

機械学習 ガイダンス

管理工学科
篠沢佳久

資料の内容

- 機械学習とは
- 講義内容
 - 講義計画, 講義の進め方

機械学習とは

講義の目的:「予測」する手法を学ぶ

最近の話題

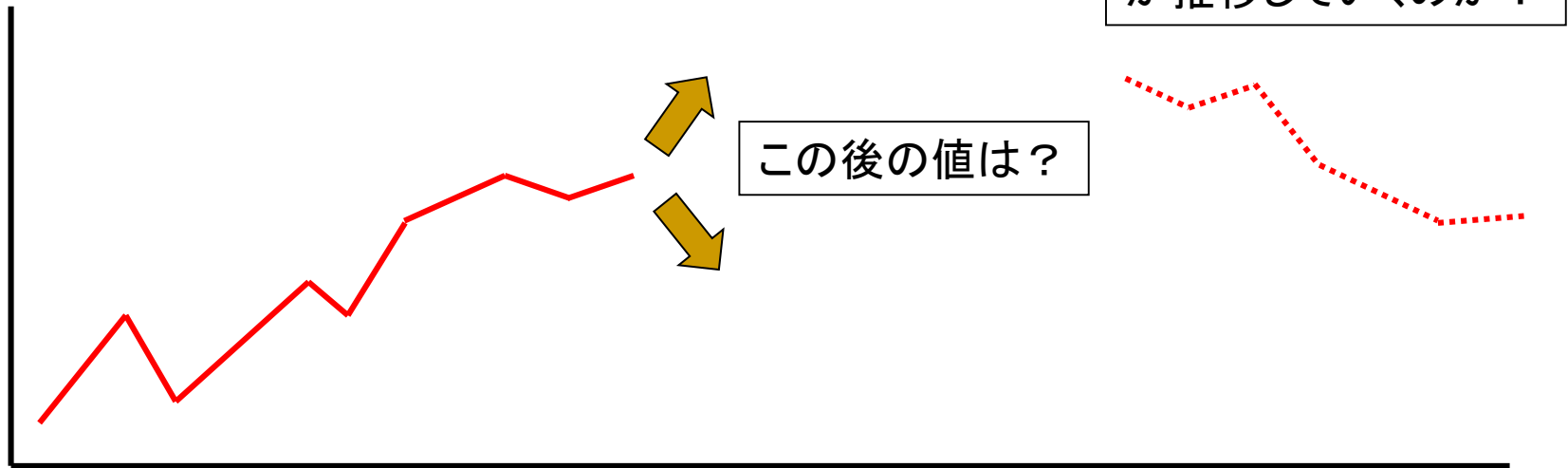
- リクナビ
 - 「内定辞退率」の予測データ
- セブン銀行
 - ATMを出金型, 入金型, バランス型に分類
 - ATMが稼働停止しないよう資金予測
- Google
 - 「深層学習」により肺がん画像を識別

講義の目的

- 「予測」する手法, アルゴリズムを学ぶ
- 何を「予測」するのか
 - 「未来」の値を予測したい
 - 「未知」のものを予測したい
 - 「似ている」ものがどれかを予測したい
- どのように「予測」するのか
 - 手法として機械学習を用いて予測
 - 道具としてscikit-learn(Python)を利用

何を「予測」するのか①

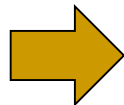
■ 未来の値を予測したい(回帰問題)



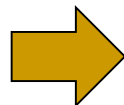
何を「予測」するのか②

- 「未知」のものを予測したい(分類問題)

これらは何でしょうか



りんご?



りんごジュース?

何を「予測」するのか③

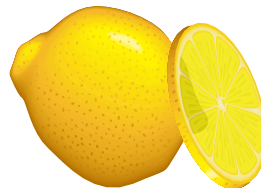
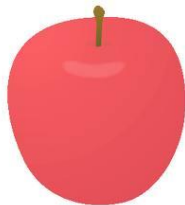
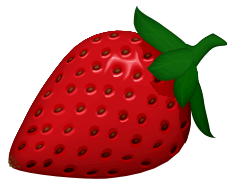
○は何でしょうか？

天は○の上に○を造らず, ○の下に○を造らずといへり

何を「予測」するのか④

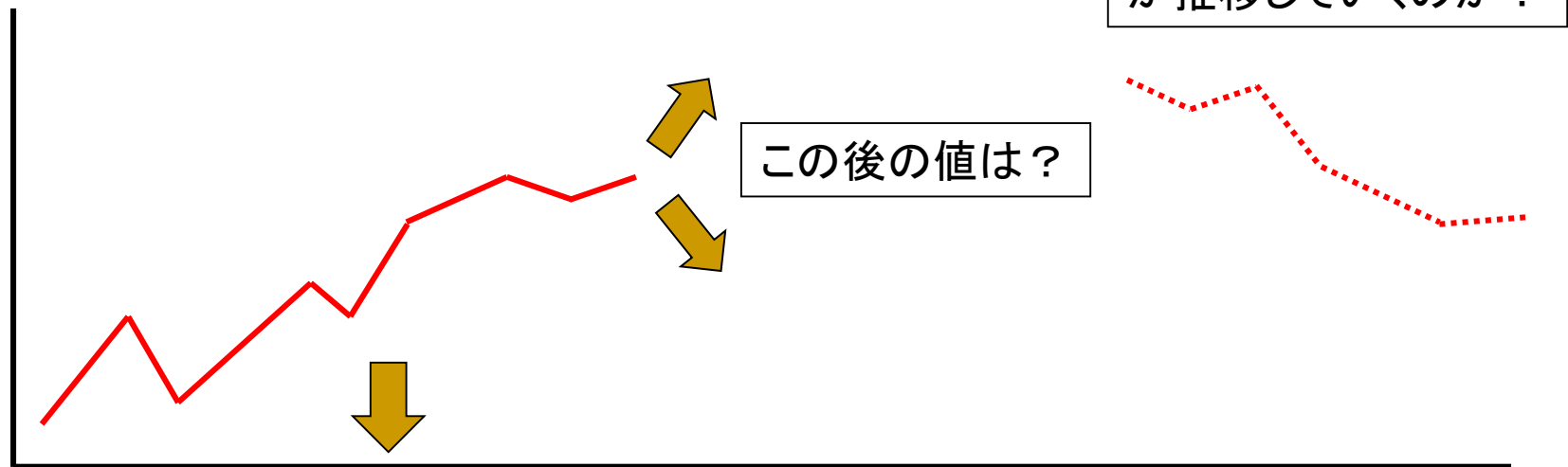
- 「似ている」ものがどれかを予測したい(クラスタリング)

