

SG09

理学におけるデータ科学 理論と実践

代表教員 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
理学研究科 連携講師

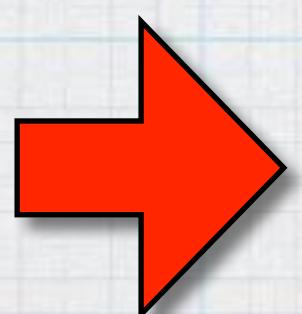
2021年4月23日

背景

機械学習的手法の研究や応用の例が増えてきている

- * 特徴量を抽出して分類や予測
- * 人の代わりにタスクを機械に実行させる

理学にもデータはいっぱい
研究にそういう手法は使えるのか？



自分で確かめてみるしかない

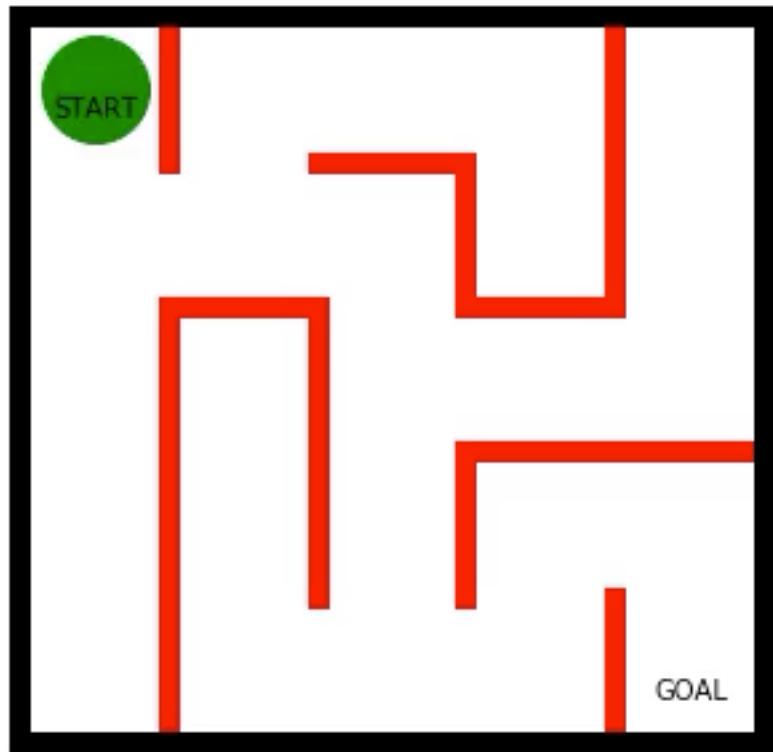
昨年度まででやったこと

- * 多層パーセプトロン・手法と理論
- * 決定木・ランダムフォレスト・手法と理論
- * AutoEncoder で次元縮約
- * Reservoir Computing で力学系の再現
- * 強化学習・方策勾配法で迷路を解く
- * 置み込みニューラルネットワークで画像解析
- * Computer Vision の方法で輪郭抽出

