**《创新实践》课程说明**



**导师简介**：李珣，毕业于浙江大学，博士，现任职于计算机学院认知与神经计算研究所，主持国家自然科学基金青年基金项目1项、浙江省自然科学基金一般项目1项，参与美国自然科学基金、美国国立卫生署基金、中国国家自然科学基金、浙江省自然科学基金项目十余项，主要研究方向：人工智能、深度学习、机器学习。项目经验：轴承和电机故障检测的时间序列深度学习应用(LSTM模型)、金融市场的深度学习和强化学习应用(LSTM + SVM + DQN)、应用于ETF指数的T+0高频交易策略模型

联系方式：13958065176, xunli@hdu.edu.cn

**项目方向：人工智能(AI)在金融科技(FinTech)的应用**

**项目内容介绍：**

随着深度学习的突破性进展，人工智能进入井喷的寒武纪。在过去的两年内，全球科技巨头在人工智能领域总投入超过300亿美元，中国企业和投资界投入80亿美元，接近一半的信息科技新产品源自人工智能技术。十九大报告更是将人工智能和大数据定位为国家战略。

红杉资本合伙人沈南鹏在JDD-2017京东金融全球数据探索者大会上指出“金融行业作为一种高度数据化的行业，是AI最好的应用行业”。AI在金融领域的应用包括：智能投资顾问(自然语言处理、量化组合投资等)、算法交易(规则挖掘、关联分析、高频交易策略挖掘等)、身份认证(人脸识别、数字签名等)、风险控制(客户画像、欺诈检测等)、精准营销(个性化推荐等)。

本项目将围绕AI与FinTech开展实践学习，涉及**数据准备**、**数据挖掘建模**、**软件开发**等几个方面，通过学习支持向量机、支持向量回归、卷积神经网络、递归神经网络、强化学习等算法和代码，构建量化投资和算法交易模型。

**项目进度安排：**

本项目将通过四个学期分阶段学习机器学习、深度学习、强化学习的基本算法和源码，搭建智能投资顾问和算法交易模型。

**第3学期：数据准备 + 机器学习实践**

本项目数据包括证券交易所的**历史和实时股票数据**，ETF指数的**历史和实时数据，**社交网络和财经网站的**文章和评论**。需要解决的问题：1、股票数据、财经评论内容的处理；2、分类和回归模型构建及应用；3、系统架构的初步设计。

**任务**：针对股票数据，应用支持向量机和支持向量回归方法构建选股模型，结合博弈论设计交易策略。

**学习内容**：

|  |  |  |
| --- | --- | --- |
| **0项目** | **0.1子项目** | **0.1.1内容** |
| **1 实践** | **1.1文件读写** | 1.1.1 txt, CSV文件读写 |
| 1.1.2 libsvm数据格式 |
| **1.2 libsvm安装** | 1.2.1 libsvm库安装 |
| 1.2.2 libsvm3.22源码 |
| **1.3 libsvm应用** | 1.3.1 svm股票数据分类 |
| 1.3.2 svr股票数据回归 |
| **1.4 库** | 1.4.1 numpy数学计算 |
| 1.4.2 matplotlib绘图 |
| 1.4.3 gnuplot |
| **2 理论** | **2.1 机器学习过程** | **2.1.1 特征提取+分类概念** |
| **2.2 神经网络** | **2.2.1 感知器** |
| **2.2.2 BP神经网络** |
| **2.3 回归分析** | **2.3.1 回归分析** |
| **2.4 支持向量机** | **2.4.1 SVM** |
| **2.5 支持向量回归** | **2.5.1 SVR** |
| **3 金融** | **3.1 资本资产定价模型** | **3.1.1 CAPM** |
| **3.1.2 随机游走** |
| **3.2 组合量化投资** | **3.2.1 量化投资组合管理** |
| **3.3 博弈论** | **3.3.1 囚徒困境** |
| **3.3.2 智猪理论** |
| **4 预处理** | **4.1 特征** | 4.1.1 股票数据特征 |
| **4.2 类别标签** | 4.2.1 股价涨跌的类别标签 |
| **4.3 标准化和归一化** | **4.3.1 标准化** |
| **4.3.2 归一化** |
| **4.4 缺失值和异常值** | 4.4.1 缺失值的处理 |
| **4.5 样本非均衡问题** | 4.5.1 欠采样和重采样 |
|  | 4.5.2 合成样本 |
| **4.6 降维** | 4.6.1 PCA等方法 |
| **5 参数优化** | **5.1 svm调参** | 5.1.1 grid.py |
| **5.2 svr调参** | 5.2.1 gridregression.py |
| **6 模型评价** | **6.1 评价指标** | 6.1.1 敏感度和特异性 |
| 6.1.2 准确率、召回率和F值 |
| 6.1.3 ROC |
| **6.2 交叉验证** | 6.2.1 K-fold cross validation |
|  |  |  |

