

Pythonで学ぶ機械学習

2019年2月2日（土）第1回～第8回
講師：渡邊 貴之

概要

- 機械学習について…
 - 人工知能を実現する手段として機械学習に注目が集まっています。
 - 機械学習では、コンピュータに大量のデータを入力し学習させることでその特徴を明らかにしたり、またモデルに当てはめることで将来の予測に繋げることができます。
 - 機械学習プログラムを作成する際には、機械学習プログラムを簡単に記述できる豊富な追加モジュールが揃ったPythonが広く利用されています。
- 本講座では…
 - 機械学習についての学びを始める初学者の方を対象として、幅広く基本的な機械学習の考え方やアルゴリズムについて講義とPythonを用いた演習を行います。
 - 教師なし学習
 - 教師あり学習
 - ニューラルネットワークの基礎など
- 参考図書
 - “データサイエンティスト養成読本 機械学習入門編”, 技術評論社
 - “Python機械学習プログラミング”, インプレス
 - “実践機械学習システム”, オライリー・ジャパン

本日の内容

● タイムテーブル

| | | | |
|-----|---------------------------------|-----|------------------------------------|
| 第1回 | 10:00 ~ 10:45 機械学習とは | 第5回 | 14:00 ~ 14:45 教師あり学習と分類モデル1 |
| 第2回 | 10:45 ~ 11:30 教師なし学習とクラスタリング | 第6回 | 14:45 ~ 15:30 教師あり学習と分類モデル2 |
| 第3回 | 11:30 ~ 12:15 教師あり学習と回帰モデル1 | 第7回 | 15:30 ~ 16:15 ニューラルネットワークの基礎 |
| 第4回 | 13:15 ~ 14:00 教師あり学習と回帰モデル2 | 第8回 | 16:15 ~ 17:00 ニューラルネットワークを用いた分類 |

※途中、休憩時間を適宜取ります

実習室利用ガイダンス

● おねがい

- 教育・研究のための利用に限ります。
- 室内での飲食は禁止されています。

● コンピュータの起動とログオン

- 別紙「講座用ユーザIDとパスワード」をお手元にご用意下さい。
- コンピュータの電源を入れ、ログオン画面にてCtrl+Alt+Deleteを押し、「ユーザーの切り替え」「他のユーザー」をクリックします。
- ログオン用IDとパスワードを入力し、矢印ボタンをクリックします。
- ログイン後、中央モニタに講師のプロジェクタと同じ画面が映ります

その他

- 自販機
 - 1Fの階段横にあります。
- 喫煙場所
 - 大学敷地内は全面禁煙です。
- 食堂
 - はばたき棟(坂を下った右側の棟)の地下1階食堂は営業しています。
 - お弁当をお持ちの方は3F地域経営研究センターまたは、1F吹き抜けのホールをご利用ください。

第1回 機械学習とは

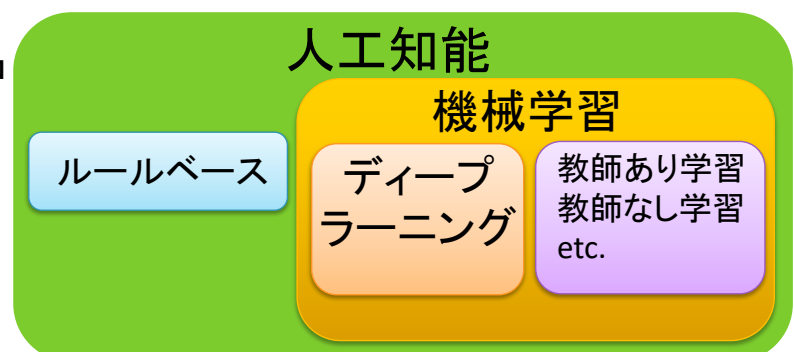
目次

- 機械学習とは
- 人工知能研究の歴史
- 機械学習の特徴と分類
- 教師あり学習
- 教師なし学習
- 中間的手法
- 学習のプロセスとモデル
- 機械学習のライブラリ
- 機械学習を行う環境を用意する
- Spyderを起動する
- 簡単なコードを記述して実行する
- 日本語環境設定