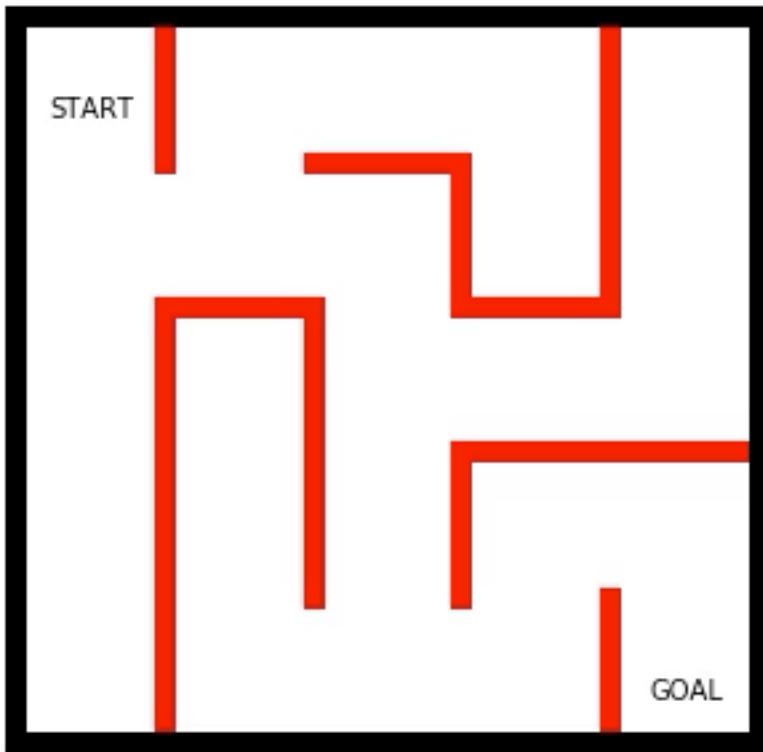
こんなことは簡単

「探索」と「学習」

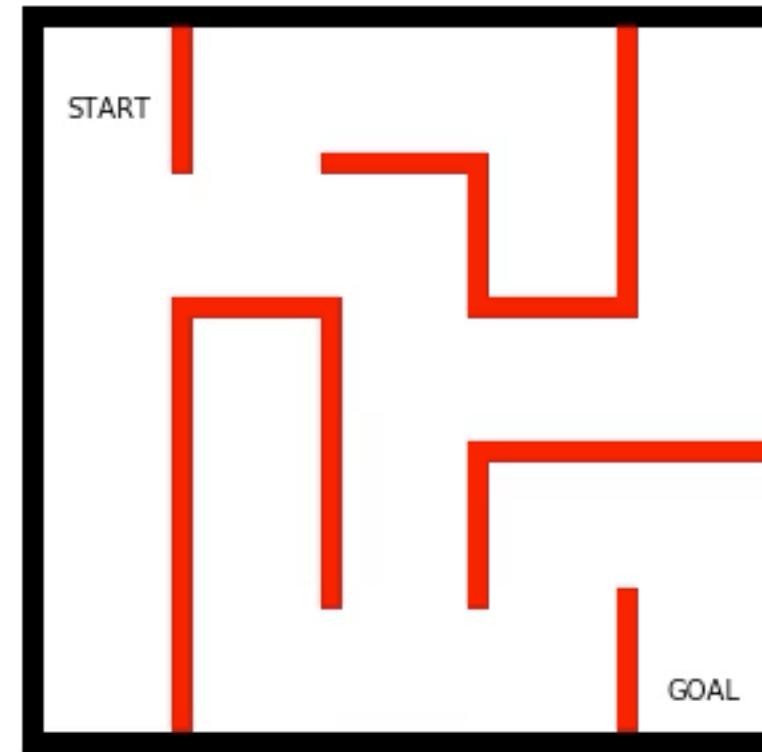