**阶段性成果**：

掌握机器学习的核心算法、项目实施流程，根据操作系统、python版本、开源项目库设计一套python程序，基于股票数据设计组合投资量化模型

|  |  |  |
| --- | --- | --- |
| **理论** | **源码** | **应用** |
| **机器学习项目流程** | **Libsvm开源项目** | **组合投资模型** |
| **支持向量机、回归** | **机器学习项目之数据读写** |  |
| **资本资产定价模型和组合投资** | **机器学习项目之特征提取** |  |
| **博弈论** | **机器学习项目之算法应用** |  |
|  | **机器学习项目之模型评价** |  |

**第4学期：深度学习算法应用**

**任务：**研究基于tensorflow的深度学习模型RNN和LSTM的应用， RNN和 LSTM做文本识别、时间序列分析。

**学习内容**：

|  |  |  |
| --- | --- | --- |
| **0项目** | **0.1子项目** | **0.1.1内容** |
| **1 实践** | **1.1 Tensorflow安装** | **1.1.1 Tensorflow1.2安装** |
| **1.1.2 Tensorflow库解读** |
| **1.2 RNN** | **1.2.1 自然语言处理** |
| **1.2.2 时间序列处理** |
| **1.3 LSTM** | **1.3.1 自然语言处理** |
| **1.3.2 时间序列处理** |
| **1.4 GPU** | **1.4.1 GPU架构** |
| **1.4.2 GPU+CPU架构下的tensorflow应用** |
| **2 理论** | **2.1 循环神经网络** | **2.1.1 循环神经网络原理** |
| **2.1.2 序列到序列架构** |
| **2.1.3 双向RNN** |
| **2.1.4 门控单元RNN** |
| **2.1.5 门控单元LSTM** |
| **3 金融** | **3.1 期权定价模型** | **3.1.1 Black-Scholes-Merton期权定价模型** |
| **3.1.2 蒙特卡罗期权定价** |
| **3.2 波动率** | **3.2.1 Garch模型** |

**阶段性成果**：

|  |  |  |
| --- | --- | --- |
| **理论** | **源码** | **应用** |
| **循环神经网络** | **Tensorflow源码** | **自然语言处理** |
| **长短记忆门控单元** | **RNN、LSTM** | **时间序列处理** |
| **期权定价模型** | **期权定价模型的python实现** |  |
| **波动率** | **Garch模型的python实现** |  |

**第5学期：深度学习和强化学习算法应用**

**任务**：研究基于tensorflow的深度学习模型CNN在文本识别的应用， 结合深度学习和强化学习设计算法交易模型。

**学习内容**：

|  |  |  |
| --- | --- | --- |
| **0项目** | **0.1子项目** | **0.1.1内容** |
| **1 实践** | **1.1 CNN** | **1.1.1 CNN源码** |
| **1.1.2 文本识别** |
| **1.2 强化学习** | **1.2.1 马尔科夫决策过程** |
| **1.2.2 蒙特卡罗强化学习** |
| **1.2.3 DQN** |
| **1.2.4 基于策略梯度的强化学习** |
| **2 理论** | **2.1 卷积神经网络** | **2.1.1 卷积** |
| **2.1.2 池化** |
| **2.2 强化学习** | **2.2.1 序贯决策问题** |
| **2.2.2 马尔科夫过程** |
| **2.2.3 蒙特卡罗统计** |
| **2.2.4 Deep Q Network** |
| **3 金融** | **3.1 高频交易** | **3.1.1 同位托管** |
| **3.1.2 统计套利** |
| **3.1.3 趋势预测** |

**阶段性成果**：

|  |  |  |
| --- | --- | --- |
| **理论** | **源码** | **应用** |
| **CNN** | **CNN** | **文本识别** |
| **强化学习** | **DQN** | **序贯决策问题的智能决策** |
| **高频交易** | **高频交易算法的python实现** |  |

**第6学期：**

**任务**: 完善底层算法，设计并优化GUI界面，提升软件性能。

**学习内容**：

|  |  |  |
| --- | --- | --- |
| **0项目** | **0.1子项目** | **0.1.1内容** |
| **1 实践** | **1.1 PyQt** | **1.1.1 PyQt可视化课程** |
|  | **1.2 完善SVM、Tensorflow、DQN代码** |  |

**最终成果**：

能够独立完成金融量化投资、算法交易、自然语言处理等项目的设计和实践，能够综合运用SVM、CNN、RNN、DQN等算法设计智能决策系统，能够应用PyQt设计可视化软件系统。

**项目协作单位：**

金融业：国泰君安、同花顺

AI技术：阿里巴巴、华为

**项目亮点：**

本项目作为Fintech的入门，将使参与者了解金融基本理论和高频交易算法设计流程，掌握当前主流的机器学习、深度学习、强化学习模型及其开源代码的应用、掌握PyQt可视化编程，学会软件系统的设计与开发。

本项目授课过程中将尽可能地邀请上述协作单位（国泰君安、同花顺、阿里巴巴、华为等）专业人员进行技术指导。

本项目所涉及的NLP、Deep Learning是当前最为热门的AI技术，通过本项目的学习将为有志于AI、Fintech的同学打下坚实的技术基础。

**面向专业：计算机科学与技术、软件工程、物联网工程**