- 7 -

機械学習とは

- 人工知能を実現するための一分野
- アーサー・サミュエル(1901-1990: 米国の計算機科学者)
 - 明示的にプログラムをしなくても学習する能力をコンピュータに与える研究分野
- 人間が自然に行なっている学習能力と同様の機能をコンピュータで実現しようとする技術・手法
- データのパターンや特徴・法則性をモデルに当てはめて(学習し)、それをもとに未知のデータに対して何かしらの予測や推定を行う
- モデル=何らかの数式やデータ構造と、その中に含まれる変更可能なパラメータ値の集合



- 8 -

人工知能研究の歴史

- 機械学習は人工知能研究の過程で発生した技術
 - 人工知能(AI)の研究は1950年代から続いており、その過程ではブームと冬の時代が交互に訪れている
 - 1960年代～1970年代: 第一次人工知能ブーム
 - コンピュータによる「推論」や「探索」が可能となり、特定の問題に対して解を明示できるようになった
 - 単純な問題は解決できても、様々な要因が絡み合うような現実社会の問題には適応できなかった
 - 人間の脳を模倣したニューラルネットワーク(人工神経網)の研究が開始されるものの計算資源不足による実用化への停滞 ⇨ 冬の時代へ
 - 1980年代: 第二次人工知能ブーム
 - コンピュータが推論を行うのに必要なルールを与えることで、その分野の専門家のように振る舞うエキスパートシステムが多数研究された
 - 膨大な情報全てをコンピュータに理解できるようにルールを記述することは困難であり、対象は特定の領域に限られた ⇨ 冬の時代へ
 - 2000年代～現在: 第三次人工知能ブーム
 - WebやIoTの進展によりビッグデータ収集が容易となり、大量のデータを活かすことのできる機械学習が注目される
 - コンピュータの爆発的大規模高速化とGPU処理を背景として、ディープラーニングがブームに

- 9 -

機械学習の特徴

- ルールベース
 - if文や探査による多彩なパターンの網羅が困難
 - 新たなルールが出てきた際に、人手によるルールの追加が必要
 - 想定外のデータに対して、対応するのが困難になる
- 機械学習
 - 人間はルールではなくモデルをコンピュータに与える
 - モデルに対してデータを入力し、データに適合するようにモデルのパラメータを自動的に調整させる
 - ルールを人間が与える必要がなく、モデルのパラメータはコンピュータ自身が学習によって習得する
 - 想定外のデータを含めて、モデルのパラメータが自動的に調整される

