

北京天演资本管理有限公司笔试说明

您好，关于笔试环节的专业测试数据请详见附件，具体说明如下：

每一个 `parquet` 为一天的数据（如可用 `pandas.read_parquet` 打开），其中 `Num` 为合约代码、`x1-x46` 为可能的因子（特征），`Vol` 为成交量、`tradeTime` 为时间戳、`MidPrice` 为中间价，研究目标是利用截至当前时刻的特征进行未来 20 个时间间隔的区间收益率预测（即预测目标为 $\text{MidPrice}_{t+20}/\text{MidPrice}_t - 1$ ，注意不要使用集合竞价阶段即 9:30:00 之前以及 14:57:00 之后的数据，不要进行跨交易日的收益率预测）。

（1）说明你觉得可能适用于该预测的模型及原因；

（2）请根据历史数据滚动划分训练集、验证集、测试集，展示所用模型在测试集上的预测性能，并分析预测的有效性；

（3）如有更多时间研究，下一步的方向和计划。

请随附整理后的研究代码。提示：请注意数据的时间序列、稀疏等特征；可以定义某一个或多个预测性能评价指标，比如 `corr(预测收益率,真实收益率)`、`R 方` 等；可以做出适当的问题简化或假设，报告应提供明确清晰的逻辑框架和说明。