

DeepLOB 运行流程

1. 数据准备

- **标的**：任意股指期货（IF/IH/IC/IM 等），其他标的还没实验。
 - **必需数据**：盘口 L1-L5 买卖价量（共 20 列）
 - **数据频率**：500 毫秒/行，仅保留交易时段
 - **整体数据数量**：
 - 小于 20 天：样本不足
 - 大于 60 天：噪音过多、运行时间过长
 - **运行时间**：约 30 分钟（取决于天数与 gpu）
-

2. 窗口设定

- **未来预测窗口**： ≤ 6 秒（推荐 1-4 秒）
 - **回看窗口**：设为未来预测窗口的 2 倍（经验最佳）
 - 窗口越长 \rightarrow 噪音越多，预测效果下降
 - 窗口越短 \rightarrow α 越小，0 和 2 区别越不明显
-

3. 标签生成 & α 搜索

1. **收益率计算**：用 mid
 2. **三分类规则**：
 - 强跌（0）： $r < -\alpha$
 - 中性（1）： $-\alpha \leq r \leq \alpha$
 - 强涨（2）： $r > \alpha$
 3. **硬性平衡条件**：任一类样本数 \leq 另一类的 2 倍
 4. **α 搜索**：
 - 程序会自动根据其他参数算出最优三类分布 α
 - 用户可根据推荐 α 手动打入
 5. **无法满足条件**：需要直接停止（class_weight / 下采样无意义）
-

4. 数据切片

- 按 W 构造样本
 - 步长 **STRIDE** = 1 (滑动一步生成一个样本)
 - 不用下采样，stride可以根据运行时间略微加大 (不建议)
 - 保留原始价格数据 → 回测时用真实成交价
-

5. 数据标准化

- **五天标准化**：舍弃掉前五天，然后之后计算前五天rolling均值、标准差并做 z-score 标准化
-

6. 测试与回测

- **目标胜率**： $\approx 50\%$
- **评估指标**：Macro-F1、混淆矩阵、逐日稳定性