ルールベース

すべて人手によるルール追加

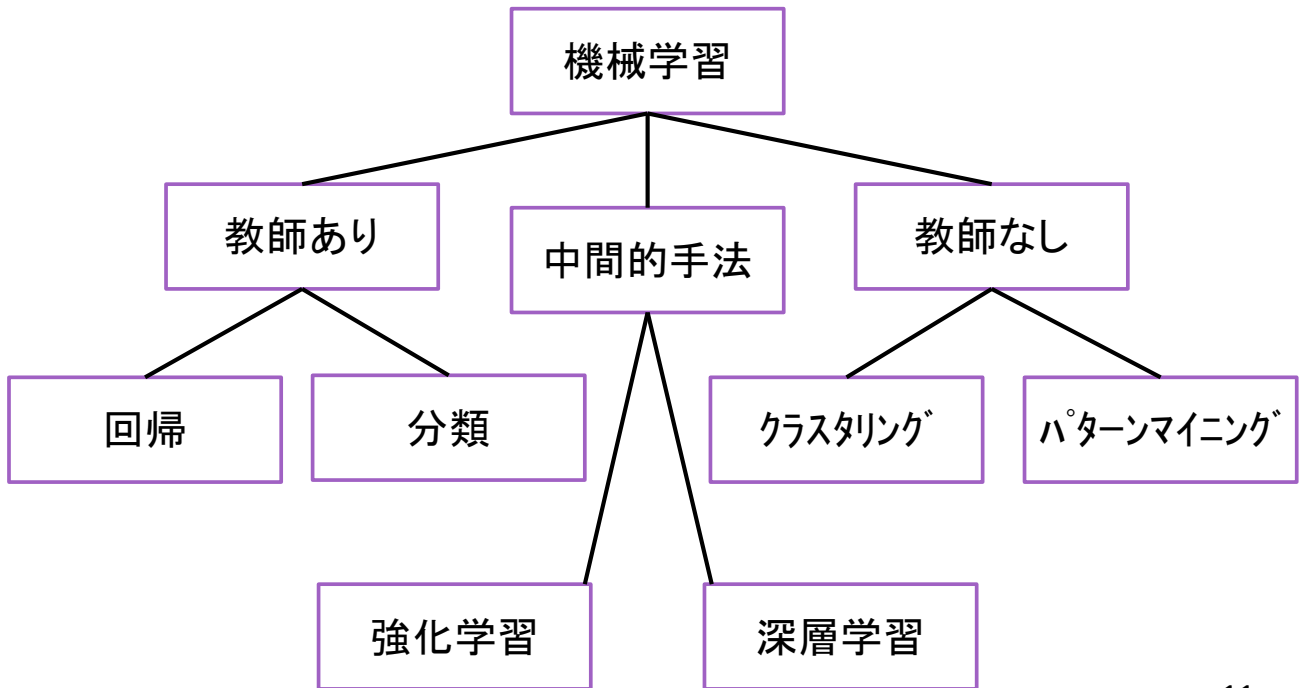
機械学習

モデルを与えてパラメータを自動調整させる

- 10 -

機械学習の分類

- 学習において、正解が付いているか(教師あり)、いないか(教師なし)で分類



- 11 -

教師あり学習

- 入力データが与えられたときに、これに対する出力を正しく予測することが目的
- データを与える際に、正解ラベルのついたデータを与えて学習させる
 - 正解ラベルが教師の役割を果たす
- 教師あり学習の種類
 - 回帰問題
 - 連続値として与えられたデータの予測に用いられる
 - 株価予測、気温の予測など
 - 手法: 線形回帰、ロジスティック回帰など
 - 分類問題
 - 学習データとして与えられた結果に基づいて、新しく与えられたデータがどのクラスに当たるのかを予測する
 - スпамメールの分類、手書き文字認識など
 - 手法: 決定木、SVM(サポートベクターマシン)、ベイズ推定など

- 12 -

教師なし学習

- データを与える際に正解ラベルは与えずに、学習を行う
- 出力する結果はあらかじめ決まっているわけではない
 - 与えられたデータにある頻出パターンや、共通項を持つクラスタを見つけ出す手法
- 教師なし学習の種類
 - クラスタリング
 - 与えられたデータから規則性など意味のある情報を見つけ出す
 - マーケティング分析、観光行動分析など
 - 手法:k-means法、最短距離法など
 - パターンマイニング
 - データ中に何度も出現するパターンや、そのパターンに基づいた規則を発見する
 - 代表事例は「ビールとおむつ」などのバスケット分析
 - 次元削減
 - 高次元のデータを低い次元の部分空間に圧縮する
 - 機械学習データの前処理など
 - 手法:主成分分析、線形判別分析など

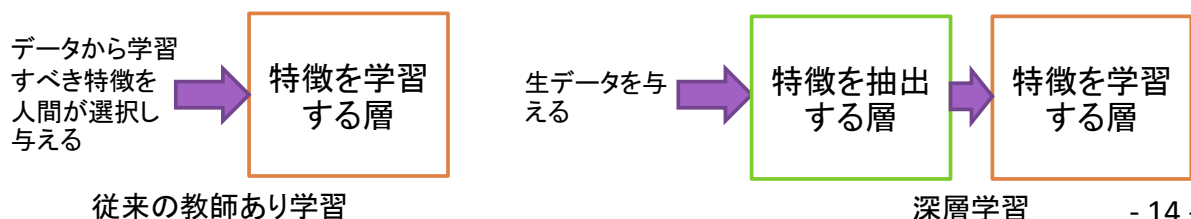
- 13 -

中間的手法

- 強化学習
 - 教師あり学習のように陽に正解が与えられるのではなく、正解を導いた際に報酬が与えられる学習方法
 - 「これが正解だから覚えてね」ではなく、正解が得られた際に「ご褒美」をあげる



- 深層学習
 - 多階層(層が深い)のニューラルネットワーク(人工神経網)をモデルとして用いた学習手法=ディープラーニング
 - 従来の機械学習ではデータ内で学習すべき特徴量を人間が考え与えていたが、深層学習ではどの特徴が有効なのかまでを機械学習する

