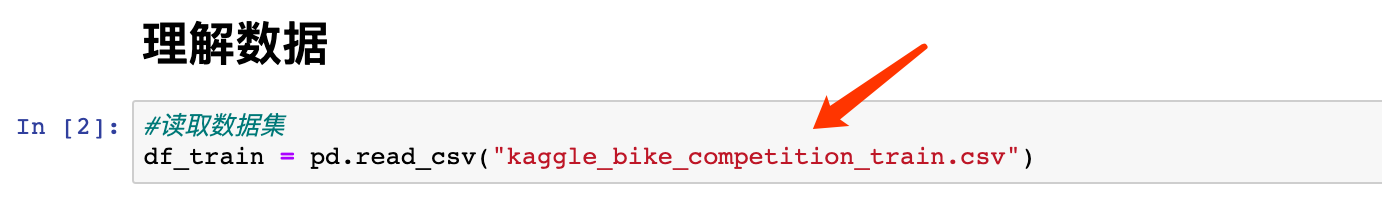
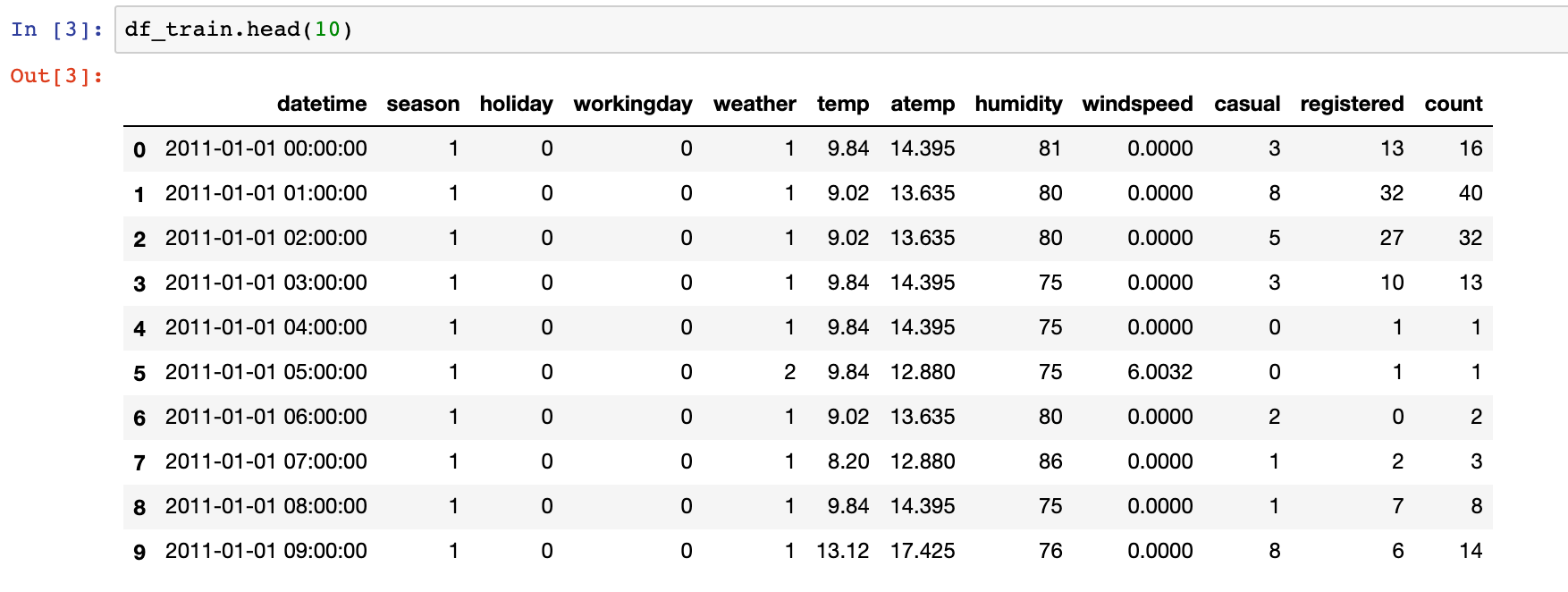
**使用说明书**

1. 本次大作业我主要是在jupyter notebook上进行操作，当然我也写了.py文件，.py文件直接点运行即可，下面主要讲解.ipynb文件的使用说明。
2. 读取数据，配置好路径

