

因果/格差/比較分析のための機械学習入門

日時: 2026 年 9 月 10 日 (木) 10:30~17:00

場所: オンライン開催 (詳細は別途ご案内)

料金: 一般 5,000 円、学生 2,500 円

講師: 川田恵介 (東京大学)

定員: 35 名 ※変更の可能性あり

■本コースの内容

本コースでは、これまで統計的因果推論や機械学習等に触れたことがない学生・研究者を念頭に、数式などを用いない、直感的な議論と R による実装に注力した議論を行います。

機械学習は幅広い学術分野及び実務組織 (政府・民間企業等)からの関心を集めてきました。特に因果効果や格差など、社会学における伝統的な研究目標への応用方法が確立されたことは、機械学習の急速な普及の原動力となっています。さらには大規模な外部データから推定された Foundation model (Gemini/ChatGPT 等の大規模言語モデルなど)を、学術研究に応用する方法についても、近年多くの優れたアイデアが提案されています。

本講義では、このような方法論的発展を取り入れた分析法について、基礎的な概念から R による実装方法までを紹介します。具体的な内容としては、(1) 回帰木系アルゴリズム (Random Forest) による線形モデル (OLS など) の補完、(2) 平均効果や格差などの推定への応用 (Double/debiased machine learning)、(3) 効果や格差の異質性分析 (BLP 推定への応用)を紹介します。また良性的過剰適合と言った近年注目されている性質や Foundation model の実践的な応用方法も紹介します。

講義内容のイメージとしては、以下の資料を参照ください。ただし実際の講義内容とは異なります。

講義レポジトリ: <https://github.com/tetokawata/NoteBalance>

教科書のサイト: <https://causalml-book.org/>

■次のような方におすすめです

- ・因果効果/格差の推定や機械学習の応用に関心のある方
- ・実装方法も同時に学ぶことで、ご自身の研究に迅速に活用されたい方
- ・根拠に基づいた定量的分析を行いたいが、数理的な議論に不慣れな方

■注意事項

- ・どなたでも参加可能です。
- ・ただし、SSJ データアーカイブのデータを利用した講義の場合、利用したデータを 9 月中に削除して頂く必要がございます。
- ・大学または公的研究機関所属の研究者・学生 (学部生も可)、SSJ データアーカイブへデータを寄託されている民間研究機関の方は、その後、研究目的で SSJ データアーカイブより申請して頂くことで利用可能です。
- ・R・Positron・必要パッケージのインストール・Project folder の作成・Project folder へのデータの格納を事前に済ませてください。登録・インストール方法を紹介した動画を事前に配布します。

■本コースの日程

- ・データ主導の定式化: 回帰木
 - ・内挿による推定精度の改善: Random Forest
 - ・平均効果/格差の推定: Double/debiased machine learning
 - ・効果の異質性の推定: Best linear projections
 - ・時間があれば、Causal Forest/Foundation model による予測の活用方法の紹介
- *進度によって内容が若干変わることがあります。