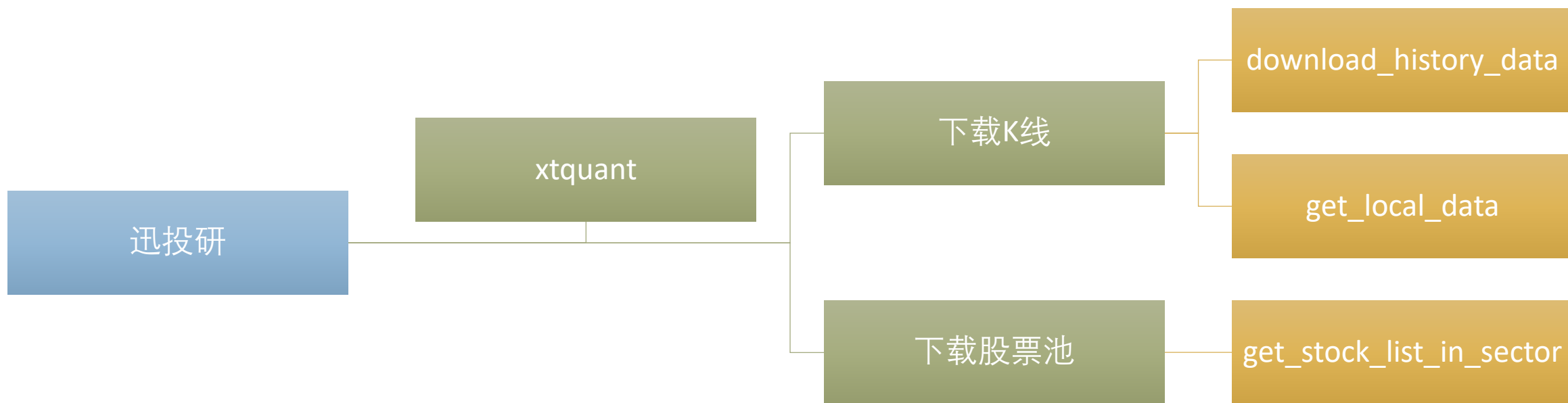


Qlib投研平台学习系列 - 3

By 陈晓优 2024-3-2

上海韦纳软件科技有限公司

数据函数梳理



交易日历数据获取

下载沪深300指数K线

- 不会有停牌情况

遍历提取交易日数据

- 基于bar.datetime

打开txt文件逐行写入

- 记得“\n”换行符

股票池信息数据更新

说明

- 股票池信息本质是指数成分数据
- 通常由中证指数公司发布，需要购买

获取

- 目前迅投研的get_stok_list_in_sector仅能获取最新数据
- 通过其他渠道准备了CSV文件，后续实盘需要自行准备

K线历史数据标准化处理

下载K线数据

- 原始K线和复权K线

计算复权因子

- 用原始K线除转换K线

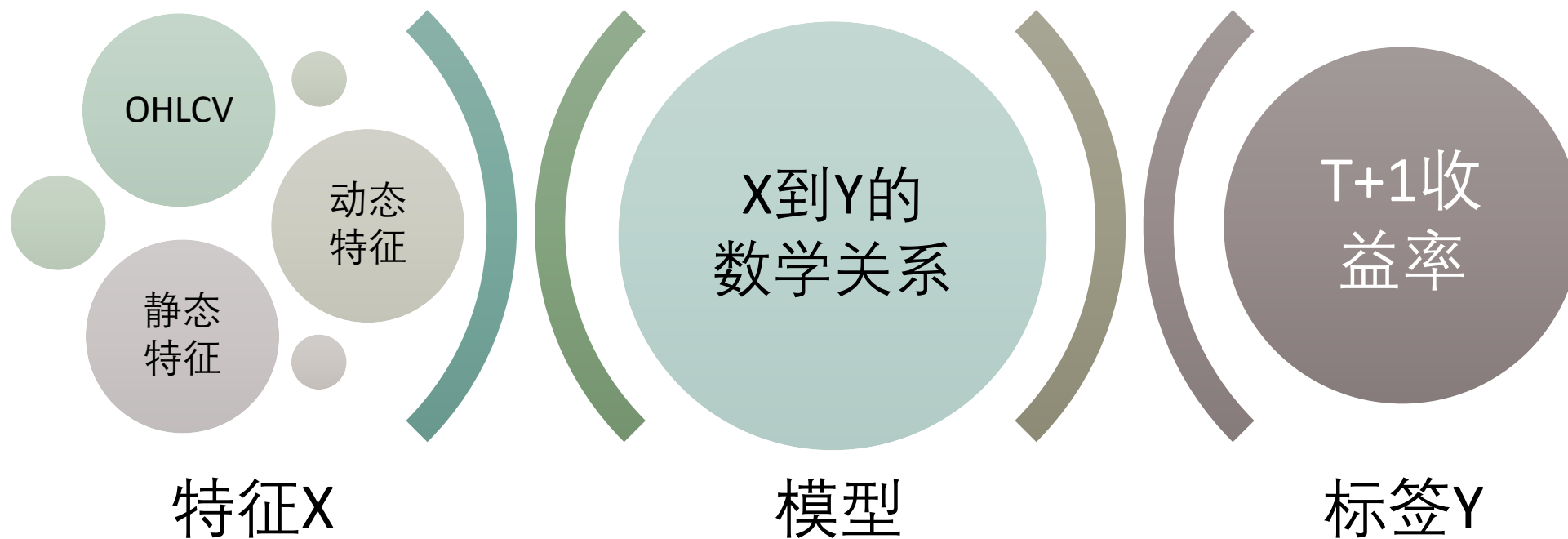
转换复权K线

- 全部除以首日收盘价

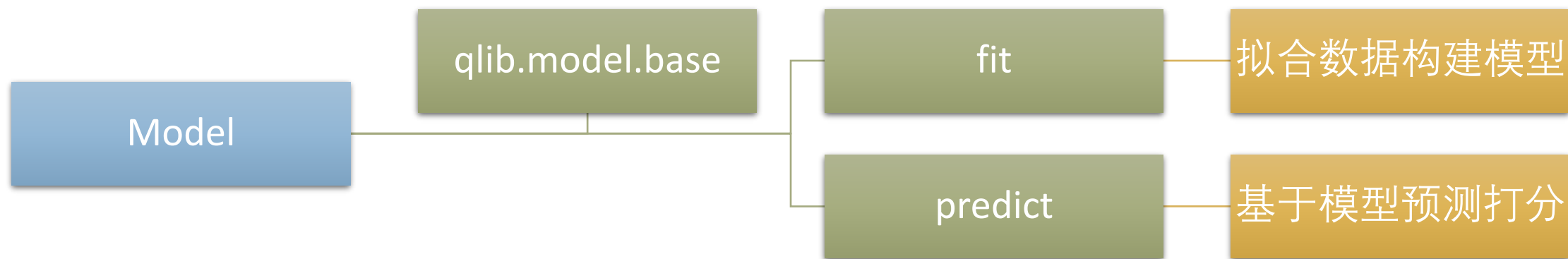
保存bin文件

- 首行放置日期索引值

模型核心应用



