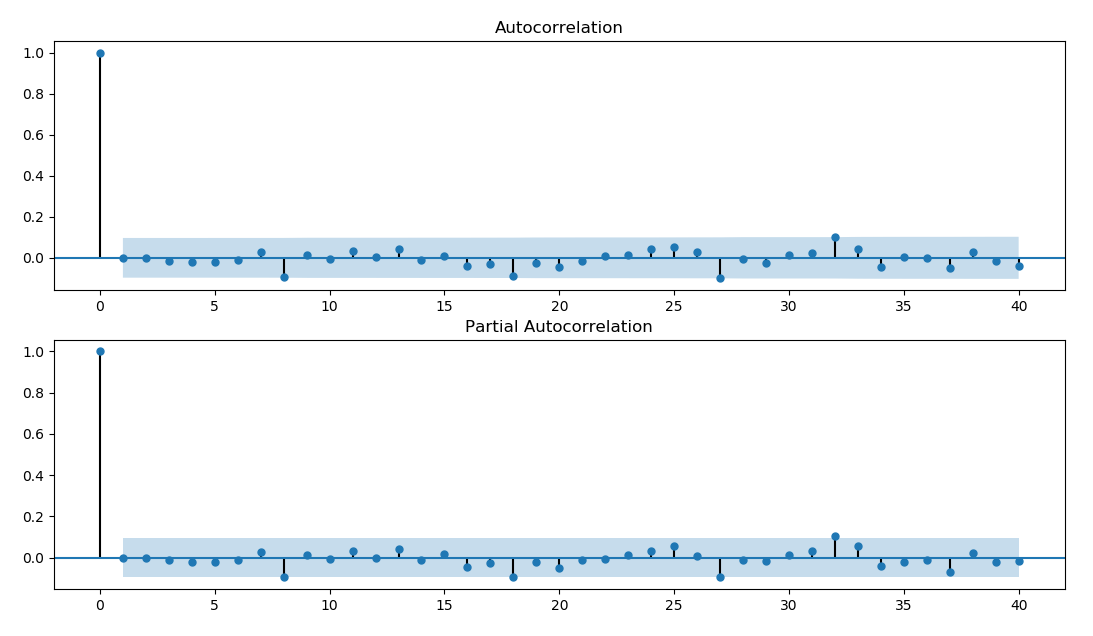


## 模型检测

在指数平滑模型下，观察ARIMA模型的残差是否是平均值为0且方差为常数的正态分布（服从零均值、方差不变的正态分布），同时也要观察连续残差是否（自）相关。

### 检验残差序列





### 做D-W检验 德宾-沃森（Durbin-Watson）检验

德宾-沃森检验,简称D-W检验，是目前检验自相关性最常用的方法，但它只使用于检验一阶自相关性。

因为自相关系数ρ的值介于-1和1之间，所以 0≤DW≤４。并且DW＝O＝＞ρ＝１　　 即存在正自相关性

DW＝４＜＝＞ρ＝－１　即存在负自相关性

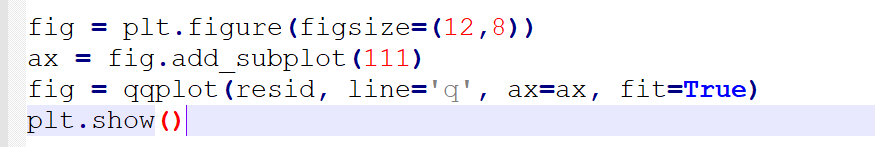
DW＝２＜＝＞ρ＝０　　即不存在（一阶）自相关性

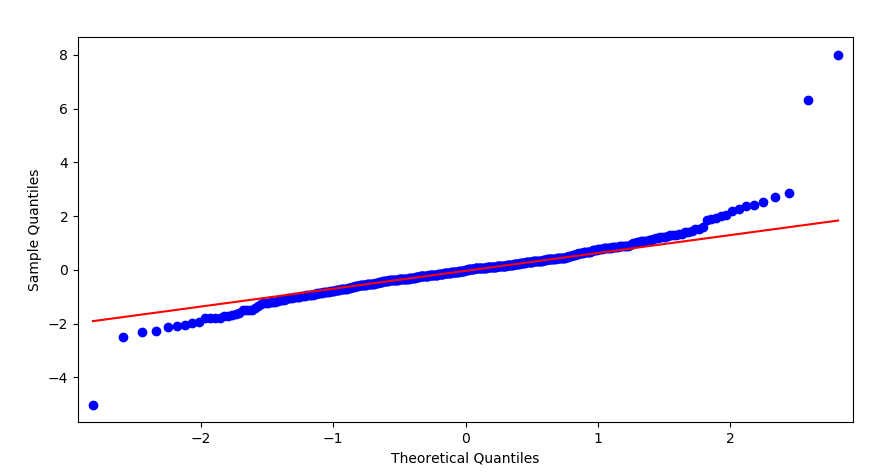
因此，当DW值显著的接近于O或４时，则存在自相关性，而接近于２时，则不存在（一阶）自相关性。这样只要知道ＤＷ统计量的概率分布，在给定的显著水平下，根据临界值的位置就可以对原假设Ｈ０进行检验。



