知能プログラミング演習 I

第 1 回: Python 入門/パーセプトロン

講義担当教員: 稲津 佑
スライド作成者: 稲津 佑，烏山 昌幸，田中 剛平，梅津 佑太

この講義について

講義の目的

- フリーの言語 python を使って深層学習の基礎を学ぶ
- 実際に自分で手を動かしてアルゴリズムを習得する
- Moodle に upload されている講義ノートを参照しながら進めること

成績評価

- 提出課題+レポート

この講義について

担当者

- 教員 稲津 inatsu.yu@nitech.ac.jp
- 教員 烏山 karasuyama@nitech.ac.jp
- 教員 田中 gtanaka@nitech.ac.jp
- TA (M2) 金山 kanayama.takuya.mllab.nit@gmail.com
- TA (M2) 野口 noguchi.shuto.mllab.nit@gmail.com
- TA (M1) 鹿谷 shikatani.rikugo.mllab.nit@gmail.com

講義内容 (予定)

1. python 入門/パーセプトロン
2. 3 層ニューラルネットワーク/深層ニューラルネットワーク
3. オートエンコーダー
4. 時系列のためのニューラルネット: リカレントニューラルネットワーク
5. 画像のためのニューラルネット: 畳み込みニューラルネットワーク I
6. 画像のためのニューラルネット: 畳み込みニューラルネットワーク II
7. 言語のためのニューラルネット: トランスフォーマー
8. 発展的な話題

休講，オンデマンド形式の実施

- 7 月 16 日は休講
- 7 月 23 日に「7. 言語のためのニューラルネット: トランスフォーマー」を実施する
- 「8. 発展的な話題」については 7 月 23 日の演習終了後にアップロードされた講義動画を視聴するオンデマンド形式とする

今日の講義内容

1. python の使い方
2. パーセプトロンの学習

python の起動と終了

- ターミナルを起動し, `python` または `ipython` を実行することで利用可能.
  - ✓ バージョンも含めて起動する場合は `python3` を実行する

- ファイルをターミナルから実行する場合,

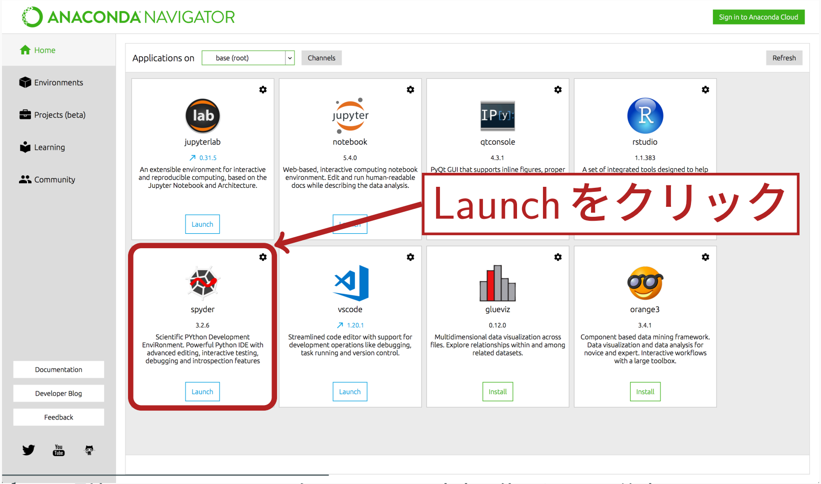
  `python　(ファイル名).py`

  などとすれば良い.

- python を終了する場合は `quit()` または `exit()` を実行

自前の PC で python を利用する場合

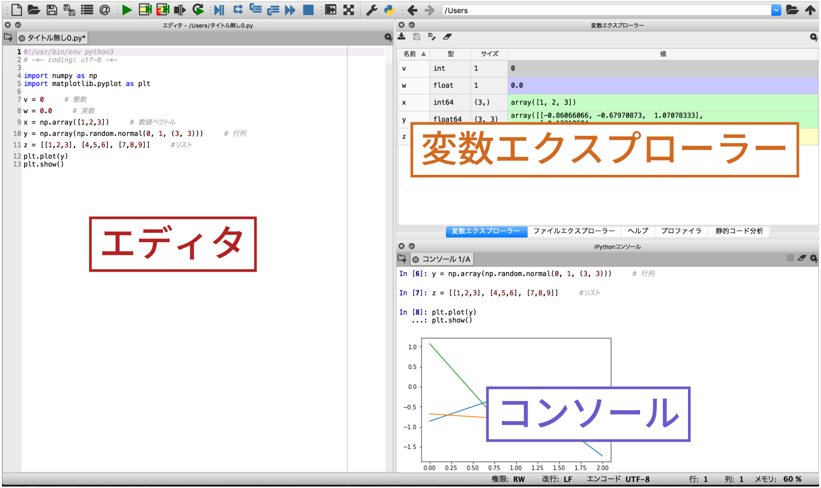
- Anaconda を自前の PC にインストールするのが便利で良い [1]
- Anaconda, spyder の順に起動し python を立ち上げる



