

基于机器学习的因果推断方法

目录

基于机器学习的因果推断方法

1. 课程概览
2. 主讲嘉宾简介
3. 课程概要
4. 报名和缴费信息
 - 缴费方式
5. 听课指南
 - 5.1 软件和课件
 - 5.2 实名制报名
6. 助教招聘
 - 说明和要求

1. 课程概览

- **时间：** 2023 年 4 月 15-16; 22 日
 - **讲授：** 9:00-12:00, 14:30-17:30; **答疑：** 17:30-18:00
- **方式：** 网络直播 + 30 天回放
- **软件：** Python
- **授课教师：** 司继春 (上海对外经济贸易大学)
- **报名链接：** <http://junquan18903405450.mikecrm.com/udInf5E>
- **课程详情：** <https://www.lianxh.cn>
- **回放安排**
 - 4 月 15-16 日课程回放：4月17 - 5月17日 (30天)
 - 4 月 22 日课程回放：4月23日-5月23日 (30天)
- **参考文献：** [点击下载](#)

2. 主讲嘉宾简介



司继春，上海对外经贸大学统计与信息学院讲师，主要研究领域为微观计量经济学、产业组织理论，成果见诸 Journal of Business and Economic Statistics、《中国人口科学》、《系统工程理论与实践》等期刊。司老师专长于机器学习，尤其是基于机器学习的因果推断前沿方法，有多个大型数据分析项目的实战经验。业余时间里，司老师也经常知乎上耐心作答，用通俗的语言普及统计和计量知识。他的知乎专栏名为「**慧航**」，关注者逾 31w，获赞超过 17w。他总能抽丝剥茧，把复杂的问题讲得清清楚楚。

3. 课程概要

因果推断要应对的一个棘手问题是「构造反事实」，其本质可以视为一个预测问题。机器学习为解决这类问题提供了诸多有效的工具(如惩罚回归、神经网络、随机森林等)，在解决遗漏变量、混淆因素、选择偏误方面凸显出传统方法无法比拟的优势。

本课程将重点讲解各种基于机器学习的因果推断方法，包括神经网络（循环神经网络和卷积神经网络）、决策树、因果树和因果森林、迁移学习和双重机器学习。

在内容安排上，重点讲解各个方法的应用场景、核心假设以及结果的分析方法和思路，以及各类方法的优劣和搭配使用。为了便于大家迁移学习，学有所得，课程中的核心方法都会借助案例论文的拆解来讲解。通过理解 Python 代码，大家将具备进一步修改和调试的能力。此外，在每个专题下，我们还精心挑选了 3-5 篇最新文献，以便大家能多个角度理解这些方法的应用场景，借他山之石攻己之玉。得益于 Python 机器学习方面的强大基础和快速发展，大家反而无需在写代码方面耗费过多的精力，更重要的是理解「它是什么」以及「如何应用」？

[点击下载所有参考文献](#)

第 1 讲 Python 基础 (3 小时)

本讲的目标是让您在短时间内初步了解 Python 的使用。与 R、Stata 相比，Python 最大的区别在于它是一门通用编程语言而非仅专注于统计分析，因此在编程中需要用计算机语言的思想进行操作。这也意味着，无论从语法层面还是编程技巧、思想和习惯方面，Python 都与 Stata、R 等都存在一定的差异。在这一讲中，我们将从最基础的数据结构开始，介绍 Python 的基础语法和读取文本文件、CSV 文件等操作。

- Python 数据类型和基本运算：数值、字符串、列表、元组、字典等
- Python 控制结构：条件、循环
- Python 中的函数和 Lambda 表达式
- Python 中的文件管理：文本文件、csv 文件、二进制文件等
- Python 语法补充：列表推断

第 2 讲 Python 中的数值计算 (3 小时)

Python 应用范围广泛，虽然不是专门的统计软件，但得益于 Numpy/Scipy/Pandas 等强大的工具，使其在数据分析、复杂运算方面的能力同样出色。本讲首先介绍 Python 的基础语法和框架，继而介绍如何在 Python 中进行科学计算和数据处理，以及 Python 的可视化工具。本讲是后续机器学习和文本分析的基础。

- Numpy 基础：向量与矩阵的操作
- Numpy 进阶：向量计算与切片
- Matplotlib 可视化：直方图、散点图、线图图形的绘制
- Pandas 基础：序列与数据框
- Pandas 进阶：切片、数据操作和处理

第 3 讲 机器学习初步 (3 小时)

随着大数据的发展，机器学习的应用已经拓展到了社会经济的各个领域。在学术界，机器学习方法也逐步应用到经济学、社会学等学科领域的研究中，为研究人员提供了更加高效、准确的分析手段。本讲将从机器学习的基础理论和概念入手，全面介绍机器学习的目标以及所面临的关键问题。我们将区分监督学习和无监督学习两种机器学习方法，并讲解它们的应用场景、优缺点和常见算法。应用层面将展示如何使用 scikit-learn 进行机器学习模型的训练，包括数据的预处理、模型的选择和调参等主题。