似ているものはどれでしょう



どのように「予測」するのか①

■ 未来の値を予測したい

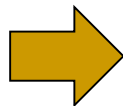


「過去のデータ」から「規則」を見つける

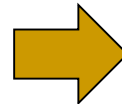
〇〇が起きた時, 値は上がる
□□が起きた時, 値は下がる

どのように「予測」するのか②

■ 「未知」のものを予測したい



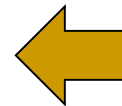
これは、
「赤い」「丸い」
ものだ



「赤い」「丸い」と言えば...
りんご？ 赤いきつね？



りんごの確率が高そう...



「りんご」と予測

どのように「予測」するのか③

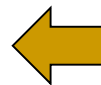
○は何でしょうか？

学問のすゝめ

天は○の上に○を造らず，
○の下に○を造らずといへり



知らない場合，正解は分からない
ただし，予測はできる



「学問のすゝめ」を知らないと
分からない

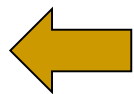


あらかじめいろいろなことを記憶
しておき，「尤もらしい語」で穴埋め

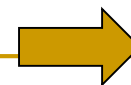
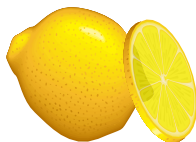
どのように「予測」するのか④

- 「似ている」ものがどれかを予測したい

大きさを基準(特徴)とした場合



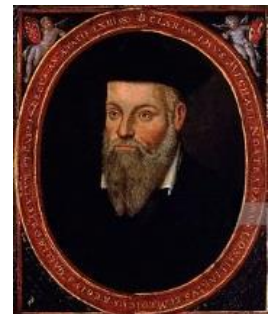
小さい



大きい

「予測」に必要なもの

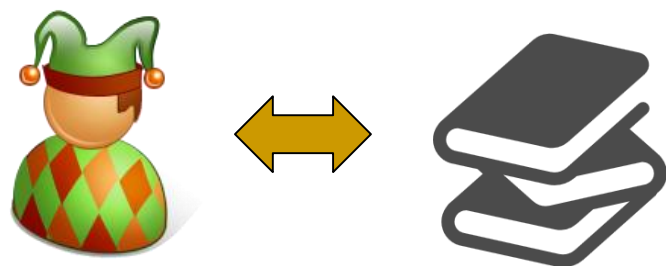
- 答えそのもの, 過去のデータ, 経験, 根拠, 規則, ...
- 規則を導出, 獲得する方法
- 「予言」との違い
 - ノストラダムス
 - 占星術
 - 地球と他の天体の関係から未来を予測*



