

## 機械学習による社会分析: 教師付き学習と因果推論への応用

日時: 2023年3月10日(水) 10:30~17:00

場所: オンライン開催

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

### ■本コースの内容

本コースでは、これまで機械学習の手法に触れたことがない学生・研究者を念頭に、サーベイデータの代表的な分析方法を紹介します。このため数式などは極力用いず、直感的なコンセプトの紹介に特化します。

機械学習への関心は学際的に高まっており、実務者からも大きな期待が寄せられています。

しかしながら社会科学における伝統的な分析手法とは、異なるルーツを有しており、実証研究への応用のハードルが高いと考えられがちです。しかしながら、伝統的な計量・統計的手法との融合が進む中で、実際には多くのコンセプトを共有しており、これまでの分析を補完する便利な\*\*工具箱\*\*としての活用法が多く提案されています。特に既存の多くの分析が、\*\*目をつぶってきた\*\*、前提とする統計モデルに推定結果が強く依存してしまう危険性を軽減することができます。

具体的には、(1) 教師付き学習 (ランダムフォレストや LASSO など) を用いて、条件付き平均値関数を“学習”する方法、(2) 社会集団を特徴づけるパラメータ (格差や因果効果など) 推論への応用、を紹介します。(1)は教師付き学習のそもそも関心である“予測モデル”構築にそのまま応用できます。対して社会科学や実務への応用においては、研究関心となる社会集団におけるパラメータの“本当”の値についての含意を明確にする必要があります。具体的には信頼区間などの推定誤差を、少ない仮定のもとで導出することが要求されます。しかしながら教師付き学習の手法のみでは、この要求に応えることができません。(2)では近年急速に進んだ機械学習と伝統的な計量手法の融合の成果を紹介し、信頼区間の近似計算が可能な手法を学びます。さらに実際のデータに対して、Python を用いて実装(sklearn, econml パッケージ等を使用)する方法も解説します。こちらも今まで Python に触れたことがない受講生を、想定します。

機械学習・Python とともに、学び始めるハードルが高い、と過剰に考えられがちです。

また学際的に発展した手法の宿命として、手法を学んだものの混乱している研究者もしばしば見かけます。しかしながらしっかりと内容を絞って順序立てて習得すれば、そのハードルは他の手法や言語と比べて決して難しいものではありません。

幅広い学生・研究者の参加をお待ちしております!!!

### ■次のような方におすすめです

- 初めて機械学習を学ぶ方
- 機械学習の格差研究・因果推論への応用に興味がある方

### ■注意事項

- Python のクラウド環境である Google colab を用いた、実習を行います。このため事前に Google ID の作成、Colab の動作確認を行なってください。動作確認用の動画を事前に公開いたします。

### ■本コースの日程

「予測モデル・社会集団を特徴づけるパラメータ・条件付き平均値・過剰適合の整理」、「教師付き学習：回帰木システムと線形モデルシステム」、「平均差・条件付き平均差の推定」、「条件付き平均差のノンパラメトリック推定」

\* 進捗によって、内容が若干変わることがあります。