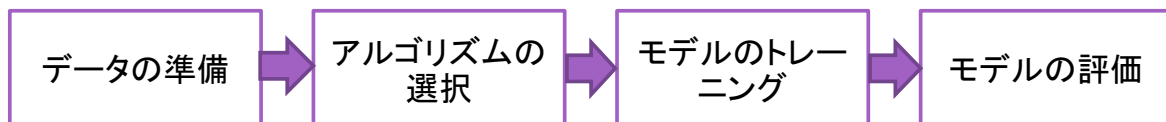


- 14 -

学習のプロセスとモデル

● 機械学習を行うための代表的な手順(教師ありの場合)

1. データの準備
 - 分析するデータを準備
 - 必要に応じてノイズを取り除いてクリーニング
 - 学習用にデータを変換
 - 教師データとテストデータに分割
2. 学習アルゴリズムの選択
 - どのような分析をおこなうかによって学習アルゴリズムは異なる
 - 適切なアルゴリズムを選ぶことでよりよいモデルが作成できる
3. 学習モデルのトレーニング
 - 教師データを用いてモデルの訓練を行う
 - パラメータのチューニング
4. 学習したモデルの評価
 - テストデータを用いて学習したモデルの評価を行う



- 15 -

機械学習のライブラリ

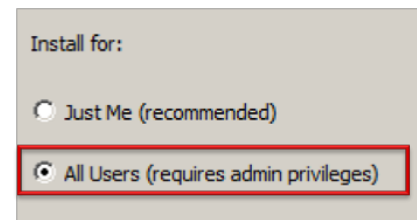
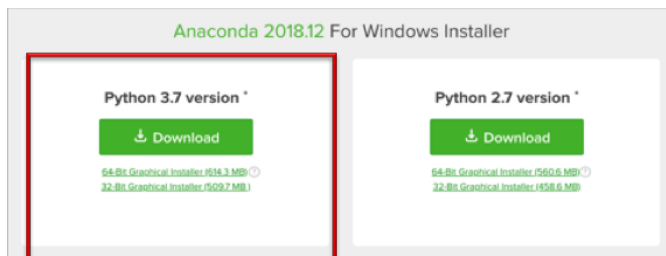
- Pythonには様々な機械学習のライブラリが用意されています
- ライブラリを用いる利点
 - それぞれのアルゴリズムを1から実装する必要がない
 - コードが簡潔に書ける
- 代表的な機械学習ライブラリ
 - scikit-learn
 - 回帰、ランダムフォレストなど様々なアルゴリズムを備えたライブラリ
 - genism
 - 様々なトピックモデルを実装したライブラリ
 - Tensor Flow
 - 代表的な深層学習ライブラリ
 - Keras
 - TensorFlowをベースにより扱いやすくした深層学習ライブラリ