- 机器学习概述：预测问题与泛化能力
- 机器学习概述：模型的评价方法
- 机器学习概述：交叉验证和正则化
- 无监督学习：聚类和嵌入 (embedding)
- 监督学习：线性回归、Logistic 回归、决策树、随机森林
- 案例
 - Lasso 与回归控制法的应用
- 参考文献：
 - Aiken E L, Bedoya G, Blumenstock J E, et al. Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. **Journal of Development Economics**, 2023, 161: 103016. [-PDF-](#)
 - Cengiz D, Dube A, Lindner A, et al. Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes. **Journal of Labor Economics**, 2022, 40(S1): S203-S247. [-PDF-](#)
 - Gilchrist, Duncan Sheppard, and Emily Glassberg Sands. "Something to talk about: Social spillovers in movie consumption." **Journal of Political Economy**, 124.5 (2016): 1339-1382. [-PDF-](#)

- Goulet Coulombe P, Leroux M, Stevanovic D, et al. How is machine learning useful for macroeconomic forecasting?. **Journal of Applied Econometrics**, 2022, 37(5): 920-964. [-PDF-](#)
- Kaniel R, Lin Z, Pelger M, et al. Machine-learning the skill of mutual fund managers. **NBER working paper**, 2022. [-PDF-](#)
- Mullainathan S, Obermeyer Z. Diagnosing physician error: A machine learning approach to low-value health care. *The Quarterly Journal of Economics*, 2022, 137(2): 679-727. [-PDF-](#)
- Thorsrud L A. Words are the new numbers: A newsy coincident index of the business cycle. **Journal of Business & Economic Statistics**, 2020, 38(2): 393-409. [-PDF-](#)

第4讲 神经网络 (3 小时)

神经网络与深度学习是机器学习的重要发展，近些年火爆的人工智能技术几乎都离不开神经网络。神经网络在处理非结构化、非线性数据，比如图像、文本等数据中有着不可比拟的优势。

本讲以 **PyTorch** 为基础，详解神经网络的基本概念、构造方法，并在此基础上介绍深度学习，如循环神经网络、卷积神经网络等的概念和使用方法，进而介绍「迁移学习」及其应用场景。

- 神经网络初步：网络结构、损失函数、激活函数
- 神经网络中的最优化方法：随机梯度下降
- 循环神经网络简介
- 卷积神经网络简介
- 迁移学习与 BERT
- 案例
 - Khachiyani A, Thomas A, Zhou H, et al. Using Neural Networks to Predict Microspatial Economic Growth. **American Economic Review: Insights**, 2022, 4(4): 491-506. (略难) [-PDF-](#)
- 参考文献：
 - Pollmann M. Causal inference for spatial treatments. **arXiv preprint arXiv:2011.00373**, 2022. [-PDF-](#)

第5讲 因果推断与机器学习：原理及基于树的方法 (3 小时)

机器学习的迅速发展也深刻影响了经济学的发展。与机器学习不同，经济学更加注重解释和因果，因而传统的机器学习方法与计量经济学的结合需要更多的理论工作。

本节将介绍计量经济学与机器学习融合的最新进展，特别是在因果推断中融合机器学习的主要方法。我们首先将简单回顾处理效应的基本知识，并介绍在无混淆分配假设下，基于树的因果推断方法（**causal forest**）。Python 中 **EconML** 包包含了这些方法及其结合，我们将介绍这些方法在 **EconML** 中的使用。

- 处理效应与无混淆分配假设
- 异质性处理效应的背景
- 因果树和因果森林
- 因果森林与匹配的比较
- 案例
 - Deryugina T, Heutel G, Miller N H, et al. The mortality and medical costs of air pollution: Evidence from changes in wind direction. **American Economic Review**, 2019, 109(12): 4178-4219. [-PDF-](#)
- 参考文献：

- Athey, Susan, and Stefan Wager. "Estimating treatment effects with causal forests: An application." **Observational Studies** 5.2 (2019): 37-51. [-PDF-](#)
- Buhl-Wiggers J, Kerwin J T, Muñoz-Morales J, et al. Some children left behind: Variation in the effects of an educational intervention. **Journal of Econometrics**, 2022. [-PDF-](#)
- Cockx B, Lechner M, Bollens J. Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium. **Labour Economics**, 2023, 80: 102306. [-PDF-](#)
- Knaus M C, Lechner M, Strittmatter A. Heterogeneous Employment Effects of Job Search Programs A Machine Learning Approach. **Journal of Human Resources**, 2022, 57(2): 597-636. [-PDF-](#)
- Knaus M C, Lechner M, Strittmatter A. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. **The Econometrics Journal**, 2021, 24(1): 134-161. [-PDF-](#)

