

## 第一次课：机器学习资产定价入门与线性模型应用

- **主题与内容结构：**介绍机器学习在资产定价中的基本概念和作用，重点关注线性模型及正则化方法在金融时间序列预测中的应用。首先简述机器学习和传统计量经济方法的差异，解释过拟合、训练集/测试集划分等基础概念。然后以标普500指数（月度或季度**超额收益**预测）为例，探讨多元回归在股市收益预测中的效果和局限（提及如Goyal和Welch对股权溢价预测的研究）。继而引入**正则化线性模型**（岭回归、LASSO等），说明其通过变量选择和防止过拟合来提升预测性能的原理，并讨论在资产定价中的意义（例如从大量宏观和技术指标中筛选有效预测信号）。
- **核心论文：**Chun et al. (2024) 《Forecasting returns with machine learning and optimizing global portfolios》之相关部分。本研究采用多种机器学习模型预测韩元/美元汇率和美韩股市收益，结果显示**LASSO和Elastic Net正则化回归在汇率和股票指数回报预测中表现最佳**，显著优于传统回归模型<sup>1</sup>。该论文权威地证明了使用正则化线性模型整合多种预测因子可提高资产回报预测的准确性，适合作为课程实例。
- **主要Python工具包与建模方法：**使用 `pandas` 与 `numpy` 处理时间序列数据（如下载标普500指数历史数据和宏观指标）；`scikit-learn` 中的线性回归模型（如 `LinearRegression`）以及正则化回归（`Ridge`, `Lasso`, `ElasticNet`）；可能涉及数据集获取的工具（如 `pandas_datareader` 或 `yfinance` 用于从Yahoo Finance获取指数和宏观数据）。模型评估指标包括均方误差（MSE）、预测的  $SR^2$ 、方向性准确率等。
- **课上实践内容建议：**指导学生从网络获取标普500指数历史数据及若干潜在预测因子（例如上一期收益率、波动率指标VIX、经济指标如短期利率或信用利差等），构建训练集和测试集。首先用OLS线性回归模型对标普500下一期收益率做预测，观察结果和存在的问题。随后应用LASSO回归模型进行相同预测，通过调整正则化强度观察变量选择和预测效果变化。让学生对比OLS与LASSO在测试集上的表现（例如MSE或方向判断准确率），体会正则化在防止过拟合和提取有效信号方面的作用<sup>1</sup>。实践过程中包括代码演示如何使用 `sklearn` 训练模型、调参（如使用交叉验证寻找LASSO的最佳正则化系数）、以及可视化预测值与实际值的对比图表。
- **课后作业设计方向：**巩固线性模型在资产定价中的应用。作业可要求学生**复现**课堂模型但更换数据情境，例如预测另一个市场指数的收益（如**KOSPI指数**或其他股指）或预测汇率（如欧元/美元）的短期变化，让学生运用线性回归和LASSO进行比较分析。作业应包含数据获取、特征整理、模型训练与调优，以及评价模型预测准确度等步骤。鼓励学生报告所选预测因子的重要性，讨论哪些变量被LASSO选择进入模型以及可能的经济含义。通过作业，学生将实践将基本机器学习方法运用于资产定价预测的完整流程。

## 第二次课：非线性模型与时间序列资产定价实践（树模型与汇率预测）

- **主题与内容结构：**本节聚焦于非线性机器学习模型（决策树及集成方法）在金融时间序列资产定价中的应用，重点以外汇汇率预测为案例。首先简述决策树模型及其优缺点，引出随机森林和梯度提升树等集成方法，解释它们如何通过集成多棵树提升预测稳健性和精度。然后将讨论金融时间序列的非线性特征，如汇率变化中的复杂模式，传统线性模型难以捕捉，此时树模型可自动捕捉特征的非线性关系和交互作用。课堂主要案例以**美元兑主要外汇汇率**（例如美元兑欧元或韩元USD/KRW）为对象，展示用随机森林/梯度提升模型进行汇率月度或季度回报率预测。讨论评价非线性模型效果的指标（如MSE、方向性预测准确率，以及通过简单策略计算投资回报表现）。
- **核心论文：**Rossi (2018) 等关于**提升回归树（BRT）应用于股市收益预测**的研究（“Predicting Stock Market Returns with Machine Learning”）是本节核心参考。Rossi论文采用Boosted Trees对美国股市超额收益进行预测，发现利用海量信息的非线性模型相比传统线性模型能够取得更高的样本外预测精度，并能带来显著投资收益改善<sup>2</sup>。这一权威结果凸显了树模型等非线性方法在资产定价预测中的价值。结合该论文，本节还会提及Chun et al. (2024)的发现：树模型（随机森林、LightGBM等）在汇率和股指预测中表现出色，也仅次于LASSO等最佳模型<sup>3</sup><sup>4</sup>。这些文献为我们选取树模型预测汇率提供了依据。