迷路を解くエージェント（強化学習）



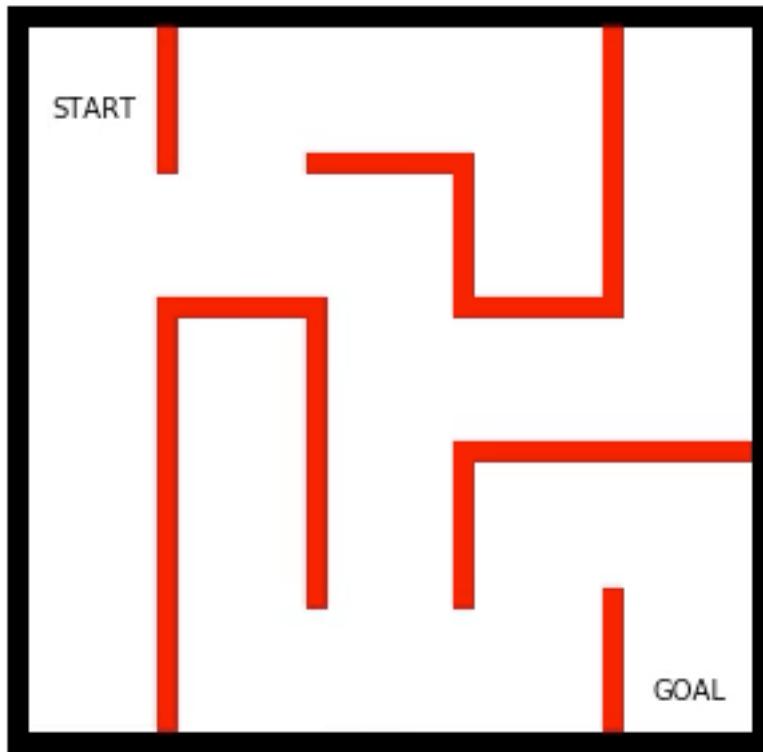
1回学習後



100回学習後