第6讲 经济学与机器学习：双重机器学习 (3小时)

双重机器学习能够很好的结合机器学习的预测、拟合能力并将其应用在计量经济学的参数估计和推断中，是一种能够灵活将机器学习与计量经济学进行融合的框架，现在也越来越多的受到理论和应用经济学的关注。

本节将介绍双重机器学习的基本思想，并介绍在因果推断中如何使用双重机器学习对异质性处理效应进行推断。此外，我们还将比较因果森林、双重机器学习等方法在因果推断中的表现以及在文献中的应用。最后，结合 EconML 介绍双重机器学习的代码实现。具体内容包括：

- 双重机器学习介绍
- 基于机器学习因果推断方法的比较与应用
- 案例
 - Deryugina T, Heutel G, Miller N H, et al. The mortality and medical costs of air pollution: Evidence from changes in wind direction. **American Economic Review**, 2019, 109(12): 4178-4219. [-PDF-](#)
- 参考文献:
 - Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, J. Robins, 2018, Double/debiased machine learning for treatment and structural parameters, **The Econometrics Journal**, 21 (1): C1-C68. [-Link-](#), [-PDF1-](#), [-PDF2-](#), [Replication](#)
 - Knaus, Michael C. "A double machine learning approach to estimate the effects of musical practice on student's skills." **Journal of the Royal Statistical Society: Series A** (Statistics in Society) 184.1 (2021): 282-300. [-PDF-](#)
 - Knaus M C, Lechner M, Strittmatter A. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. **The Econometrics Journal**, 2021, 24(1): 134-161. [-PDF-](#)
 - Knaus M C. Double machine learning-based programme evaluation under unconfoundedness. **The Econometrics Journal**, 2022, 25(3): 602-627. [-PDF-](#)

4. 报名和缴费信息

- **主办方：**太原君泉教育咨询有限公司
- **标准费用：**2900 元/班/人
- **优惠方案：**(以下各项优惠不能叠加使用)
 - **专题课/现场班老学员报名：**9 折, 2610 元/人
 - **学生(需提供学生证/卡照片)：**9 折, 2610 元/人
 - **会员报名：**85 折, 2465 元/人
- **联系方式：**
 - 邮箱: wjx004@sina.com
 - 王老师: 18903405450 (微信同号); 李老师: 18636102467 (微信同号)

报名链接: <http://junquan18903405450.mikecrm.com/udlnf5E>

长按/扫描二维码报名:



缴费方式

方式 1: 对公转账

- 户名: 太原君泉教育咨询有限公司
- 账号: 35117530000023891 (山西省太原市晋商银行南中环支行)
- **温馨提示:** 对公转账时, 请务必提供「**汇款人姓名-单位**」信息, 以便确认。

方式 2: 扫码支付



温馨提示： 扫码支付后，请将「付款记录」截屏发给王老师：18903405450 (微信同号)

5. 听课指南

5.1 软件和课件

听课软件： 本次课程可以在手机，ipad，平板以及 windows/Mac 系统的电脑上听课 (**台式机除外**)。

特别提示：

- 为保护讲师的知识产权和您的账户安全，系统会自动在您观看的视频中嵌入您的「用户名」信息。
- 一个账号绑定一个设备，且听课电脑不能外接显示屏，请大家提前准备好自己的听课设备。
- 本课程为虚拟产品，**一经报名，不得退换**。
- 为保护知识产权，课程不允许以任何形式录屏及传播。

5.2 实名制报名

本次课程实行实名参与，具体要求如下：

- 高校老师/同学报名时需向连享会课程负责人 **提供真实姓名，并附教师证/学生证图片**；
- 研究所及其他单位报名需提供 **能够证明姓名以及工作单位的证明**；
- 报名即默认同意「**连享会版权保护协议条款**」。

6. 助教招聘

说明和要求

- **名额:** 10 名
- **任务:**
 - **A. 课前准备:** 协助完成 3 篇介绍 Python 和计量经济学基础知识的文档;
 - **B. 开课前答疑:** 协助学员安装课件和软件, 在微信群中回答一些常见问题;
 - **C. 上课期间答疑:** 针对前一天学习的内容, 在微信群中答疑 (8:00-9:00, 19:00-22:00);
 - **Note:** 下午 5:30-6:00 的课后答疑由主讲教师负责。
- **要求:** 热心、尽职, 熟悉 Stata 或 Python 的基本语法和常用命令, 能对常见问题进行解答和记录。优先考虑熟悉 Python 的申请人。
- **特别说明:** 往期按期完成任务的助教可以直接联系连老师直录。
- **截止时间:** 2023 年 4 月 1 日 (将于 4 月 3 日公布遴选结果于连享会主页 lianxh.cn)

申请链接: < <https://www.wjx.top/vm/eGy4OYA.aspx#> >

扫码填写助教申请资料:

