第四章 代理人及航空公司销售预测

4.1 整体思路

1）对数据进行预处理

2）预测各销售商的所有待预测时段的销售额及销售量

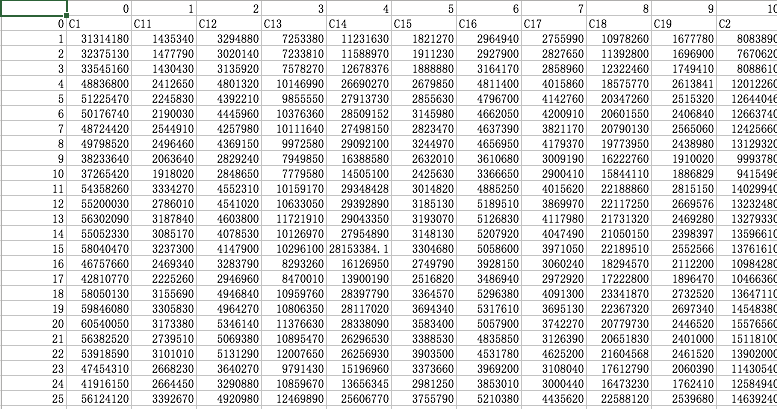
3）预测各销售商的所有待预测时段的销售对象及销售比例

4）以预测的销售额为基础将其分配到各个销售对象，形成最终预测结果

5）导出数据

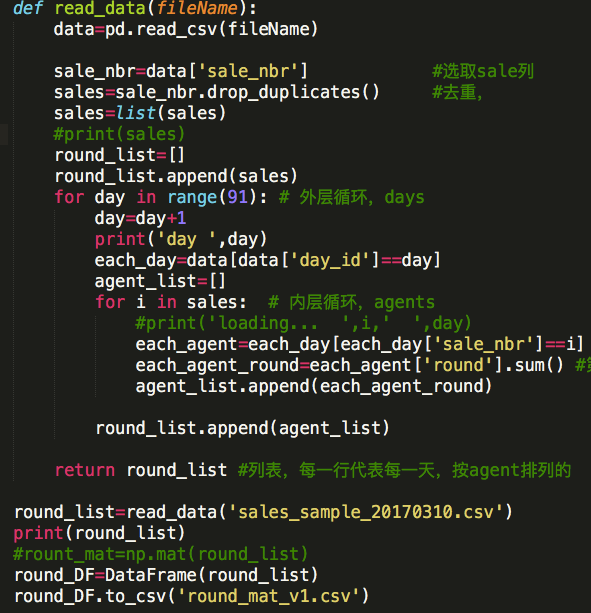
4.2 数据处理

1）给定的初始数据格式混乱不易处理，首先将其整理为矩阵形式：



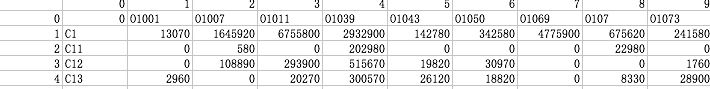
其中，第一行记录所有出现过的销售商，共7397条；第一列记录出现的天数，共91条。

代码如下：



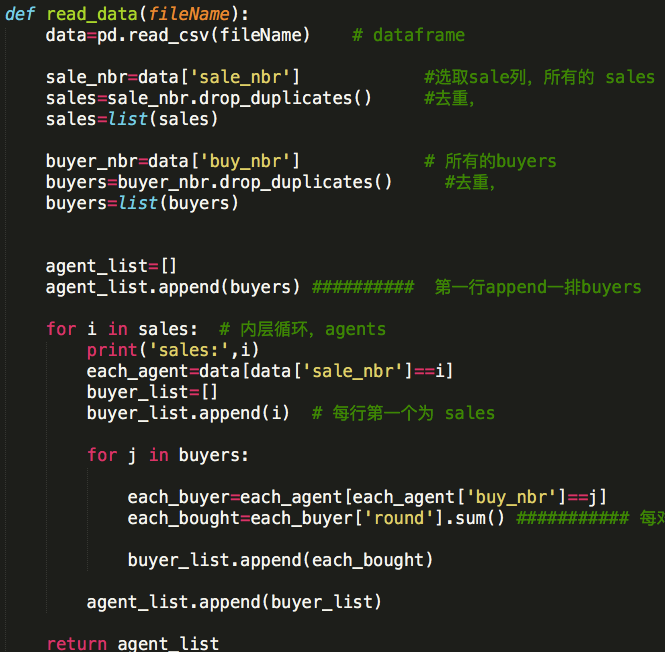
2）其次，求出每对sale-buyer之间的链接关系，即每个sale对应的buyer分别是谁，分别在什么时间段购买了多少。

数据如下：



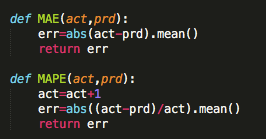
其中，第一行记录所有buyer，第一列记录所有sales。

代码如下：

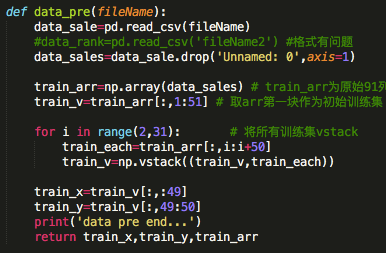


4.3 模型预测

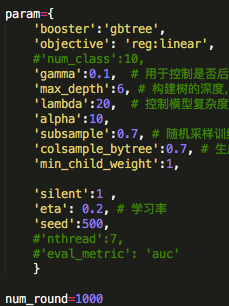
1）用当前机器学习中最热门的xgboost模型对未来销售情况进行预测，引入测评使用的loos function进行调整：



