

Rで学ぶ機械学習入門：予測から因果推論まで

日時：2021年3月5日（金）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：40名

■本コースの内容

「機械学習の社会科学への応用」について基礎から学ぶコースです。機械学習は、データを“柔軟”に用いることで、性能の良い予測モデルを構築することを可能にします。すでに標準的な toolkit の一つとして学部/大学院レベルの計量経済学の教科書でも紹介されるなど、社会科学においても急速に活用が拡大しております。

本コースでは、機械学習の基礎から社会科学の伝統的な関心である因果推論/比較研究への応用まで、高度な数学を用いずに紹介します。現状でデータを扱っている研究者・分析者のみならず、これからデータ分析をやりたい院生・学生の参加も歓迎いたします。

具体的内容としては、線形予測モデルを用いる LASSO・Ridge 推定および予測木モデルを用いる Bagging・Random Forest・Boosting について紹介いたします。さらに機械学習の、(1) 平均効果推定、(2) 条件付き平均効果に基づく効果の異質性探索、への応用法について紹介します。これらの手法はとくに多数のコントロール変数が活用可能な状況において比較優位性を持ちます。

さらに無料の統計ソフトである R と SSJ データアーカイブに寄託されている個票データを用いて、実際に自分の手で機械学習を用いた実証研究を体験してもらいます。こちらについても基礎的な内容から始めますが、R や因果推論に初めて触れられる方は、3月4日に開催される「Rで学ぶ統計的因果推論入門」コースも併せて受講されることをお勧めします。

■次のような方におすすめです

- ・機械学習、とくに社会科学研究への応用に関心がある方
- ・「機械学習/AIを用いたデータの解析結果」を批判的に検討したい方
- ・新しい統計ソフトの習熟に忌避感の無い方

■注意事項

- ・SSJ データアーカイブのデータを利用するため、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のみを対象とします。
- ・R および Rstudio がインストールされたパソコン等をご準備ください。

■本コースの日程

「過剰適合問題」、「データ分割」、「回帰系アルゴリズム」、「決定/回帰木系アルゴリズム」、「コントロール変数法への応用」、「効果の異質性分析への応用」

* 進度によって、内容が若干変わることがあります。