模型模板类



三大类模型

线性模型

LinearModel

sklearn

决策树

LGBModel

lightgbm

神经网络

DNNModelPytorch

pytorch

以线性模型为例

qlib.contrib.model.linear

创建模型类

- 继承Model父类

fit函数

- 训练集数据拟合

__init__函数

- 模型配置参数传入

predict函数

- 测试集数据预测



线性回归算法

OLS

- 最小二乘法回归

NNLS

- 非负最小二乘回归

RIDGE

- 岭回归

LASSO

- 最小绝对收缩和选择回归

模型配置参数

estimator

- 回归算法名称缩写

alpha

- 回归正则化参数

fit_intercept

- 是否要拟合截距项

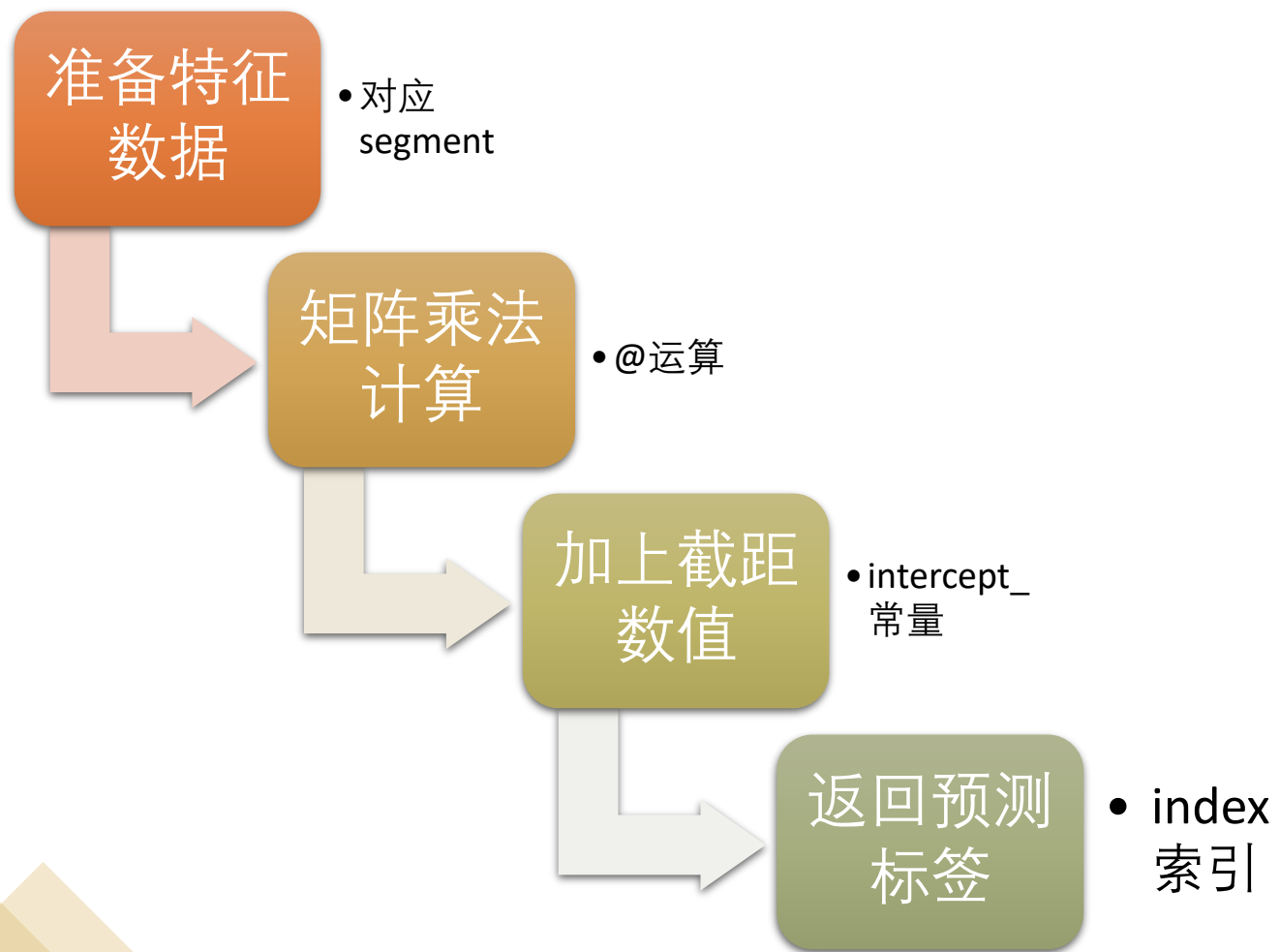
include_valid

- 是否要包含校验集

fit函数逻辑流程



predict函数逻辑流程



已集成的预测模型对比

社区结果

- 位于examples/benchmarks目录下
- 包括整理好的表格以及测试脚本配置

简单总结

- 决策树类模型普遍不错：LightGBM、XGBoost、CatBoost
- 不少神经网络类模型的效果还不如简单的线性模型

显卡GPU支持



PyTorch

CUDNN

CUDA

显卡驱动

显卡硬件

阿里云GPU服务器

服务器购买

- GPU计算型实例，Nvidia A10，按需付费选最便宜的
- Ubuntu 22.04系统，记得选择自动安装CUDA驱动

环境安装

- 重要!!! 安装2.0.1版本的PyTorch: `pip install torch==2.0.1`
- 检查CUDA可以正常使用: `import torch; torch.cuda.is_available()`
- 安装图形界面（参考之前Ubuntu社区活动）和Qlib相关模块

运行程序

- 使用命令工具下载测试数据
- 上传MLP的测试Notebook并运行回测投研任务



- 24年【机器学习CTA】小班课将在3月底开始
- VeighNa社区公众号: [vnpy-community](#)
- 各种入门教程、进阶经验、公告发布等等