*ただし, 科学的には正しくありません

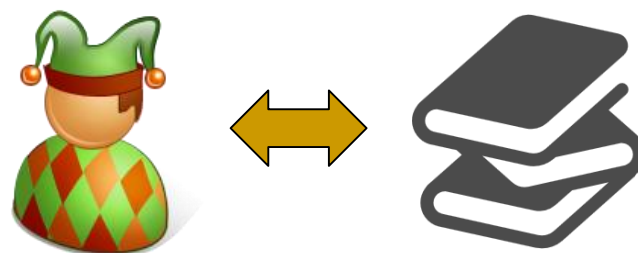
人間はどのように「予測」できるようになるのか？

■ 人間の場合



① 過去のデータから多くの規則を獲得

② 規則を獲得する手法を考案

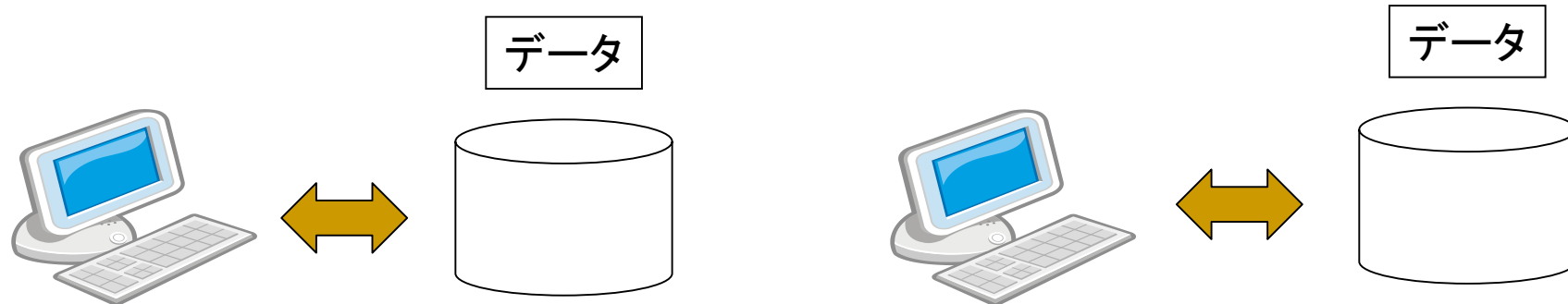


③ 過去のデータを記憶

学習

計算機に「予測」させるためには？

■ 計算機の場合



① 過去のデータから多くの規則を獲得
ただし、規則を獲得する方法は与える

② 過去のデータを記憶

大量データから規則を獲得し、予測を行なう手法、アルゴリズム

機械学習

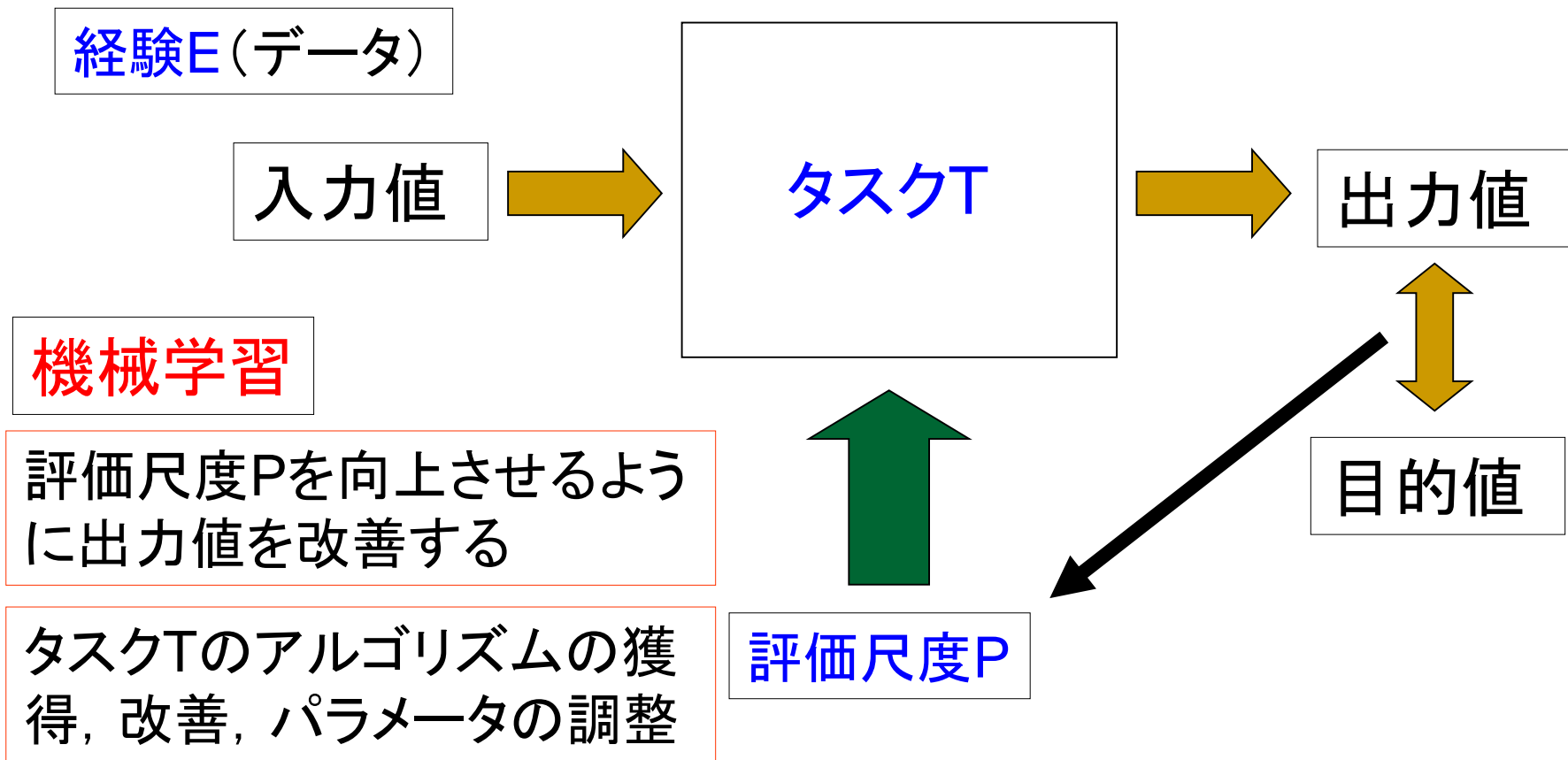
機械学習 (Machine Learning)

- Arthur Samuel (1959)
 - 明示的にプログラムされなくとも、自分で学習する能力
 - チェッカープログラム
 - Samuel Checkers-playing Program
 - デモ
 - <https://www.gamesforthebrain.com/japanese/checkers/>

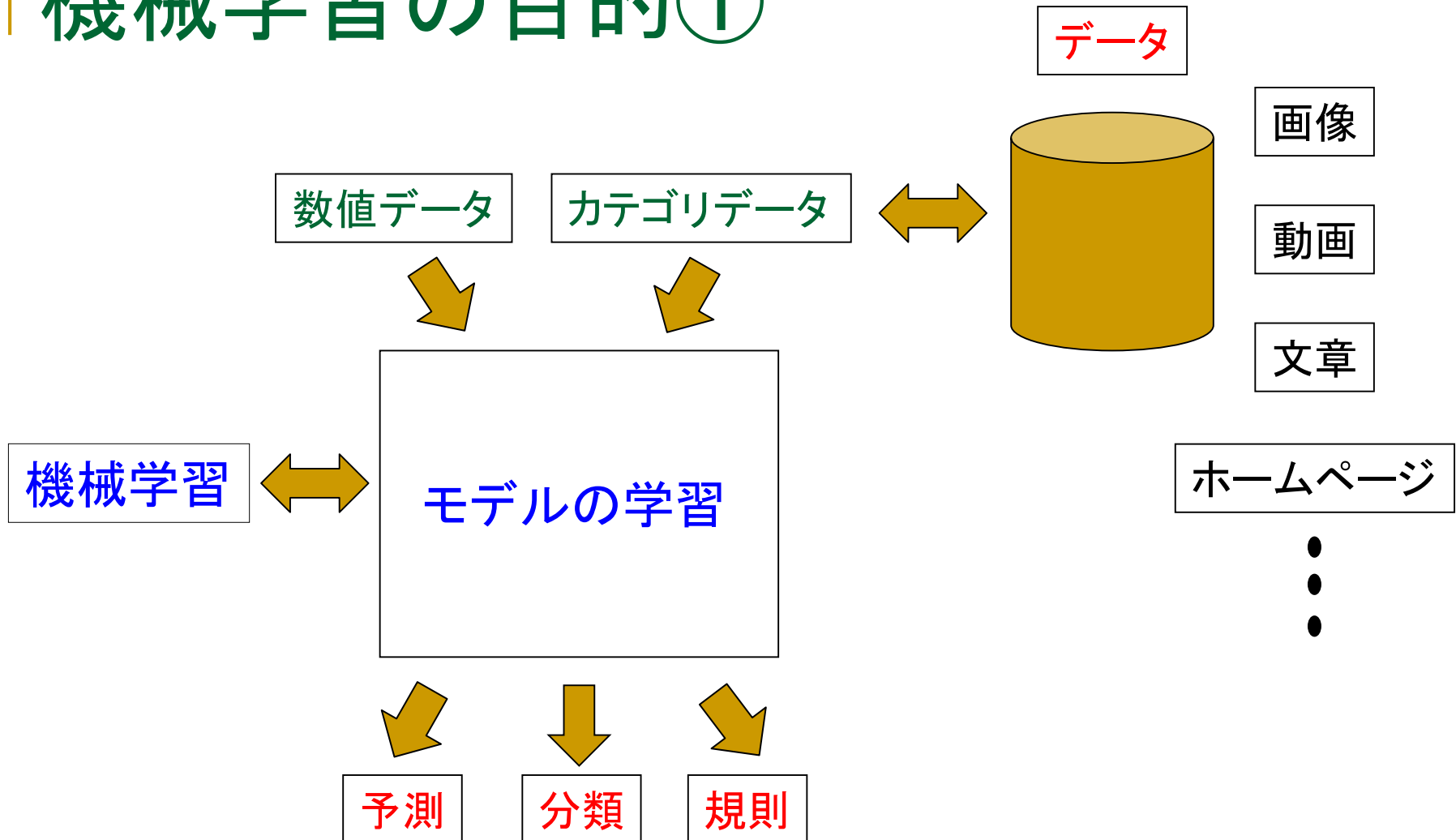
機械学習とは①

- Thomas Mitchellの定義(1997)
 - 構成要素
 - 経験E
 - タスクT
 - 評価尺度P
 - あるタスクTについて, 評価尺度Pで測られたタスクの実行能力が経験Eを通じて向上
 - 経験Eより学習