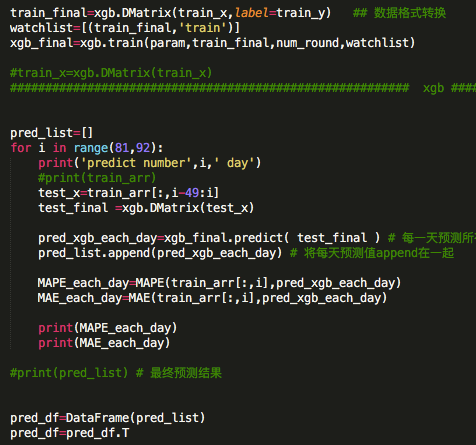
2）对数据进行再次处理，得到模型可以使用的矩阵形式：



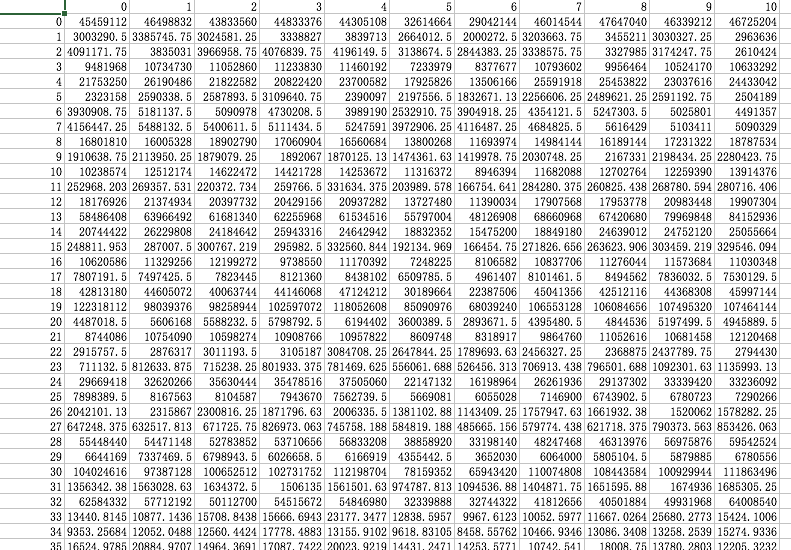
3）模型使用的参数：



4）模型使用：



5）结果：

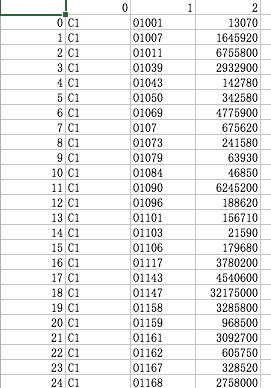


其中，第一排为天数，即待预测的81至91天；第一列记录所有销售商。每一条数据表示某销售商在某天的预测销量。

4.4 导出结果

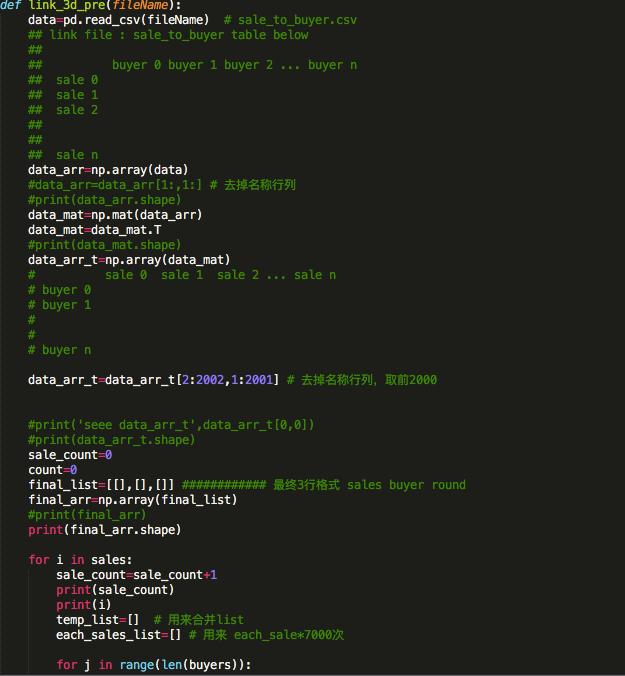
1）将预测结果与真实值之间的链接进行匹配

由得到的链接结果，即sales-buyer对的链接，将数据初步转化为待提交的格式：



其中，第一列为所有销售商，第二列为某sales对应的所有buyers，第三列为sales-buyer对的销售额（总量）。

代码为：



2）将预测结果与真实值之间的链接进行匹配

由预测的待预测日的链接关系和预测的销售额及销售量，将二者结合起来，得到最终第81至91天的所有sales-buyers的预测结果。

格式如下：



代码如下：