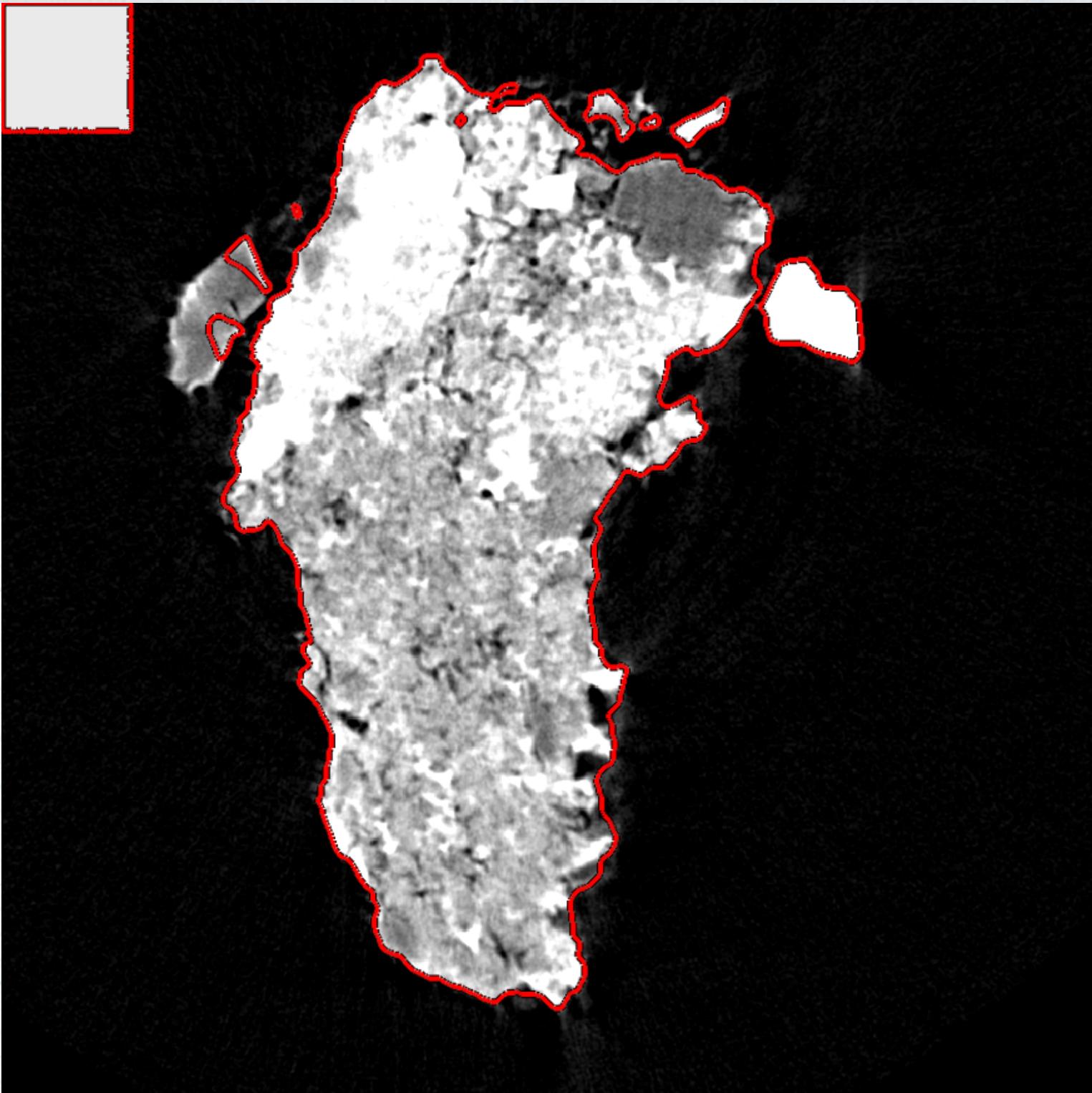
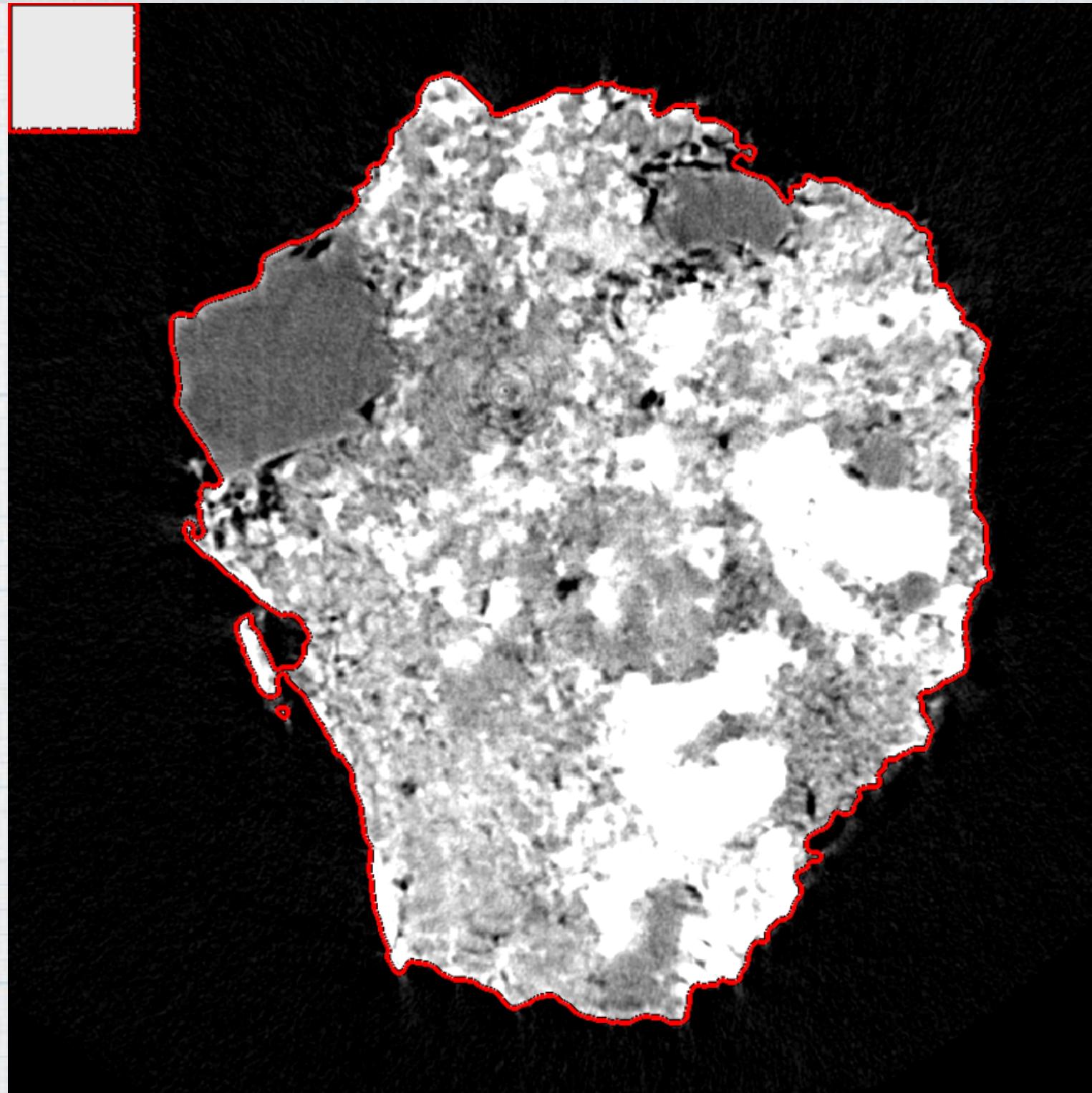
500回学習後