[1] CSE 環境ではこのスライドと次のスライドの内容は使えないので注意

spyder

- エディタでアルゴリズムを作成 → コンソールで実行
- 定義済みの変数は変数エクスプローラーで表示される



numpy と matplotlib

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

import numpy as np
import matplotlib.pyplot as plt

```

- numpy: ベクトルや行列演算のための数値計算モジュール
- matplotlib: グラフ描画ライブラリ

python のデータタイプ

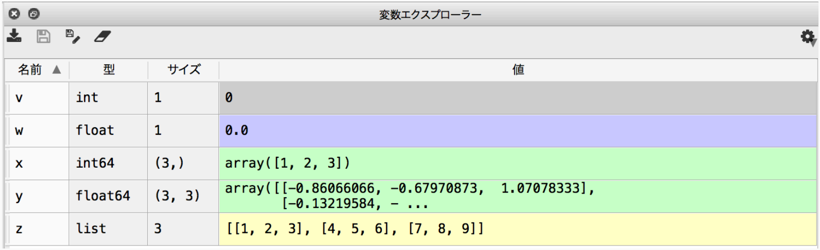
- 数値

  v = 0　# 整数 (int),　w = 0.0　# 実数 (float)
- ベクトル

  x = np.array([1, 2, 3])　# 数値ベクトル
- 行列

  y = np.array(np.random.normal(0, 1, (3, 3)))　# 行列
- リスト

  z = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]　# リスト


変数エクスプローラー

名前 ▲	型	サイズ	値
v	int	1	0
w	float	1	0.0
x	int64	(3,)	array([1, 2, 3])
y	float64	(3, 3)	array([[-0.86066066, -0.67970873, 1.07078333], [-0.13219584, - ...
z	list	3	[[1, 2, 3], [4, 5, 6], [7, 8, 9]]

ベクトル演算 I

```
In [2]: x = np.array([1, 2, 3], float)

In [3]: x[0]
Out[3]: 1.0

In [4]: x + 0.5
Out[4]: array([1.5, 2.5, 3.5])

In [5]: 2*x      # 要素ごとの定数倍
Out[5]: array([2., 4., 6.])

In [6]: x**3      # 要素ごとのべき
Out[6]: array([ 1.,  8., 27.])

In [7]: x*x      # 要素ごとの積
Out[7]: array([1., 4., 9.])
```

ベクトル演算 II

```
In [8]: y = np.array([4, 5, 6], float)

In [9]: x + y      # 要素ごとの和
Out[9]: array([5., 7., 9.])

In [10]: np.dot(x, y)      # 内積
Out[10]: 32.0

In [11]: np.outer(x, y)      # 行列積
Out[11]:
array([[ 4.,  5.,  6.],
       [ 8., 10., 12.],
       [12., 15., 18.]])

In [12]: np.outer(y, x)      # 行列積
Out[12]:
array([[ 4.,  8., 12.],
       [ 5., 10., 15.],
       [ 6., 12., 18.]])
```

行列演算

```
In [2]: A = np.random.normal(0, 1, (3,3))     # 3*3行列

In [3]: B = np.random.normal(0, 1, (2,3))     # 2*3行列

In [4]: A
Out[4]:
array([[-0.54402056, -2.25872705, -0.84008799],
       [-0.45872658, -0.65348438,  0.14641837],
       [-1.9583066 ,  0.11243239,  1.5944507 ]])

In [5]: A.T      # 転置
Out[5]:
array([[-0.54402056, -0.45872658, -1.9583066 ],
       [-2.25872705, -0.65348438,  0.11243239],
       [-0.84008799,  0.14641837,  1.5944507 ]])

In [6]: np.dot(B, A)     # 積BA
Out[6]:
array([[-2.68199056,  0.70173615,  2.38960981],
       [-1.84463277, -2.56791097, -0.06458365]])

In [7]: np.dot(A, B)     # Aの列数とBの行数が異なるのでエラーが表示される
Traceback (most recent call last):

  File "<ipython-input-7-bb50fe1a139b>", line 1, in <module>
    np.dot(A, B)     # Aの列数とBの行数が異なるのでエラーが表示される

