**2022秋冬学期量化投资作业汇总**

**何老师作业：**

**20221116作业：**

基于课堂展示的“机器学习选股范例”代码进行改造：

1. 原始代码为静态训练，可否改成滚动训练，如每次使用最近N个月数据训练，每半年更新一次模型？（提示：增加循环即可）
2. 原始代码未涉及交叉验证调参，验证集未实际发挥作用，可否添加调参模块？（提示：sklearn有现成工具）
3. 原始代码仅使用单一模型训练，其他机器学习方法表现如何，这些方法内部相关性如何？可否对低相关模型进行集成，进一步提升表现？（提示：绝大部分方法sklearn有现成工具；神经网络可使用tf、pytorch等编写，有一定难度）
4. 原始策略在2017年后经历了较长的超额收益回撤（相对沪深300），可能是什么原因？有什么方法可以改进？

作业可使用Jupyter Notebook完成，请大家独立完成。作业提交日期为2022年11月30日，请将作业分别交给所在组的助教。

**20221123作业：**

理解课堂展示的“NN-demo-ZJU”项目，并进行改造：

1. “04\_train\_model\_nn\_static.py”主程序中，定义模型时设置了一系列超参数（d\_feat、hidden\_size等），叙述这些超参数的含义和作用。
2. 针对“04\_train\_model\_nn\_static.py”进行扩展，测试下列关键超参数的不同取值对模型的影响，并分析原因：
3. hidden\_size
4. lr
5. “model/pytorch\_nn.py”是核心模型方法，其中神经网络模型在NNModel类中定义，叙述网络结构，网络总共有多少自由参数？每层各有多少自由参数？

BN层课堂未介绍，可搜索学习相关原理：

https://pytorch.org/docs/stable/generated/torch.nn.BatchNorm1d.html

https://zhuanlan.zhihu.com/p/24810318

1. 针对“model/pytorch\_nn.py”进行扩展，将损失函数改为：
2. IC值。
3. 上节课介绍的半衰加权MSE。每个截面对股票收益排序，高收益股票给更高权重，半衰期可设为0.5，即收益排名第一的股票权重为1，排名中位数的股票权重为0.5，排名后25%分位数的股票权重为0.25。
4. 上节课介绍的logistic有序回归，类别数可设为5。

考察这些损失函数下，模型IC值表现。

**20221130作业：**

20221116课堂作业的第一项任务中，我们对xgboost选股模型进行了滚动训练，如每半年重新训练一次模型。请参考课堂展示的demo\_3\_explainable案例，针对历史上不同时刻训练得到的模型，计算各特征重要性、shap值等解释性指标的变化情况，并尝试分析原因。

本次作业和第二次作业提交时间为2022年12月21日

**20221207作业：**

理解课堂展示的“demo\_4\_gan”项目，并将原始GAN改造为WGAN-GP。

参考资料：

原理部分可参考：Gulrajani, I. , Ahmed, F. , Arjovsky, M. , Dumoulin, V. , & Courville, A. . (2017). Improved training of wasserstein gans.

代码部分可参考：Pardo, F.D. (2019). Enriching Financial Datasets with Generative Adversarial Networks.

https://github.com/FernandoDeMeer/Mitigating-Overfitting-Experiment

本次作业没有看上去那么难，给四个提示：

1. 原始GAN模型核心代码为core\_gan.py，可以仿写一个core\_wgan\_gp.py。
2. 网络结构：D需要删除什么组件？
3. 损失函数：原loss调用nn.BCELoss()，新loss不需要调用函数，更简单。
4. GP梯度惩罚项：可参考github，写函数d\_loss\_grad = compute\_gradient\_penalty(D, real\_data, fake\_data)

本次作业提交日期为2023年1月11日

**彭老师作业：**

多因子模型

* 基于python或其他语言编写组合回测代码，输出收益序列及关键指标
* 借鉴国泰191或其他研报构建10-15个alpha因子（越多越好）
* 通过线性模型和至少一个机器学习模型进行因子组合，分别合成信号
* 通过分层回测或回归考察两模型融合后的信号表现差异

Barra风险模型

* 基于python或其他语言编写组合回测代码，输出收益序列及关键指标
* 完成barra风险因子中的价量因子部分
* 选定一个股票池（如中证500）进行多元线性回归，获取行业因子收益和风险因子收益
* 通过因子收益数据获取因子协方差矩阵，每月底或其他周期进行一次优化（以最小化预期风险为目标）
* 通过回测对比考察传统协方差矩阵样本外模型风险预测能力