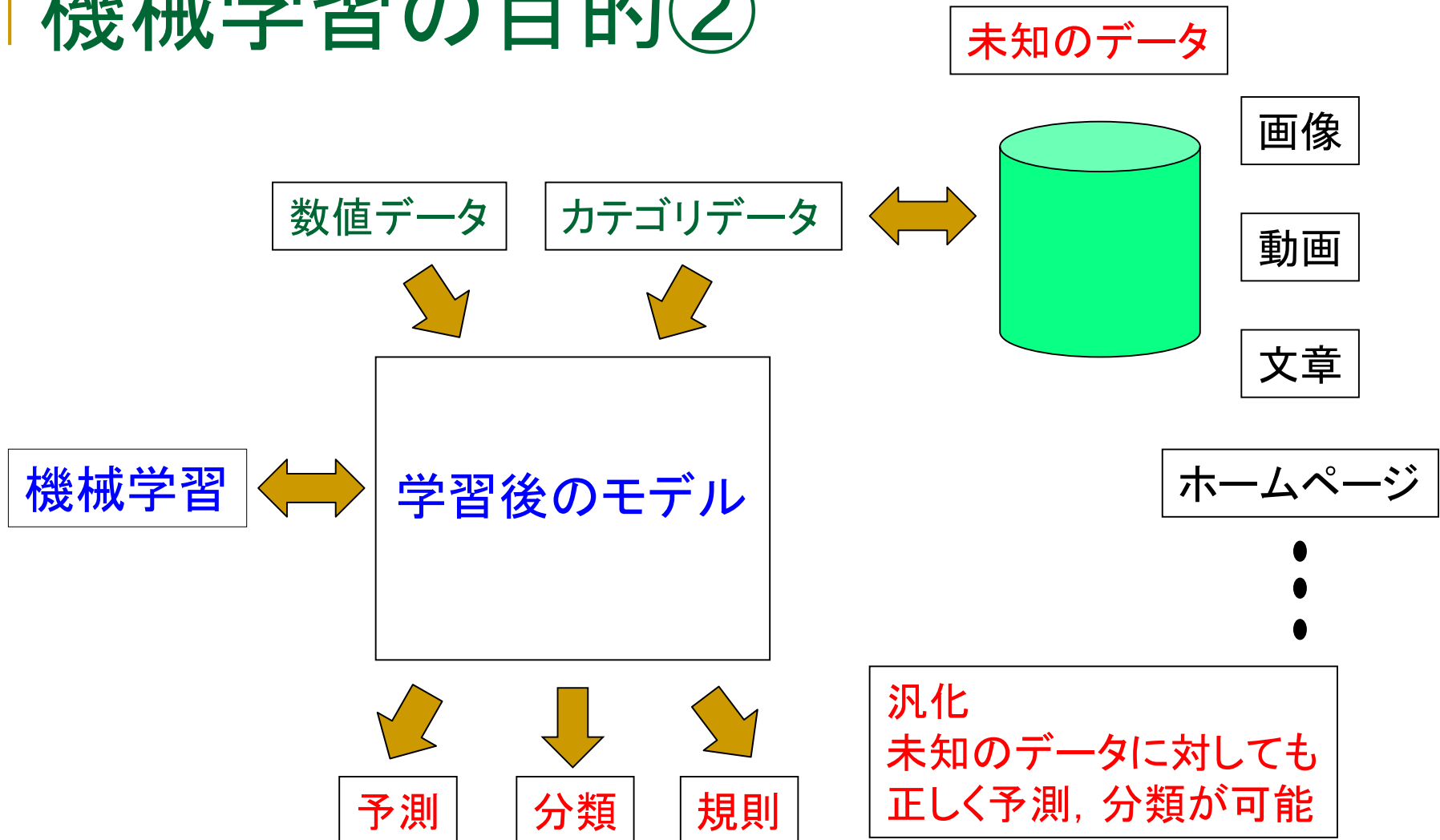
機械学習とは②



機械学習の目的①



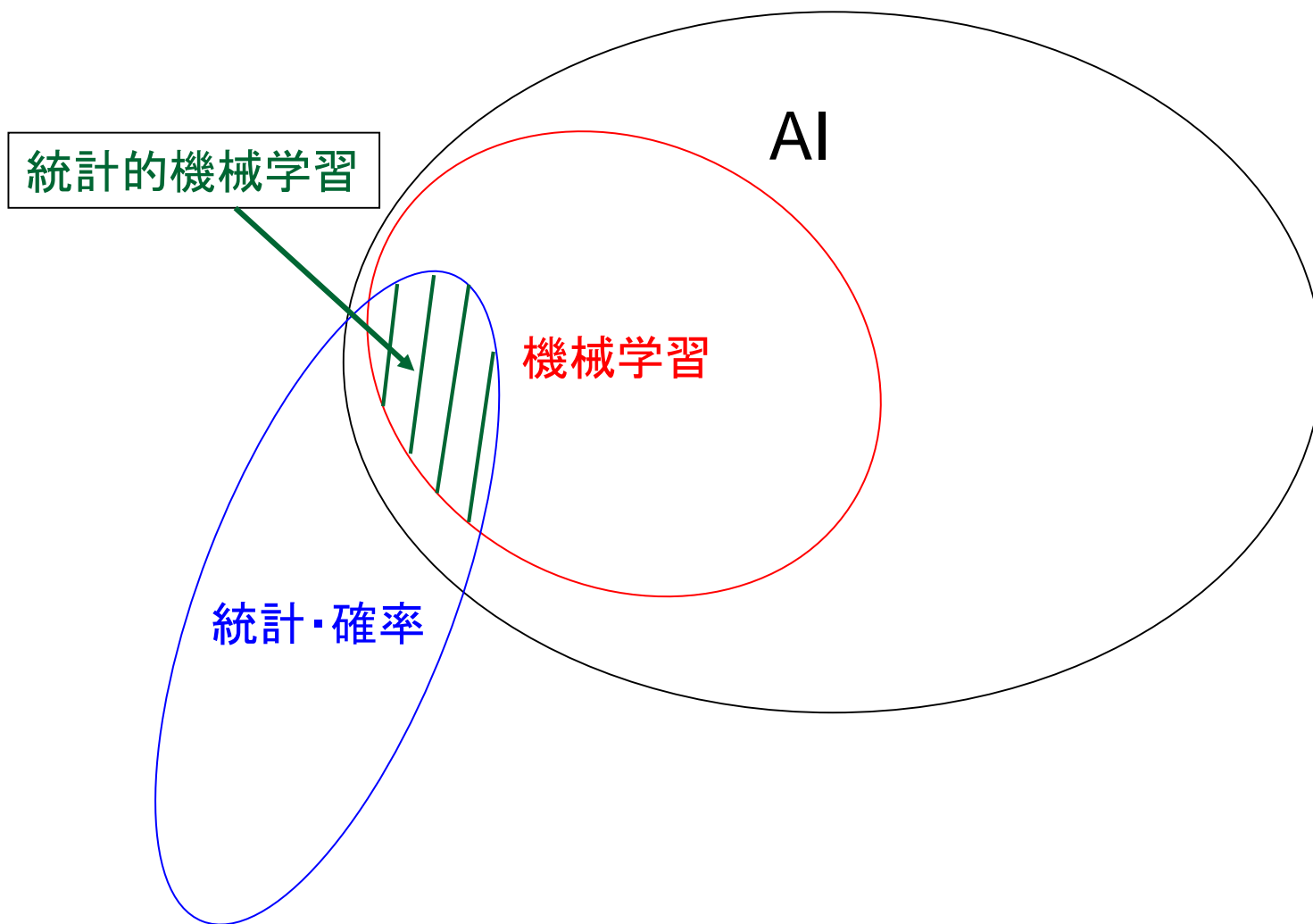
機械学習の目的②



機械学習の種類

- 教師あり学習 (supervised learning)
 - 出力すべき正解値 (正解ラベル) を提示しながら行なう学習方式
 - (例) 線形回帰
- 教師なし学習 (unsupervised Learning)
 - 出力すべき正解値を提示しない学習方式
 - (例) 主成分分析
- 半教師あり学習 (semi-supervised learning)
 - 正解値のついたデータが少ない場合, 未知データに正解値をつけながら行なう学習方式

人工知能 (Artificial Intelligence) ブーム



機械学習と人工知能

- John McCarthy
 - ダートマス会議(1956)
 - 第一次(?)AIブーム
- 計算機に人間の知能を持たせようとする試み
- 機械学習は「人工知能」の分野とも関連が深い
- 近年は第三次(?)AIブーム
 - 深層学習(deep learning)

