**数据集**

数据集（下载链接见文末附件）来自某只特定的股票，数据中第一二三四五列分别表示

1. 明天收盘价相对今天收盘价的回报率
2. 今天开盘价相对昨天收盘价的回报率
3. 今天最高价相对昨天收盘价的回报率
4. 今天最低价相对昨天收盘价的回报率
5. 今天收盘价相对昨天收盘价的回报率

从上而下对应着时序从过去到现在

**问题**

请使用Python，根据上述的数据集，尝试用第二三四五列数据构建任何有效的模型，预测第一列数据。

需要注意的是，你当然可以使用任何历史上已经发生过的数据来做预测，而不仅限于使用当日的数据。但是尤其要小心的是不能使用任何未来的数据，也即在当前的数据行后面的任何数据。

对于所采用的方法包括但不限于如下

* 基于规则的交易
* 设计和实现有预测力的特征和指标
* 线性回归等统计学方法
* 神经网络等机器学习类方法

**提交**

需要提交两个文件：

* python文件
* 和实验报告

**python文件**需要满足如下要求

1. 要求能够在与dataset.csv格式相同的文件上直接运行，并给出每一行的预测值：  
   *python3 your\_model\_file.py dataset.csv > output\_file*
2. output\_file每一行格式：  
   *sample\_id,prediction\_value*  
   *1,0.58*  
   *2,-0.33*  
   *3,1.39*

**实验报告文件**需要包含如下内容

分析并解释结果，采用但不仅限于如下指标

* 交叉检验的正确率
* R 方统计量
* 盈亏的夏普率
* 预测的正确率

要求：

1. 想法明确
2. 清晰的描述流程
3. 客观的评价
4. **最后预测的准确率高低不是我们评价好坏的唯一指标，过程和思路同样重要。**