

崔卓毅

+86 17368718862 carlcui3824@gmail.com

教育经历

南京大学，工程管理学院	中国，南京
金融工程学 GPA: 4.53/5.00 (前 10%)	2021.8 至今
<ul style="list-style-type: none">核心课程: 面向对象编程(C++), 数据库和信息系统, 数据结构与算法, 概率论与数理统计, 金融工程学, 固定收益证券, 应用统计, 系统建模与仿真, 数值分析	
加州大学伯克利分校 (交换)	美国，伯克利
金融工程学 GPA: 3.96/4.00	2023.1-2023.12
<ul style="list-style-type: none">核心课程: 机器学习, 深度学习, 人工智能导论, 随机过程, 资产定价和投资组合, 计量金融学, 计算机程序构造与解释, 离散数学	

实习经历

量桥投资管理有限公司	中国，上海
策略部，股票量化实习生	2024.1-2024.3
<ul style="list-style-type: none">编程实现股票因子回测框架构建，实现因子主要指标（夏普比率、最大回撤、超额收益率等）计算、中性化处理、分组收益率图表绘制等功能。阅读研报并整理出超过 100 个股票市场高频因子，并将其归类为量价相关性因子、反转因子、动量因子、内幕交易因子等。利用 dolphinDB 处理 level2 行情数据，并对因子做降频处理，实现超 20 个高频因子的构建与回测。	

科研经历

芝加哥大学，布斯商学院	美国，芝加哥
科研助理(远程)	2023.10 至今
<ul style="list-style-type: none">训练自编码器 autoencoder 对高噪声数据进行特征提取，在此环境下探索不同初始化、归一化的方法对训练稳定性的影响，并比较 early_stopping 和 drop_out 在避免模型过拟合方面的效果。利用特征值比例检验方法研究 PCA 在数据降维时的最优参数选择，并比较 PCA 和不同核函数下的 KernelPCA 在特征提取方面的表现。研究在信噪比极低的情况下，L1、L2 正则化运用于线性回归和神经网络时避免过拟合的效果，并与基准模型 OLS 进行比较，发现 L1 正则化在此情境下对信号的误判率较高，不适合进行弱信号识别。使用网格搜索和 optuna 对神经网络、随机森林和 XGBOOST 等模型进行参数调整，用于实证研究具有大量预测变量的经济学数据集。	

项目经历

Transformer 上下文学习机制研究	2023.10-2023.12
<ul style="list-style-type: none">通过与其他回归模型（OLS, XGBoost 等）对比,研究 transformer 在上下文中学习简单函数类的能力，并调查了模型在不同函数类混合样本中学习和区分不同任务的能力。编写程序实现对多维线性函数、决策树、K-NN 等函数的模拟，生成了训练模型所需的数据。搭建了以 GPT-2 为基础的 Trasnformer 模型，并在 GPU 上利用生成的数据集进行模型训练。在验证阶段绘制均方误差图，描述模型误差和上下文样本数量之间的关系。	

LLM 检测个人信息

2024.2-2024.3

- 利用大语言模型自动检测学生作文中的个人信息，减少学生个人信息泄露。
- 对文本数据清洗整理后进行分词并添加标签，利用分词后的数据对 DeBERTa-v3 模型进行全量微调 (FFT)，实现在测试集上 0.963 的 F5_score。
- 提出对文本数据预先切割，并利用 stride 来缩减输入文本长度的方法训练模型，避免了由模型分词器大小限制导致的信息流失，将训练集上的 F5_score 提高 0.13。

校园组织

南京大学英语俱乐部 口语角负责人	2021.9-2023.1
<ul style="list-style-type: none">组织策划口语角圣诞节、万圣节活动，在活动中担任话题发起人，引导成员展开讨论负责口语角活动的话题撰写工作，产出当周活动中所要讨论的话题	

相关技能

- 编程: 熟练使用 python,c++, linux
- 深度学习: 熟练掌握深度学习相关知识，熟悉 pytorch 和 tensorflow 两种主流深度学习框架，有过使用 w&b 训练和微调大模型的经验
- 数据分析: 熟练使用 python 常用的数据分析工具和算法（numpy, pandas, matplotlib, seaborn 等）