学習終了後
(1384回)

方策勾配法は必ず最短経路に収束する
(ベルマン方程式の不動点の一意性)

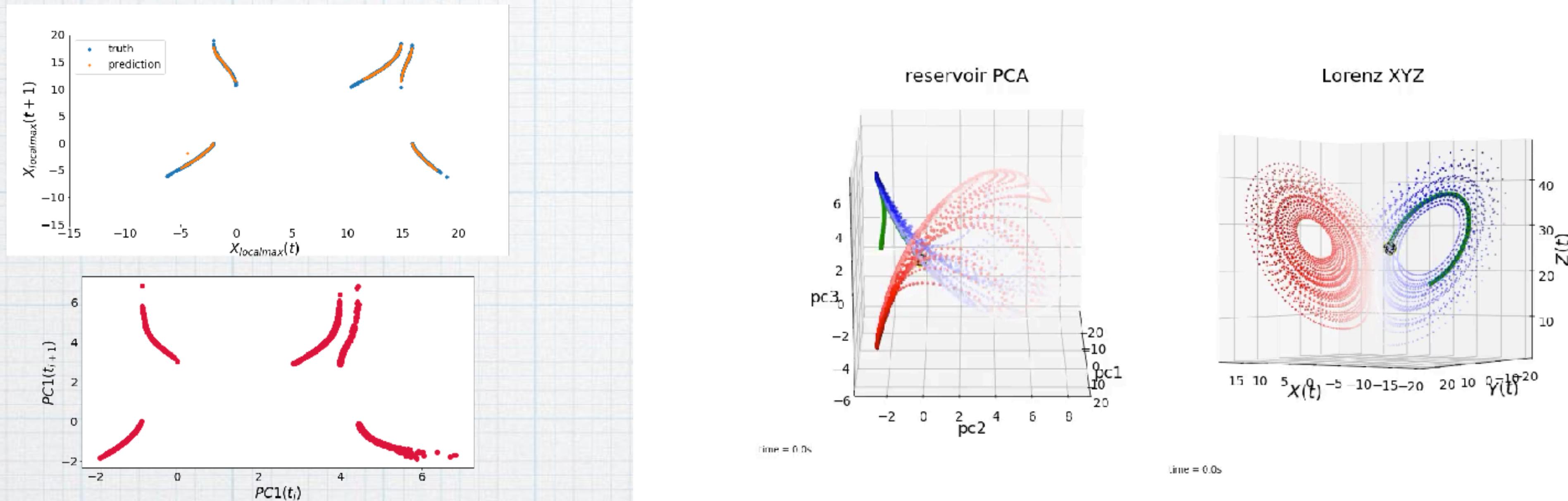
例：画像の輪郭抽出



土・山明先生（名誉教授・地質学鉱物学教室）のスライドより

こういう画像の位置合わせして中身のセグメンテーションして...
機械学習的手法よりは Computer Vision の手法の方がいいかも

例：Reservoir Computing



再帰的ニューラルネットワークの学習方法の1つ
学習できるときのノードの構造を調べた

感想

機械学習を試すこと自体は
すでにどの分野でも進められてきた

いや、クソ大変

機械学習手法を用いて（必要に応じて工夫して）
課題を解決することはまあ何とかなるでしょ

理論的に「機械学習がなぜうまくいくか」を
考えるのが楽しい

→ 今年度は機械学習のお教室はやらない

このSGでやりたいこと

なぜ対象の機械学習モデルは
学習ができるのかを知りたい

ニューラルネットワークには
universality があるからとか
つまらんことは言うなよ

このSGでやりたいこと

なぜ対象の機械学習モデルは
学習ができるのか

これは難しいので

対象の機械学習モデルが
どういう設定のときに学習ができるか

これもまだ
難しいので

うまく学習できるときに
特徴的な構造は何か？

これらから
じっくり考える

昨年度は

素粒子と凝縮系の博士課程の学生と
統計力学モデルのアナロジから
Reservoir Computing の学習性能に迫ってみた

平均場近似！

経路積分！

くりこみ群方程式！

今年度は

とにかく

「機械学習がなぜうまくいくか」を理論的に考える

数学や物理だけで済むか？ そうじゃないだろう？

具体的な機械学習の題材（例えば画像解析）でも
「いつうまくいか」をきちんと考へるには
ドメイン知識が必要じゃないか？

何らかの意味での理論が大事だろ？

実施要領

- * コロナであってもなくとも関係なく全部リモートでやります
- * なので定期的に集まることはしない
- * ツールを使う
 1. Slack で情報共有
 2. Miro (or Esa) で進捗確認
 3. Zoom も必要に応じて

だから積極的に活動できる人でないとムリ

特に代表教員は忙しいと放置するタイプ。

PC必須

理学におけるデータ科学
理論と実践

參加教員

代表教員 中野直人

(以下敬称略・順不同)

宮路智行 (数学・数理解析専攻, 力学系)

楠岡誠一郎 (数学・数理解析専攻, 確率論)

谷村吉隆 (化学専攻, 統計力学)