

理学におけるデータ科学実践 機械学習で自然科学を読み解けるか

中野直人

国際高等教育院附属データ科学イノベーション教育研

究センター 特定講師

理学研究科 連携講師

理学におけるデータ科学実践：機械学習で自然科学を読み解けるか

参加教員： 中野直人
余田成男（地球惑星科学専攻）

．．．

実験や観測からデータを取得してそれを解析することは、自然科学のいずれの分野においても普遍的に必要とされることである。昨今は機械学習などのデータ科学的手法が発達し、計算機能力の向上に伴い、様々な目的でデータが潜在的に持つ情報の抽出が試みられている。しかし、実際の理学研究においてもこれらの手法はどこまで有効なのであろうか。本SGでは、データ科学的手法を簡単なものから習得して、データ科学を身近なものにすることを目的とする。さらに、既存手法との比較による再確認、既存手法では得られない新たな理解、ここで学ぶ新しい手法ではできないこと、手法の数学的構造、などを整理することで、各手法の本質の理解を目指す。

具体的なSGの進め方としては、まず各手法の大まかなサーベイを行ってデータ科学を俯瞰した後、サンプルデータで手法に慣れるようにする。その後、数名ごとの班に別れて機械学習的手法を実践する。各自が実習課題を持つこととし、月に1・2回のペースで結果を持ち寄り全体で議論をおこなう。数学的構造について検討し、手法の改善や対象範囲を理論的に整理する班も設置可能とする。解析に用いるデータは、各自の研究に関わるデータを持ち寄るか、興味のある分野のデータの提供を受けるなどする。本SGでは機械学習の手法に主眼をおくため、RやPythonを用いて実習を行うが、初めのサーベイ時にチュートリアルを行うなど、計算環境の準備も行う。専門分野毎に得意とする理論や手法があるため、本SGを介して専攻横断的に方法論の共有も行えると尚良い。

理学におけるデータ科学実践：機械学習で自然科学を読み解けるか

参加教員： 中野直人
余田成男（地球惑星科学専攻）

• • •

実験や観測からデータを取得してそれを解析することは、自然科学のいずれの分野においても普遍的に必要とされ、そのために様々な手法が開発されてきた。近年は機械学習などのデータ科学的な手法が飛躍的に計算機能力の向上に伴い、様々な分野の物理学研究に活用されるようになった。簡単なものから複雑なものまで、比較による再構築や、既存手法の数学的

**理学にまつわるデータで
機械学習してみないか？**

具体的なSGの進め方としては、まず各手法の大まかなサーベイを行ってデータ科学を俯瞰した後、サンプルデータで手法に慣れるようにする。その後、数名ごとの班に別れて機械学習的手法を実践する。各自が実習課題を持つこととし、月に1・2回のペースで結果を持ち寄り全体で議論をおこなう。数学的構造について検討し、手法の改善や対象範囲を理論的に整理する班も設置可能とする。解析に用いるデータは、各自の研究に関わるデータを持ち寄るか、興味のある分野のデータの提供を受けるなどとする。本SGでは機械学習の手法に主眼をおくため、RやPythonを用いて実習を行うが、初めのサーベイ時にチュートリアルを行うなど、計算環境の準備も行う。専門分野毎に得意とする理論や手法があるため、本SGを介して専攻横断的に方法論の共有も行えると尚良い。

やること

- * 興味あるデータを持ち寄ったり，提供してもらったりしてデータ解析する
- * 基本的にグループ作業
進捗状況や成果を全体発表で共有
- * ツールと手法のチュートリアルします
~ Python (Jupyter notebook)
~ Scikit-learn
とりあえず慣れる

やれることの例

- * 地磁気のデータ

理学研究科附属地磁気世界資料解析センターご提供

- * バイオインフォマティクス

Cancer Program Resource Gatewayから取れる

- * 気象庁55年長期再解析

気象庁JRAプロジェクトのページ

- * 手書き文字認識でとにかく遊ぶ

MNIST

- * モデルパラメタリゼーション

とにかくいろいろな回帰やニューラルネットワーク
などもろもろの手法をためす

実施要領

* PCもちこみしてください

なくても大丈夫！ 最悪スマホでもできる
Google の Colaboratory

なくても大丈夫！ SG「コンピュータでとことん遊ぶ」
と連携

* 曜日時限は参加者で相談

個人的には火曜日午後遅めがいいです

* ペースは月1・2回くらい

持ち帰って作業してもらいます

簡単な進捗報告と手法のチュートリアル

メッセージ

- * 専攻が異なると、違ったデータ解析手法を使うことがあります
 - 仲良くなって共有しよう
- * とりあえず慣れることが重要
 - 難しそうとかでおじけづかないで
 - 片手間でいいです 最悪ライブラリが全て解決
fit! predict!
- * 機械学習や計算機に関係するSGは多いのでうまく協力しよう

おまちしています