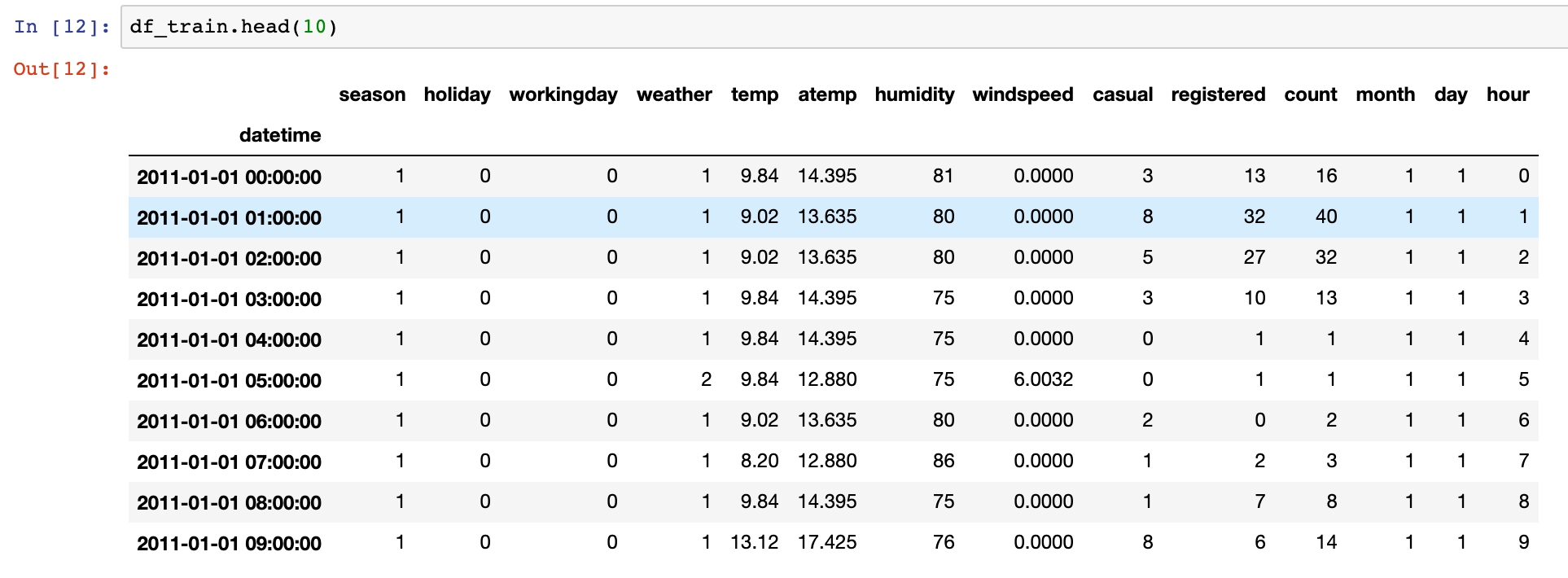




三、数据清洗



关键步骤：将datetime这一列设为索引（方便后续将结果写入文件）



另外，需要将特征和标签进行拆分

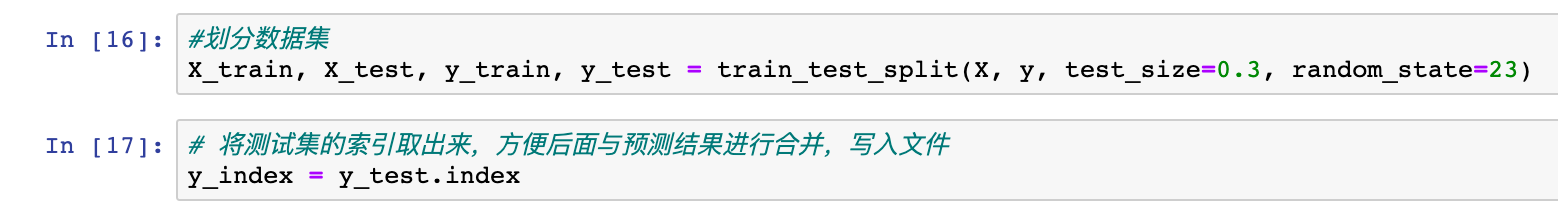
****

1. 构建模型和模型评估

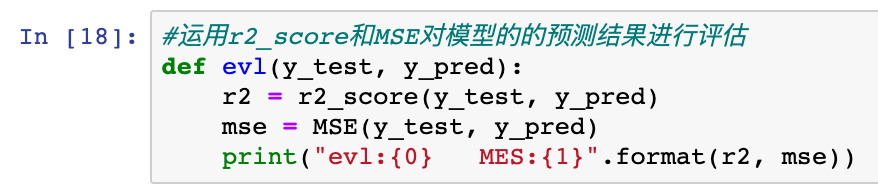
导入sklearn包