- 16 -

機械学習を行う環境を用意する

Pythonの入手方法

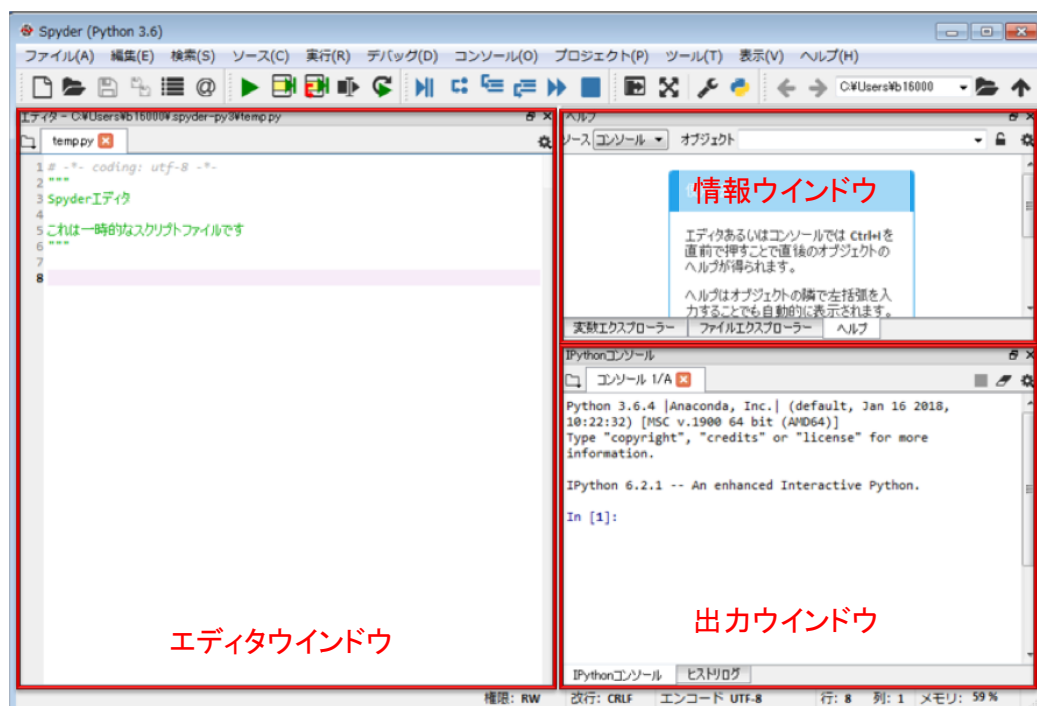
- この講座では、データサイエンス向けに作成されたPythonパッケージであるAnacondaを使用します
 - データサイエンス用の多くのモジュールが標準でインストールされるのでとても便利です
- 手順
 - Webブラウザで「<https://www.anaconda.com/download/#windows>」を開きます
 - この講座では、Pythonのバージョン3.xを使用するため、3.7をダウンロードしてください(実習室ではあらかじめインストールされています)
 - ダウンロード時にメールアドレスの入力を促す画面が出ますが、入力は不要です(入力・送信しなくてもダウンロードできます)
 - インストール時の「Install for」では、「All Users」を選択したと仮定して演習を進めます



インストール設定 - 17 -

Spyderを起動する

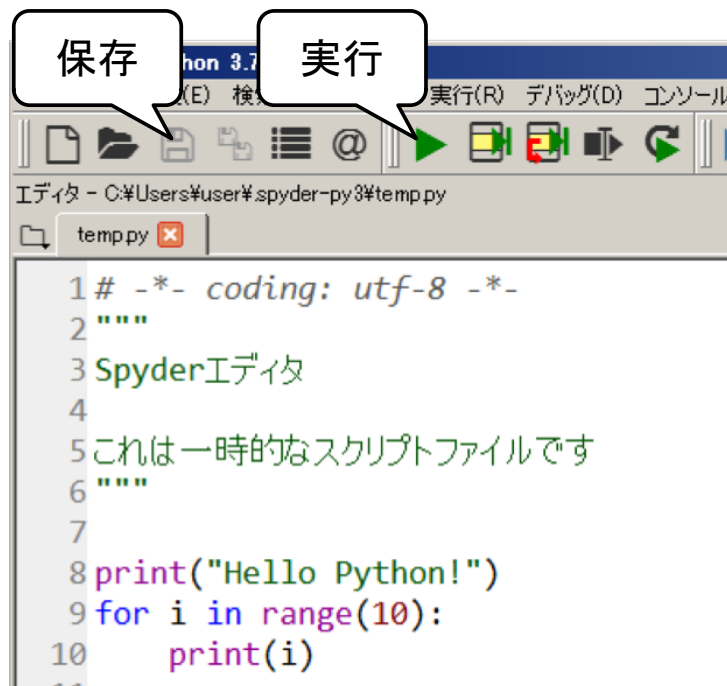
- [スタート]-[すべてのプログラム]-[Anaconda3]を開く
- Spyderを起動する



- 18 -

簡単なコードを記述して実行する

- 下記のような2行のprint関数とfor文を記述し保存、実行する。
 - 右下の出カウインドウ(コンソール)の結果を確認する。



- 19 -

Spyderの日本語環境設定(1)

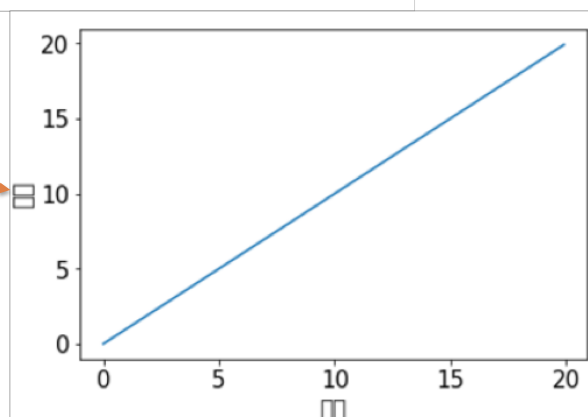
- 初期状態ではmatplotlibを使用してグラフを描く際に、日本語が文字化けする

```
import numpy as np
import matplotlib.pyplot as plt

xp = [x * 0.1 for x in range(0, 200)]
plt.plot(xp, xp)
plt.xlabel("横軸")
plt.ylabel("縦軸")
```

軸ラベルの設定

文字化け!