機械学習と統計学

■ 統計学

- データの母集団の分布が分かれば、未知のデータも予測できる



- データの母集団の分布を知りたい
- ただし、利用できるデータは限られている



- 母集団の分布を推定し、予測

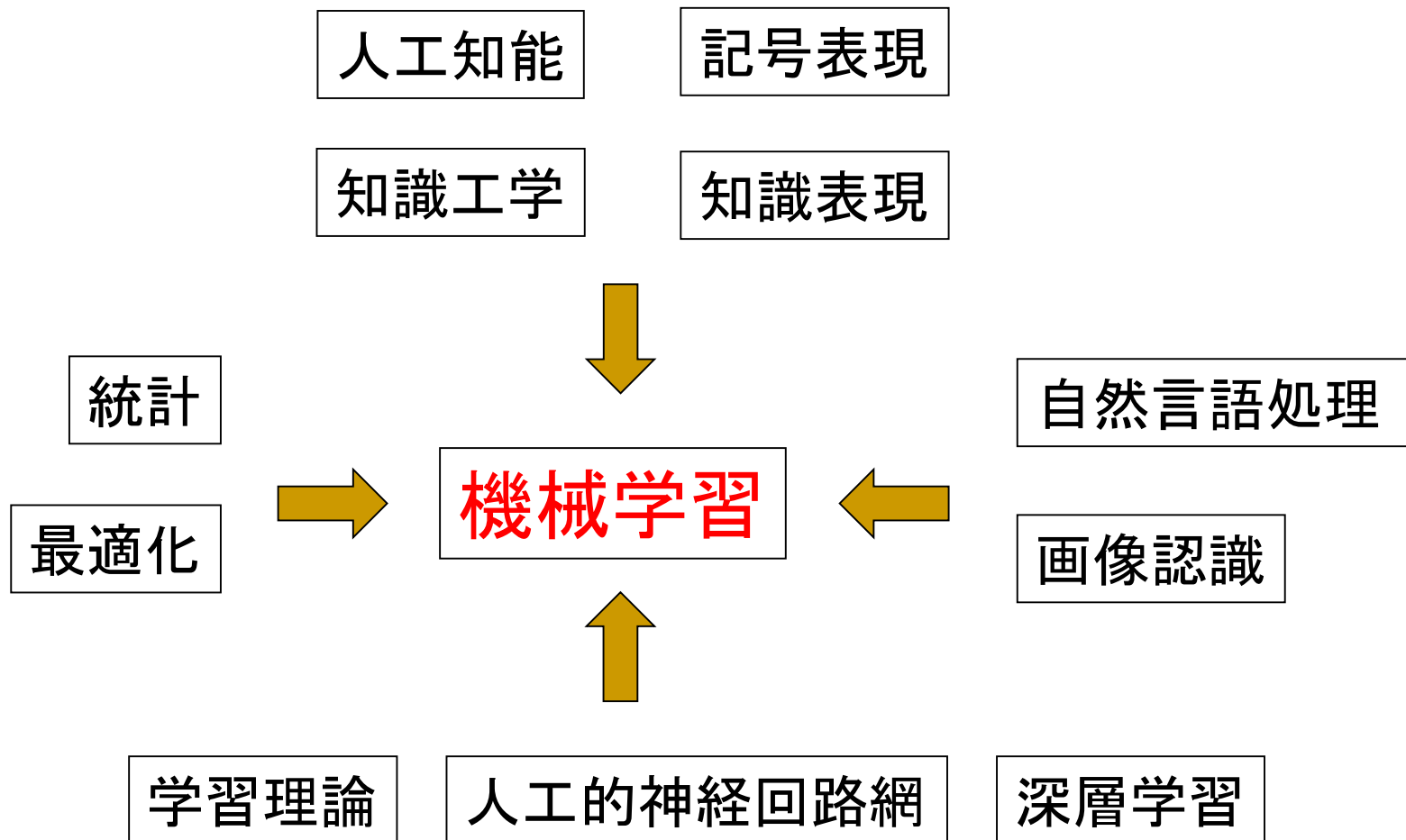
統計的機械学習

機械学習と数理モデル

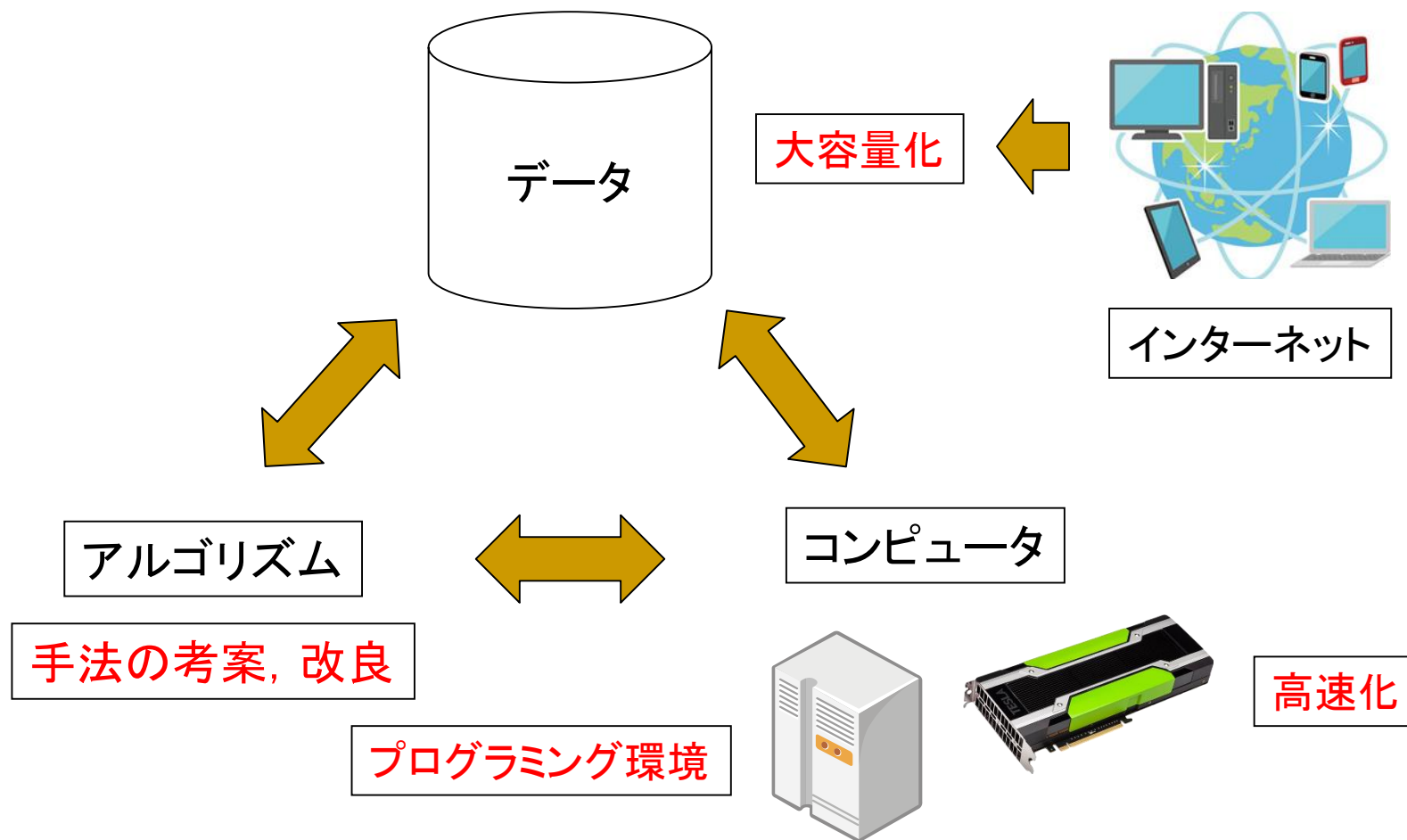
■ 最適化

- 入力データが正しく予測された場合, 間違って予測された合, それらを評価する指標(評価基準)を決める
- 評価基準を最大(もしくは最小)にするという「最適化問題」に帰着できる

機械学習をとりまく分野



機械学習をとりまく環境



機械学習を行なうためには？

- 大量のデータが必要
 - さらに、正解値（正解ラベル）が必要
 - （例）顔画像からの性別判定
 - 顔画像ごとに、男性か、女性かの正解ラベルをつけないといけない
- 大量のデータを学習するには？
 - 高性能な計算機が必要

講義の目的

- 機械学習を学ぶことによって,
 - 「未来」の値を予測する
 - 「未知」のものを予測する
 - 「似ている」ものがどれかを予測する
- 現実的なデータを対象として,
 - scikit-learn(Python)によるプログラミング

講義で扱う手法

- ロジスティック回帰
- 最近傍法, k近傍法
- 決定木
- ベイズ決定則
- 識別関数法
 - デルタルール, パーセプトロン, サポートベクターマシン
- アンサンブル学習
 - バギング, ブースティング
- ニューラルネットワーク(深層学習)
 - 誤差逆伝播則, 畳み込みニューラルネットワーク
- クラスタリング