### 观察是否符合正态分布

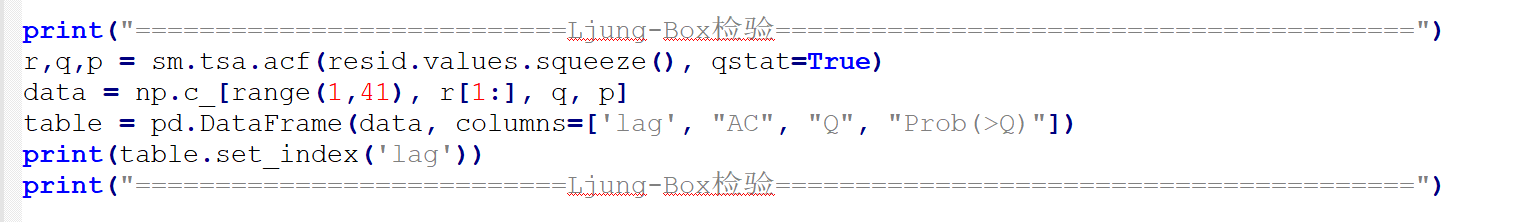
这里使用QQ图，它用于直观验证一组数据是否来自某个分布，或者验证某两组数据是否来自同一（族）分布。

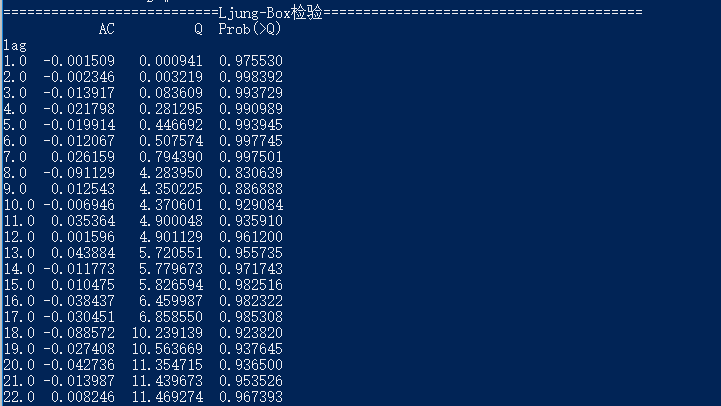




### Ljung-Box检验

Ljung-Box test是对randomness的检验,或者说是对时间序列是否存在滞后相关的一种统计检验。对于滞后相关的检验，我们常常采用的方法还包括计算ACF和PCAF并观察其图像，但是无论是ACF还是PACF都仅仅考虑是否存在某一特定滞后阶数的相关。LB检验则是基于一系列滞后阶数，判断序列总体的相关性或者说随机性是否存在。时间序列中一个最基本的模型就是高斯白噪声序列。而对于ARIMA模型，其残差被假定为高斯白噪声序列，所以当我们用ARIMA模型去拟合数据时，拟合后我们要对残差的估计序列进行LB检验，判断其是否是高斯白噪声，如果不是，那么就说明ARIMA模型也许并不是一个适合样本的模型。检验的结果就是看最后一列前十二行的检验概率（一般观察滞后1~12阶），如果检验概率小于给定的显著性水平，比如0.05、0.10等就拒绝原假设，其原假设是相关系数为零。就结果来看，如果取显著性水平为0.05，那么相关系数与零没有显著差异，即为白噪声序列。