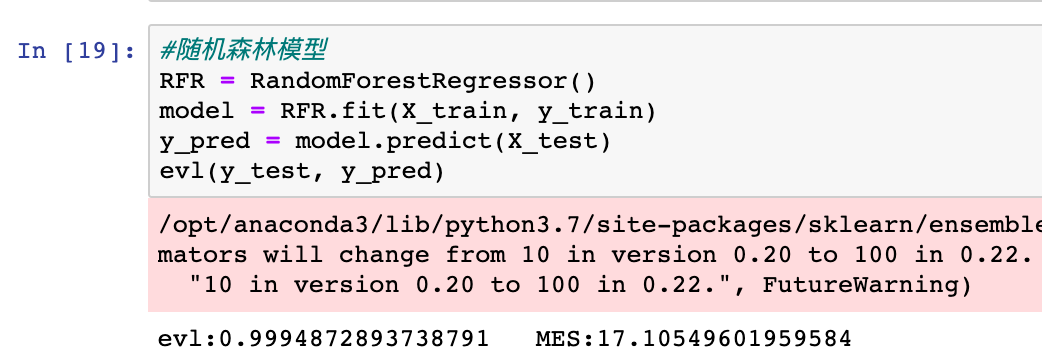
划分数据集



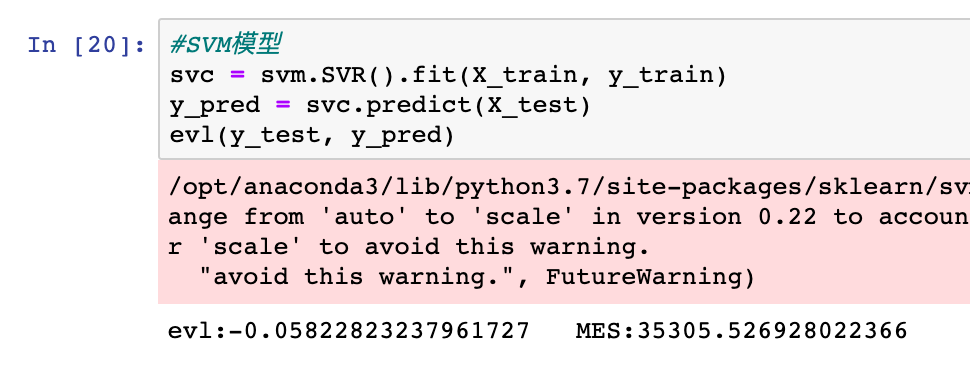
使用r2\_score来评估模型拟合度，MSE为模型的损失Loss



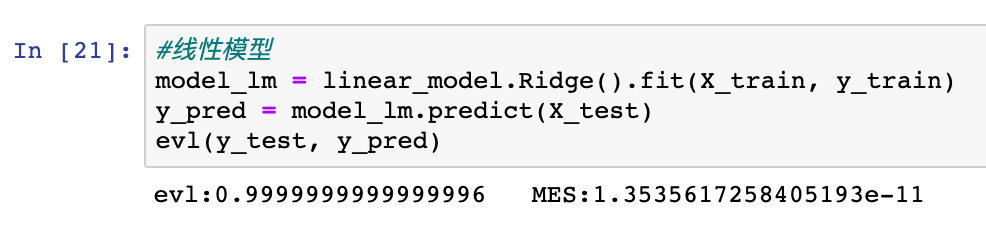
随机森林模型结果：



SVM模型结果：



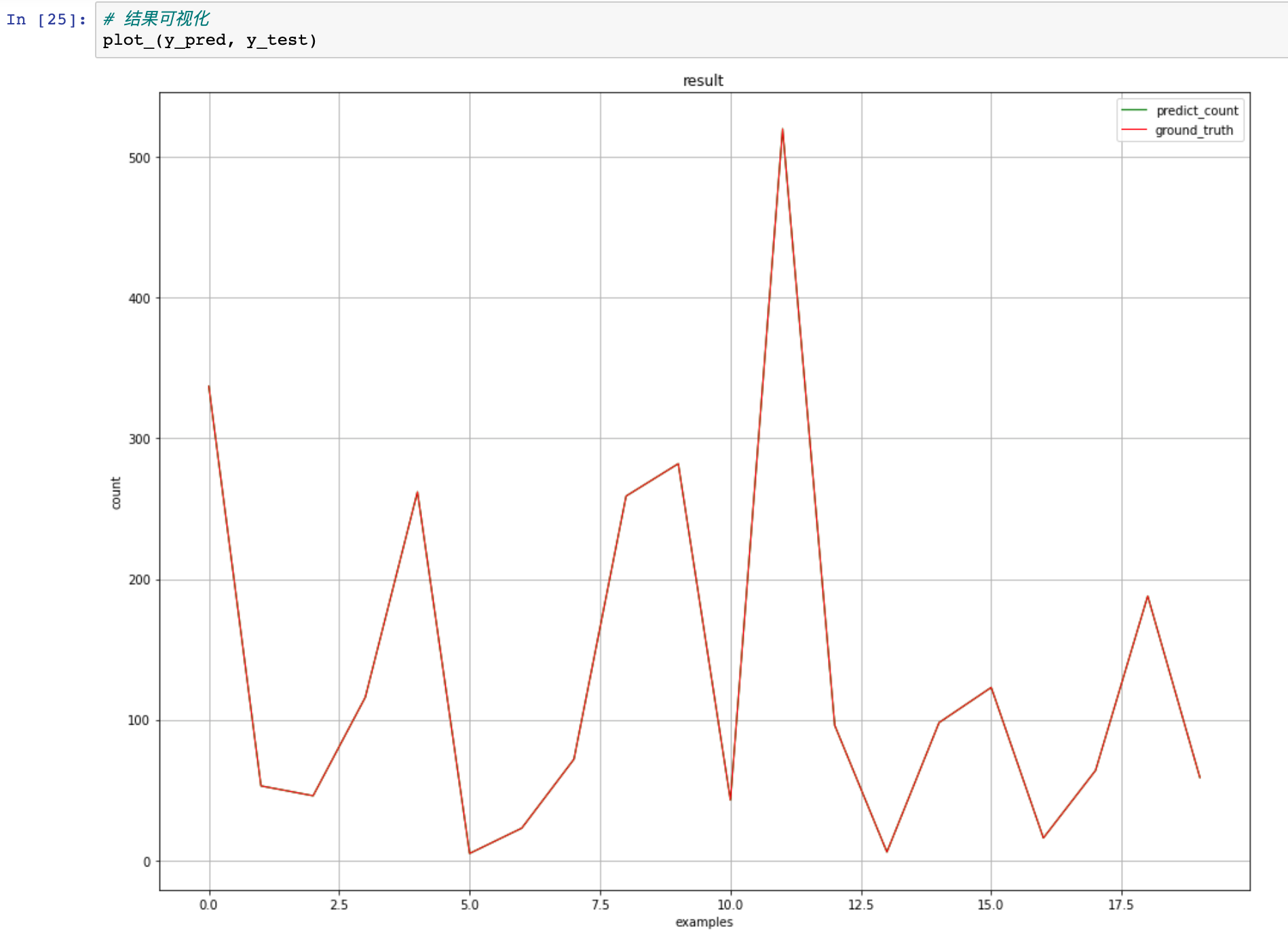