- 20 -

Spyderの日本語環境設定(2)

- matplotlibの日本語フォントの文字化け対策
 - 独立行政法人情報処理推進機構が配布している日本語フォントをインストールする(IPAフォント)
 - <http://pana4405.u-shizuoka-ken.ac.jp/pyk2>からIPAフォント(ipaexg00301.zip)をダウンロードする。
 - ダウンロードしたファイルをダブルクリック展開し、ipaexg.ttfをC:\ProgramData\Anaconda3\Lib\site-packages\matplotlib\mpl-data\fonts\ttfに入れる。
 - spyderを終了する。
 - [スタート]-[コンピュータ]を開き、C:\Users\自分のユーザアカウント名(今回は、kouza)のmatplotlibフォルダを削除する
 - spyderを再起動する

※正式な配布サイトは<https://ipafont.ipa.go.jp/node26>

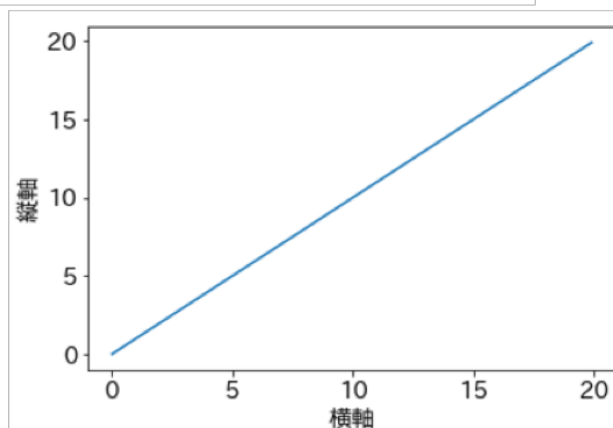
- 21 -

Spyderの日本語環境設定(3)

```
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams["font.family"] = "IPAexGothic"
plt.rcParams['font.size'] = 15
```

← IPAフォントの指定

```
xp = [x * 0.1 for x in range(0, 200)]
plt.plot(xp, xp)
plt.xlabel("横軸")
plt.ylabel("縦軸")
```



- 22 -

課題1

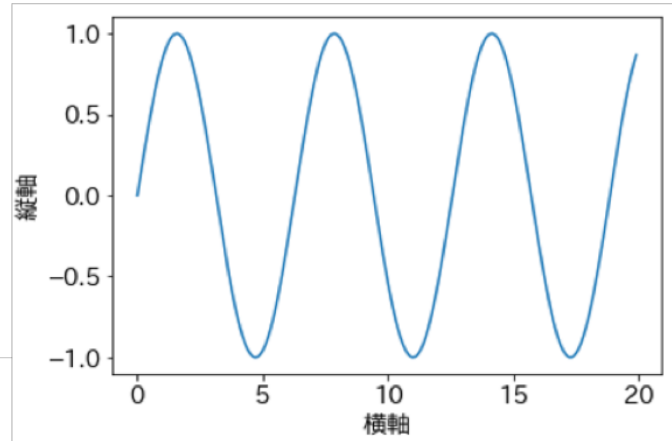
- matplotlibを用いて、

$$0 \leq x \leq 20$$

において、

$$y = \sin(x)$$

のグラフを描いてください。



```
import numpy as np
import math
import matplotlib.pyplot as plt
plt.rcParams["font.family"] = "IPAexGothic"
plt.rcParams['font.size'] = 15
```

```
xp = [x * 0.1 for x in range(0, 200)]
plt.plot(xp, ここを考える)
plt.xlabel("横軸")
plt.ylabel("縦軸")
```

numpyを使わない場合と使う場合