- **主要Python工具包与建模方法：**使用 `scikit-learn` 的决策树与随机森林实现模型训练（如 `DecisionTreeRegressor`，`RandomForestRegressor`），以及梯度提升算法库（如 `LightGBM` 的Python接口或 `sklearn` 的 `GradientBoostingRegressor`）。数据处理仍依赖 `pandas`，用于特征工程可能会用到 `numpy` 及相关时间序列处理库（如生成滞后变量、滚动均值等特征）。如果涉及调整超参数（树的深度、森林树数、学习率等），可用 `sklearn.model_selection.GridSearchCV` 等工具。评估方法包括计算预测误差、方向性判断准确率，及简单策略回溯（例如预测汇率升值贬值并模拟持有相应头寸计算收益）来评估经济意义。
- **课上实践内容建议：**以汇率时间序列预测为主线设计实践。教师可事先准备或带领学生获取汇率数据（例如通过 `yfinance` 获取美元兑欧元日线或月线数据，或使用央行公开数据）。实践步骤包括：对原始汇率数据进行处理（计算对数收益率或百分比变化），构造特征矩阵——如最近若干期的汇率变化、移动平均、波动率指标，甚至可加入股市指数变化等相关市场指标作为特征。然后让学生训练一个**随机森林**模型来预测下一期汇率变动方向或幅度。通过调整随机森林的树数和深度来观察模型在训练集和验证集上的表现，强调避免过拟合。随后训练梯度提升决策树（例如使用 `LightGBM`），比较其与随机森林的预测效果。课堂中应实时计算模型的**样本外**预测误差和准确率，展示非线性模型可能捕捉到的一些模式。例如，可视化模型预测值与实际值的时间序列对比，或绘制变量重要性图表以解释哪些特征在模型中最重要的。让学生体验在金融时间序列中，非线性机器学习模型如何发现传统方法难以察觉的模式，并**验证这些模型在预测准确度上的提升和潜在投资收益**（例如简单计算若根据模型信号进行交易的绩效）。
- **课后作业设计方向：**巩固树模型和集成学习在资产定价中的应用。作业可以要求学生选择**另一类金融时间序列**进行预测练习，如预测**市场波动率指数(VIX)**的变化，或预测**另一货币对（如欧元/美元）的汇率走势**。学生需应用课堂所学的随机森林或梯度提升模型完成从数据获取、特征提取到模型训练、评估的流程。要求对比至少一种基准（例如线性回归或ARIMA模型）的表现，以凸显机器学习非线性模型的效果差异。作业报告应包含模型参数选择过程、变量重要性分析，并探讨模型结果的经济含义（例如模型认为哪些因素对预测目标最重要，可能反映了怎样的市场现象）。通过该作业，学生将进一步锻炼使用Python构建非线性预测模型并应用于真实资产定价问题的能力。

### 第三次课：深度学习与资产定价前沿（横截面股票定价应用）

- **主题与内容结构：**本节聚焦深度学习方法在资产定价中的应用，特别是股票横截面定价问题。首先回顾传统横截面资产定价（如Fama-French因子模型）的框架，指出其局限：特定函数形式、因子数量有限，难以捕捉复杂的非线性关系和大量特征信息。引出机器学习在横截面定价中的新进展：利用大量公司特征（基本面指标、技术指标等）预测个股收益，挖掘新的因子结构。重点介绍**神经网络**模型如何用于拟合资产定价函数，例如多层前馈神经网络捕捉特征与预期收益间的非线性映射，或简单介绍时间序列模型LSTM在金融序列模式提取上的优势。课程核心讨论一篇经典论文：Gu、Kelly和Xiu (2020) 《Empirical Asset Pricing via Machine Learning》，其中比较了多种机器学习方法在股票横截面预期收益预测中的效果<sup>5</sup>。论文结论指出：**树模型和深度神经网络是表现最优的方法**，能够显著提高资产风险溢价的预测精度，其原因在于捕捉了传统线性方法遗漏的非线性特征交互<sup>6</sup>；不同方法识别出相似的主要预测信号（如动量、流动性、波动率等）<sup>7</sup>，显示机器学习能够从海量因子中自动提取有效信息。这些结论体现了深度学习在资产定价研究中的巨大潜力。
- **推荐经典论文：**除上述Gu等(2020)外，可参考Chen、Pelger和Zhu (2019) 《Deep Learning in Asset Pricing》等前沿研究。后者利用深度神经网络直接估计资产定价模型（SDF），在无套利约束下结合**GAN**和**LSTM**网络提取宏观状态和非线性因子，大幅提高定价精度。这类权威研究拓展了资产定价的视角，但由于理论复杂，课堂将以直观理解和简单案例为主，不深陷数学细节。综合来看，Gu等人的论文适合复现其中实证部分技术路径，而Chen等的工作可作为拓展介绍前沿方向。
- **主要Python工具包与建模方法：**深度学习部分将使用 `TensorFlow/Keras` 或 `PyTorch` 构建和训练简单的神经网络模型。此外，`scikit-learn` 可用于较浅层的神经网络（如 `MLPRegressor`）以及数据预处理管道。学生将用到 `pandas` 处理大规模截面数据，例如多只股票的截面特征矩阵。可能涉及的建模方法包括：多层感知机（MLP）用于回归预测股票下月收益，简单的**LSTM**用于捕捉资产价格时间序列模式（若涉及时序依赖），以及模型正则化技术（如Dropout、早停等）防止过拟合。由于学生缺乏机器学习背景，本次实践中的神经网络模型将控制在基础水平（例如一到两层隐藏层的MLP），侧重于让学生了解其基本工作流程。