线性模型结果：



数据可视化：

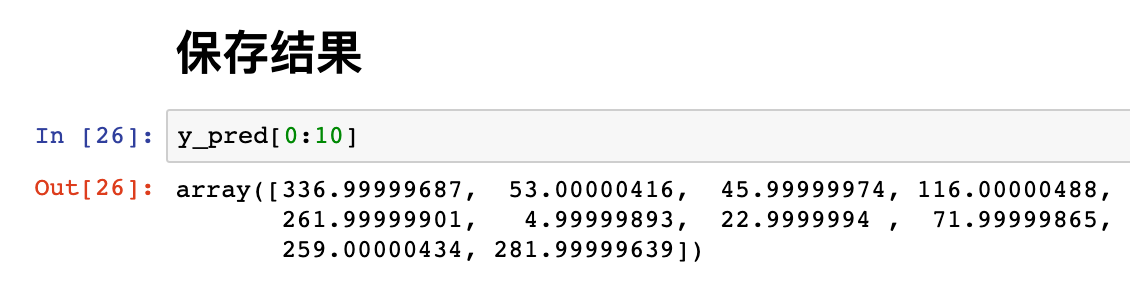


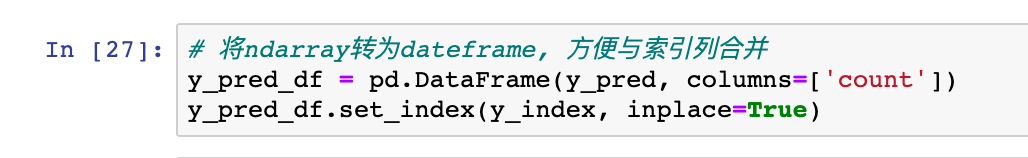
可视化结果（线性模型）：



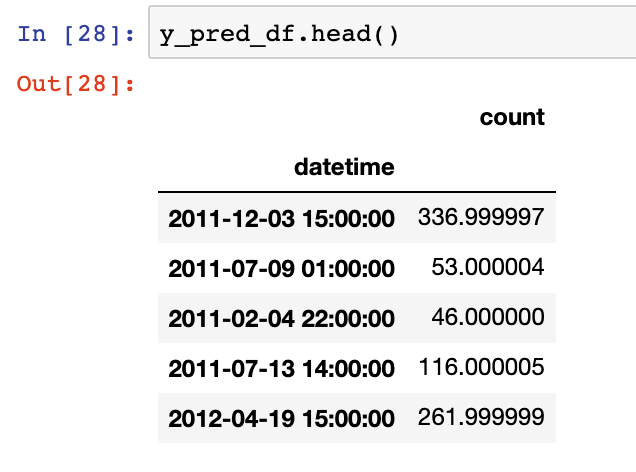
1. 保存结果

模型预测的结果的数据类型为ndarray，需要将其转为DataFrame