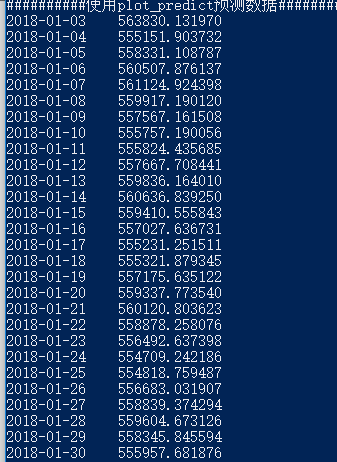




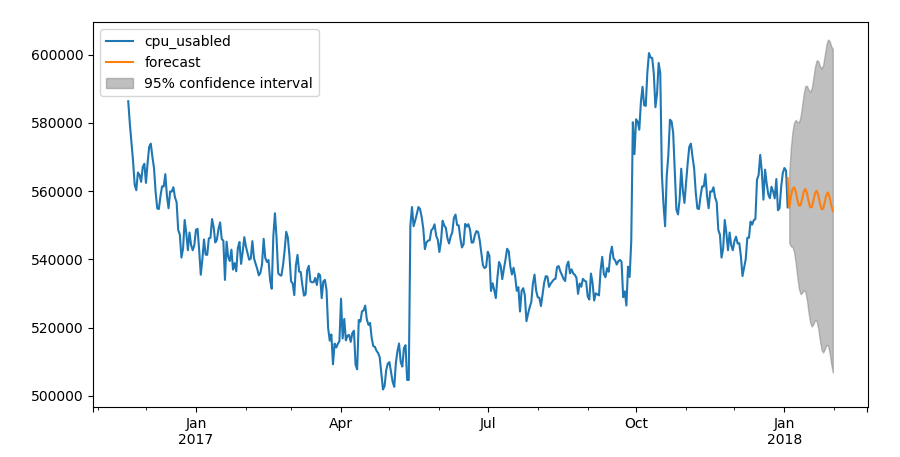
## 数据预测

预测范围2018-01-03到2018-01-31，其中2018-01-03到2018-01-23中的数据为拟合数据

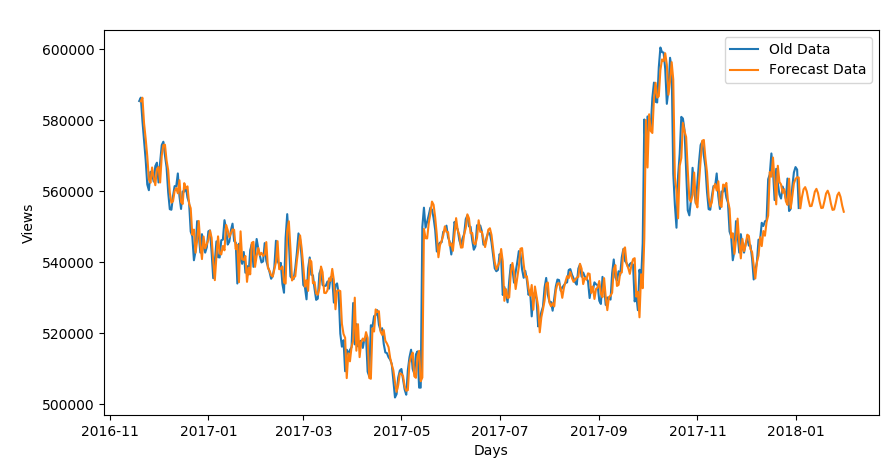
预测数据：



预测曲线：

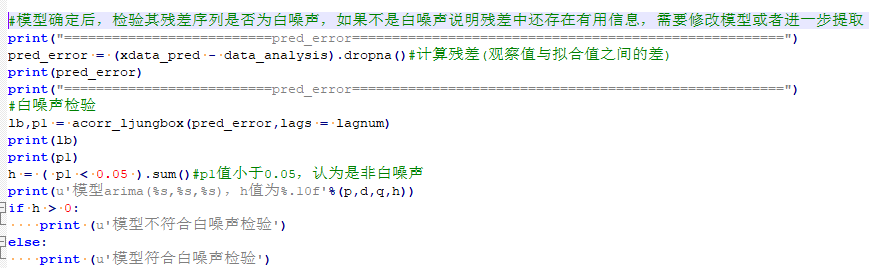


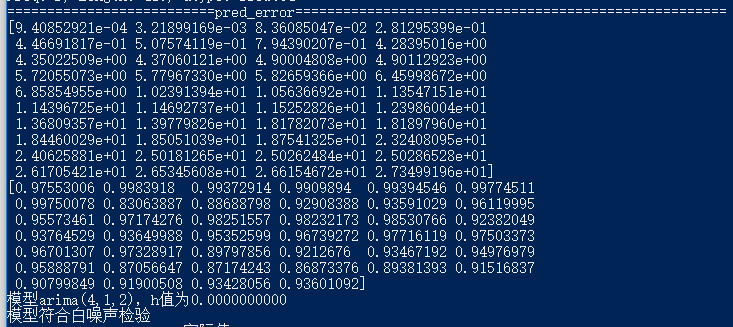
拟合曲线：



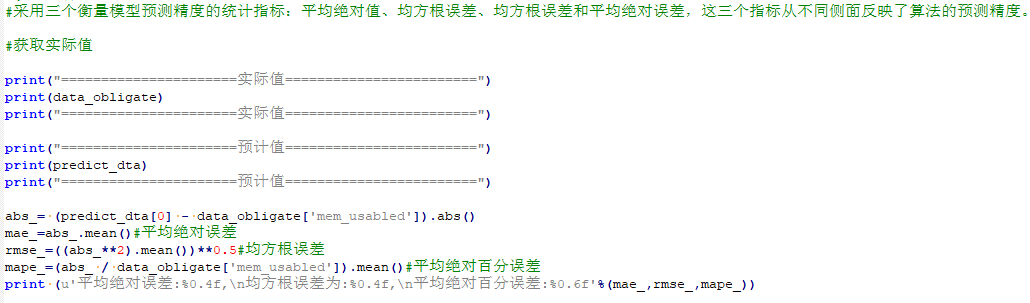
Arima数据拟合度图形

## 数据校对





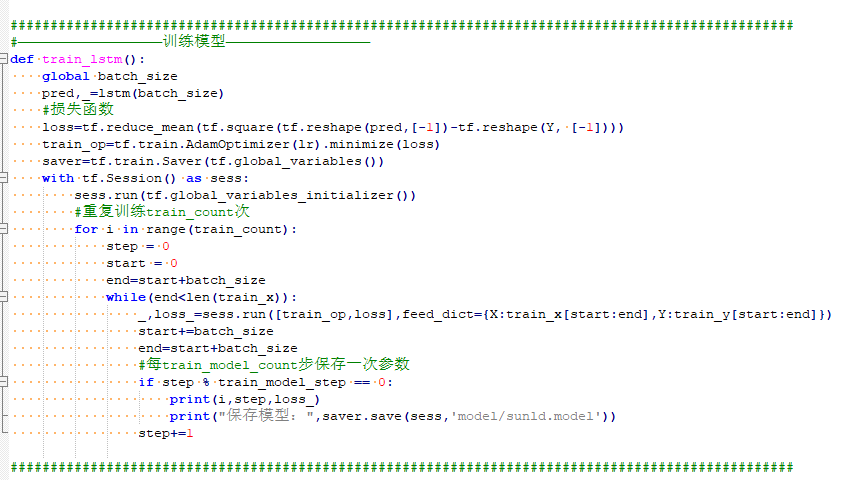