講義内容

講義計画

講義の進め方, 講義資料

講義計画①

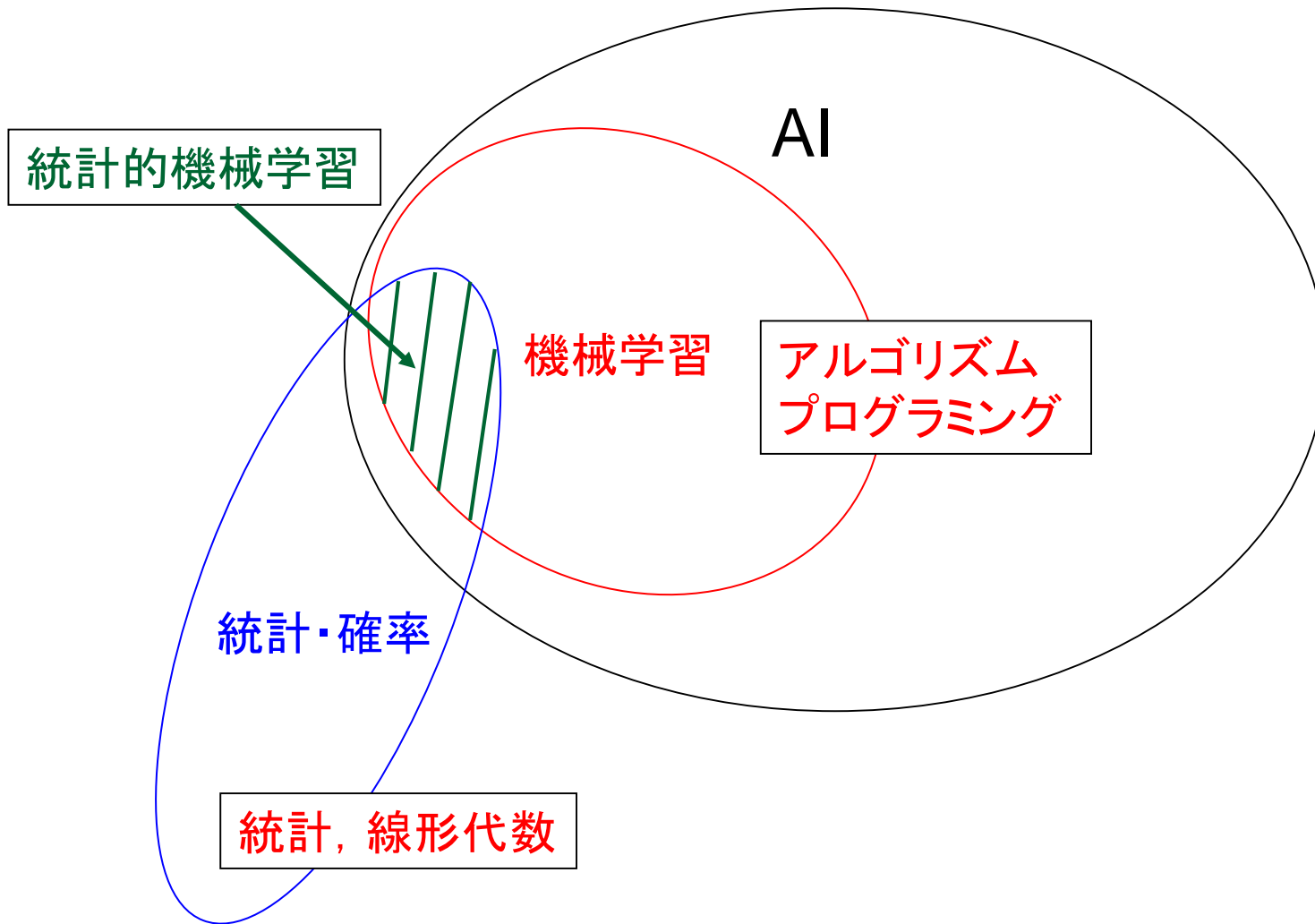
第一回	9月23日	ガイダンス Pythonについて
第二回	9月30日	機械学習入門① 分類問題と回帰問題
第三回	10月7日	機械学習入門② scikit-learnによる機械学習
第四回	10月21日	最近傍法, k近傍法
第五回	10月28日	決定木
第六回	11月4日	生成モデル① 統計的機械学習の基礎
第七回	11月11日	生成モデル② ベイズ決定則

講義内容は進捗によって変わることがあります

講義計画②

第八回	11月18日	識別モデル デルタルール, パーセプトロン
第九回	12月2日	サポートベクターマシン
第十回	12月9日	アンサンブル学習 バギング, ブースティング
第十一回	12月16日	ニューラルネットワーク① 階層型ニューラルネットワークの学習
第十二回	12月23日	ニューラルネットワーク② 畳み込みニューラルネットワーク
第十三回	1月6日	クラスタリング k-平均法
第十四回	1月20日	講義のまとめ

講義に必要な知識



講義資料

■ 教科書

- なし
- 参考書は適宜示す

■ 資料

- <http://lecture.comp.ae.keio.ac.jp/ML2019/>
- 講義に関する連絡は上記のURL上に掲載する

講義の進め方

■ 講義＋実習

- 最後の20分～30分程度は実習時間
- （行なわない回もあります）

■ 実習内容

- Python (scikit-learn)
- 表計算ソフト (MS-Excel)
- 計算問題

評価方法

- 講義中の演習問題
- レポート
 - 二回を予定(三田祭前, 冬休み前)
- 最終回にまとめの演習問題

講義に関する質問

■ 講義に関する情報

- <http://lecture.comp.ae.keio.ac.jp/ML2019/>

■ 質問

- 電子メール: shino@ ae.keio.ac.jp
- 篠沢の居室: 23-624

Pythonの補習

- 補習日
 - 9/28(土) 3時限目, 4時限目
 - 教室:12-109
- 内容
 - 本年度の第一回COM実験
- 補習を受ける場合, 9/27(金)13時まで shino@ae.keio.ac.jp宛にメールを送って下さい