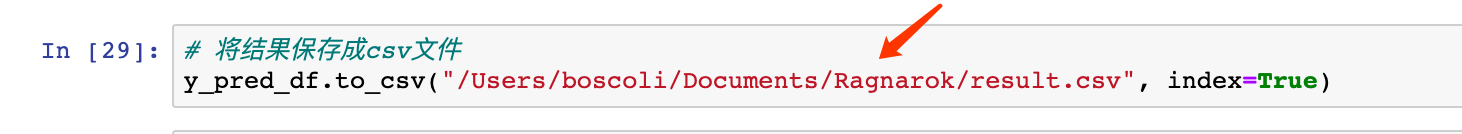




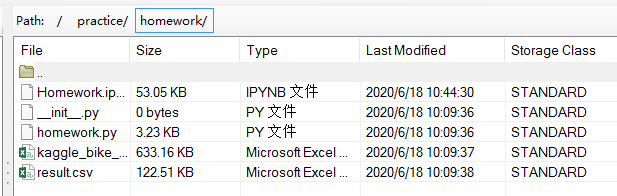
转换后结果如下：



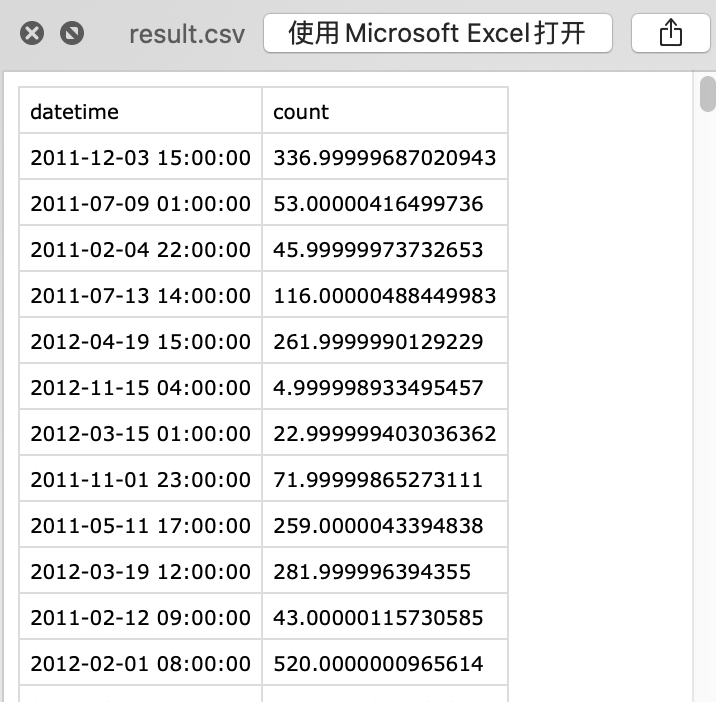
利用.to\_csv方法保存到指定路径



查看S3对应的路径是否保存结果文件



查看result.csv文件



1. 结果分析：上述结果说明线性模型和随机森林模型肯能过拟合了，而SVM则欠拟合