## 模型评价





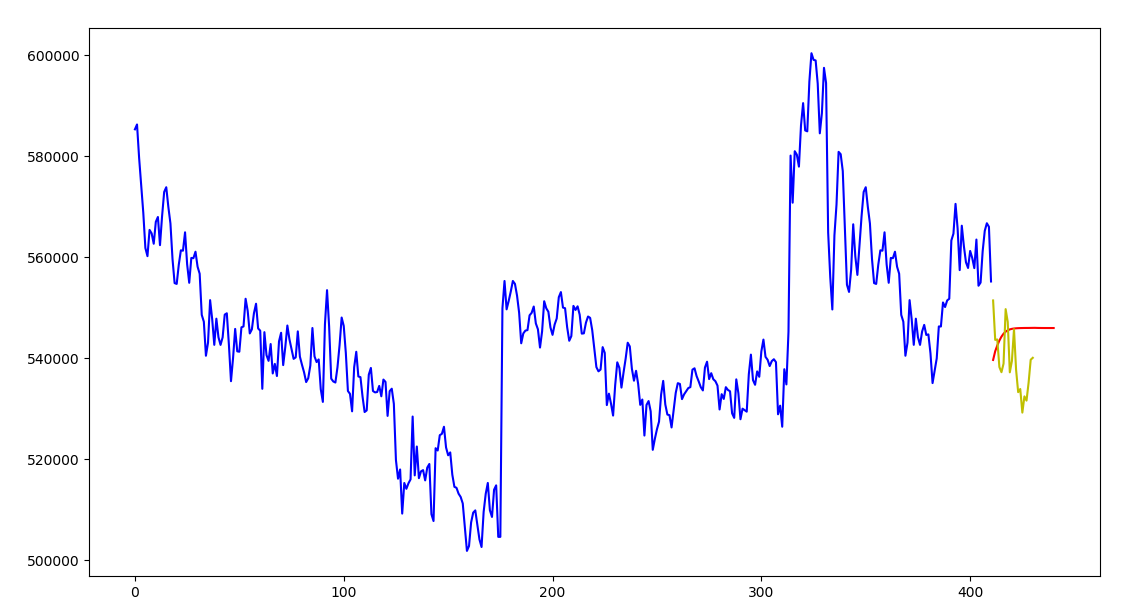
# Lstm处理步骤

## 生成训练模型



## 完成数据预测





Lstm数据拟合度

## 模型校对

预测数据



偏差：



# 比对

|  |  |  |
| --- | --- | --- |
| 算法 | Arima | Lstm |
| 拟合度 | 查看“Arima数据拟合度  ”图形，可以发现数据拟合度比较高 | 查看“Lstm数据拟合度”图形（红色和黄色重合部分），拟合度较低 |
| 工作量 | 3人月 | 6人月（需要系统化调研Lstm） |
| 技术难度 | 低 | 高 |
| 数据模型 | 匹配度低 | 匹配度高 |
| 多因子输入 | 支持率低 | 支持率高 |