- **课上实践内容建议：**实践环节以**股票横截面收益预测**为场景。教师可提供事先整理好的小型数据集，例如标普500成分股的简化特征-收益率数据（可能包含每只股票的市值、市净率、动量收益率等特征，以及随时间滚动的未来1期超额收益率作为标签）。首先，引导学生使用 `scikit-learn` 对截面数据做基础处理，如标准化特征、拆分训练和测试集（注意时间顺序防止未来信息泄漏）。然后构建一个**简单神经网络模型**（使用Keras顺序模型或 `MLPRegressor`）对股票截面收益进行回归预测。演示如何设置输入层与隐藏层大小、激活函数（如ReLU）、优化算法（如Adam）、以及训练迭代过程。让学生观察训练集和测试集上的误差随迭代的变化，理解模型拟合程度。训练完成后，评估模型的截面预测能力，例如计算**截面  $SR^2$** 或者根据预测排序构建多空组合看其平均收益，以评估模型的经济价值。接着，可与传统线性基准比较：例如将相同特征集用线性回归或Fama-French回归预测，比较其解释力。若时间充裕，可进一步让学生调试网络结构或参数（比如增加隐藏单元或训练轮次）观察对性能的影响，从中体会模型复杂度与过拟合的权衡。整个实践突出**深度学习模型在处理高维非线性资产定价问题上的优势**，但也让学生认识到其训练难点和对数据规模要求。
- **课后作业设计方向：**深化对横截面机器学习资产定价的理解和应用。作业可以要求学生**复现或拓展论文实证**：例如利用提供的股票截面数据，尝试不同的机器学习方法预测股票回报（如随机森林 vs. 神经网络），比较它们的表现差异，并与线性模型做基准比较。也可让学生选取一个特定月份，利用训练好的模型预测该月所有股票的相对回报并构建一个多空投资组合，评估其收益表现（这模拟了利用机器学习选股的过程）。作业报告需包含：所用模型的说明、参数设置和调优过程、截面预测效果评价（如  $SR^2$ 、均方误差、信息比率等），以及对重要预测因子的分析（可借鉴Gu等论文中的做法，分析模型认为最重要的特征是否与已知因子如动量等一致）。通过作业，学生将加深对深度学习和复杂模型在资产定价中应用的认识，并积累将前沿研究思想转化为编程实践的经验。

**教学总结：**经过三次课的学习，学生将从实践中掌握基本的机器学习方法（线性回归、正则化、决策树/集成方法、神经网络）在资产定价中的应用技能。他们将理解机器学习如何利用**时间序列数据**提升对标普500指数等市场的收益预测，以及在**横截面数据**上发掘股票风险溢价的能力。课程通过经典论文引导和代码实作相结合的方式，培养学生将金融理论与数据科学工具融合的思维方式，为日后独立运用机器学习解决金融实际问题打下基础。

**参考文献：**课程各节内容均有经典研究作支撑，例如Gu et al. (2020)证明机器学习方法可将投资业绩提升一倍<sup>5</sup>；Chun et al. (2024)显示正则化回归在汇率和股指预测中的优势<sup>1</sup>；Rossi (2018)等工作则突出非线性模型捕捉复杂模式的价值<sup>2</sup>。这些权威来源和实证结果将贯穿课堂，确保课程内容的前沿性和可信度。学生通过阅读并复现这些研究的方法，将获得将理论应用于实践的宝贵经验。

---

<sup>1</sup> <sup>3</sup> <sup>4</sup> Forecasting returns with machine learning and optimizing global portfolios: evidence from the Korean and U.S. stock markets | Financial Innovation | Full Text  
<https://jfin-swufe.springeropen.com/articles/10.1186/s40854-024-00648-w>

<sup>2</sup> Machine\_Learning  
<https://mendoza.nd.edu/wp-content/uploads/2019/07/2018-Alberto-Rossi-Fall-Seminar-Paper-1-Stock-Market-Returns.pdf>

<sup>5</sup> <sup>6</sup> <sup>7</sup> Empirical Asset Pricing via Machine Learning by Shihao Gu, Bryan T. Kelly, Dacheng Xiu :: SSRN  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3159577](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3159577)