

Section 5

- 機械学習エンジニアへの道

機械学習技術者に求められるもの

- インフラ
 - データの取得、管理、計算資源の確保ができる
- 実践
 - 対象業務を深く理解し、収集するデータの質を高めることができる
 - どの程度の性能で、何ができるかを見極めることができる
- 理論
 - 性能勝負の段階で、いかに多くの引き出しを持っているか

機械学習技術者に求められるもの

- インフラ・実践・理論のすべてをカバーできる人は少ない
- 上記3つのいずれかに軸足を置き、他の2つは何をやっているかがわかる程度になっておく
- 技術は変化するので、継続的な勉強が必要

機械学習のインフラ

- データの取得
 - サンプルデータのような十分に整えられたデータが得られることはあまりない
- 音声・画像・言語
 - それぞれのツールの精通することが必要
- プライバシー保護が必要なデータ
 - 匿名化等の処理が必要

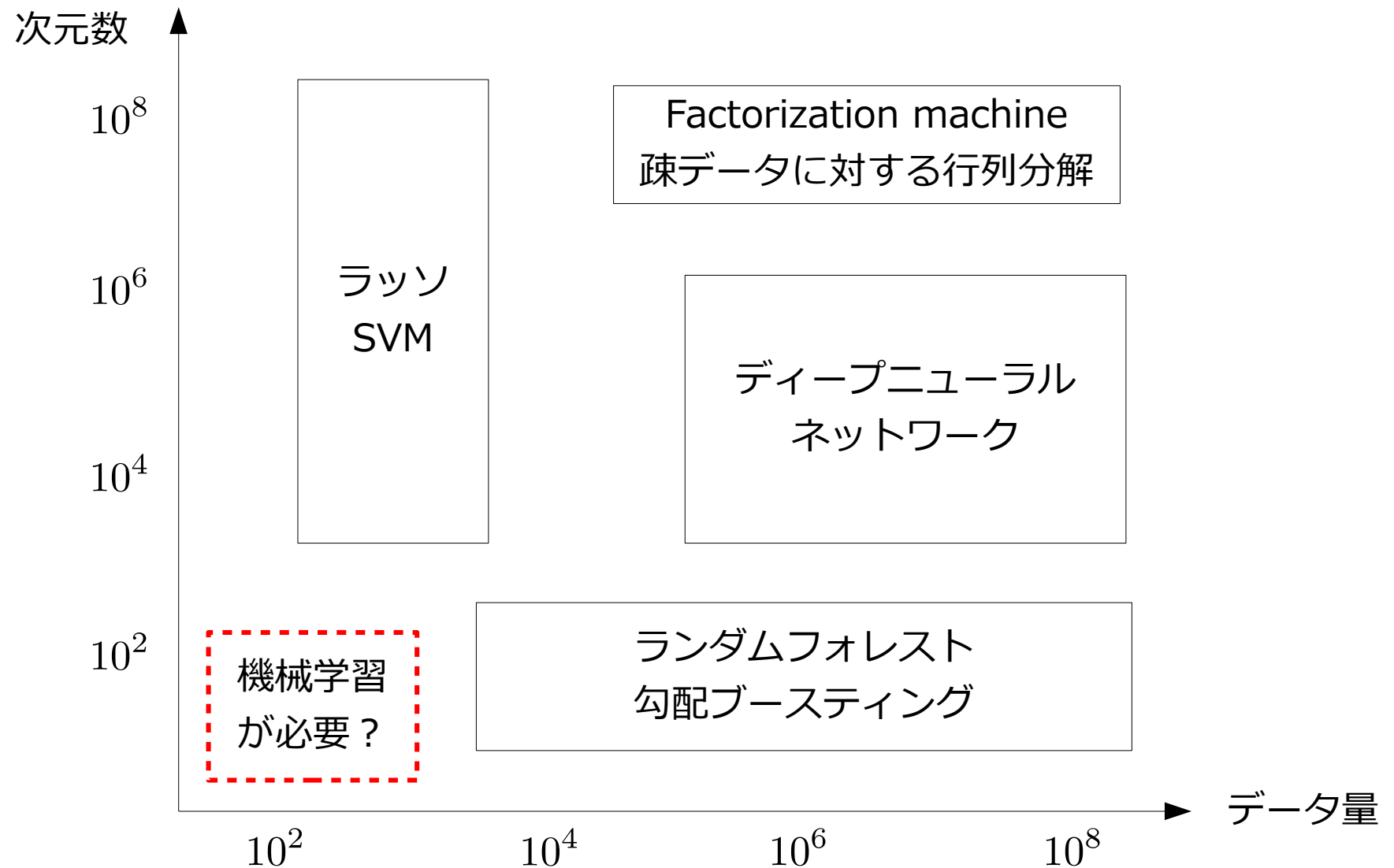


機械学習のインフラ

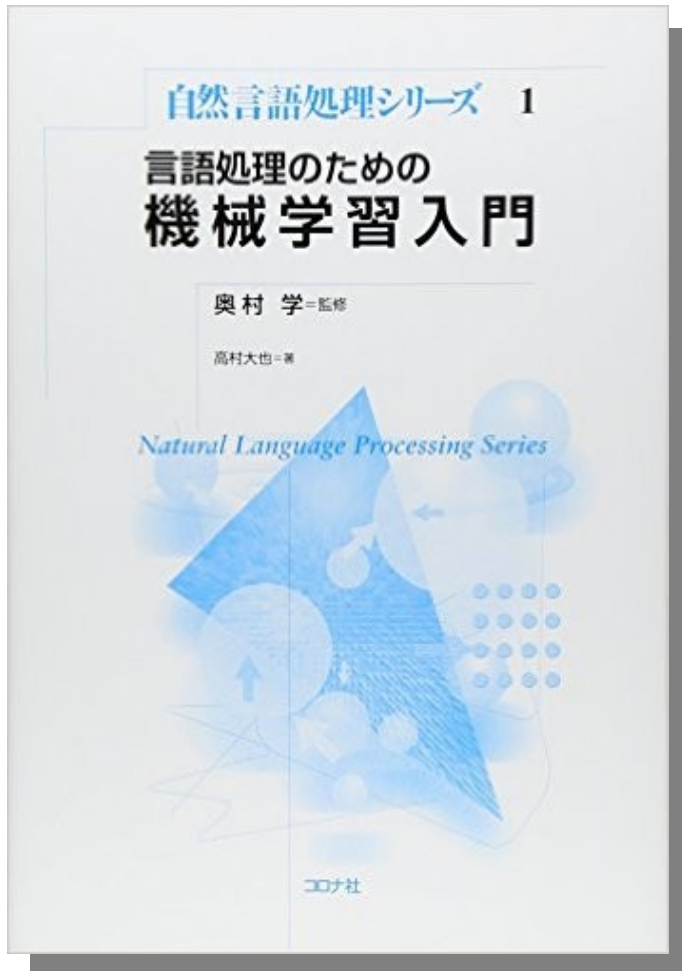
- データの管理
 - 関係者間でどのようにしてデータを共有するか
- 計算資源の確保
 - オンプレミス
 - GPU を用いた計算環境の構築
 - クラウド
 - 最新情報の共有が必要

機械学習の実践

- 問題を正確に定義し、適用する手法を見極める



推薦図書



1. 必要な数学的知識
2. 文書および単語の数学的表現
3. クラスタリング
4. 分類
5. 系列ラベリング
6. 実験の仕方など

高村著 コロナ社 , 2010

推薦図書



第 1 章 数学の準備

第 2 章 Python の準備

第 3 章 ニューラルネットワーク

第 4 章 ディープニューラルネットワーク

第 5 章 リカレントニューラルネットワーク

第 6 章 リカレントニューラルネットワーク
の応用

巢籠著 マイナビ出版, 2017

推薦コース

- オンライン学習 (udemy)



- Python で機械学習 : scikit-learn で学ぶ識別入門

- <https://www.udemy.com/python-scikit-learn/>

- 講師 : 玉木 徹

- 合計 9 時間のビデオレクチャー

機械学習の理論を学ぶ

- ベース
 - 統計学、最適化数学
- 英語
 - 良質な情報を得られる機会が増える

推薦図書



平井著 森北出版，2012

第 1 章 はじめに

第 2 章 識別規則と学習法の概要

第 3 章 ベイズの識別規則

第 4 章 確率モデルと識別関数

第 5 章 k 最近傍法（kNN 法）

第 6 章 線形識別関数

第 7 章 パーセプトロン型学習規則

第 8 章 サポートベクトルマシン

第 9 章 部分空間法

第 10 章 クラスタリング

第 11 章 識別器の組み合わせによる性能強化

推薦図書



杉山著 講談社，2013

第 I 部 はじめに

第 1 章 機械学習とは

第 2 章 学習モデル

第 II 部 教師付き回帰

第 3 章 最小二乗学習

第 4 章 制約付き最小二乗学習

第 5 章 スパース学習

第 6 章 ロバスト学習

第 III 部 教師付き分類

第 7 章 最小二乗学習に基づく分類

第 8 章 サポートベクトル分類

第 9 章 アンサンブル分類

第 10 章 確率的分類

第 11 章 系列データの分類

第 IV 部 教師なし学習

第 12 章 異常検出

第 13 章 教師なし次元削減

第 14 章 クラスタリング

第 V 部 発展的課題

第 15 章 オンライン学習

第 16 章 半教師付き学習

第 17 章 教師付き次元削減

第 18 章 転移学習

第 19 章 マルチタスク学習

第 VI 部 おわりに

第 20 章 まとめと今後の展望

Section5 のまとめ

- 機械学習エンジニアへの道
 - インフラ、実践、理論のいずれかに精通する
- インフラ
 - データの取得・利用に関する法整備の動向にも注意
- 実践
 - 問題に応じた手法の適用からチューニングまで
- 理論
 - 継続して勉強する環境を整える