ValueError: shapes (3,3) and (2,3) not aligned: 3 (dim 1) != 2 (dim 0)
```

for 文と if 文

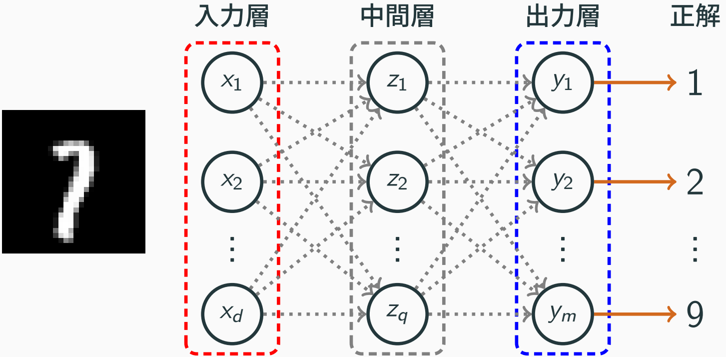
- $i = 0, 1, \ldots, 9$ まで順番に足す $(x = 0 + 1 + \cdots + 9)$

```
In [1]: x = 0
   ...: for i in range(0, 10):
   ...:     x = x+i
   ...:
   ...:

In [2]: x
Out[2]: 45
```

- j が偶数なら y に $j/2$ を加え，それ以外なら $(j+1)/2$ を加える

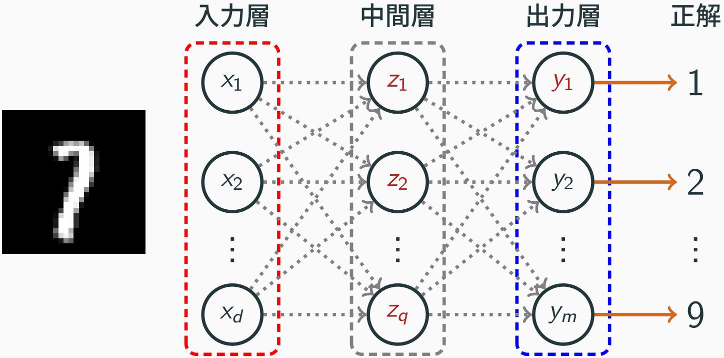
```
In [3]: y = 0
   ...: for j in range(0, 10):
   ...:     if j % 2 == 0:
   ...:      y = y + j/2
   ...:     else:
   ...:      y = y + (j+1)/2
   ...:
   ...:

In [4]: y
Out[4]: 25.0
```

ニューラルネットワーク



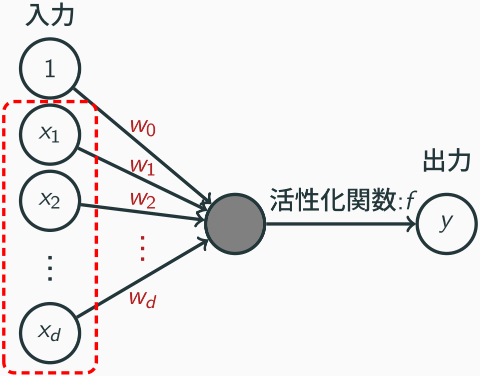
- ネットワークから得られた出力を正解と比較してモデルを学習

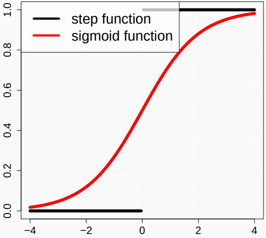
ニューラルネットワーク



- ネットワークから得られた出力を正解と比較してモデルを学習
- 各ユニットにつながるネットワークはパーセプトロンと呼ばれる

(一層) パーセプトロンによる分類器の学習





活性化関数

右図のような活性化関数 f を用いて

$$y \approx f(w_0 + w_1 x_1 + \cdots + w_d x_d)$$

によって入力と出力の関係をモデル化し，データから重み $w_0, w_1, \ldots, w_d$ を学習

ここでは二値分類 $y \in \{1, 0\}$ を仮定

以降では，活性化関数は微分可能な sigmoid 関数（右図赤線）を使用

誤差関数

誤差関数 (クロスエントロピー)

$$E(\boldsymbol{w}) = -y \log f(\boldsymbol{w}^\top \boldsymbol{x}) - (1-y)\log(1 - f(\boldsymbol{w}^\top \boldsymbol{x}))$$

- ある $\boldsymbol{x}$ と y のペアに対する誤差
  - if $y = 1 \to E(\boldsymbol{w}) = -\log f(\boldsymbol{w}^\top \boldsymbol{x})$
    このとき，$f(\boldsymbol{w}^\top \boldsymbol{x})$ が大きいほど誤差 $E(\boldsymbol{w})$ は小
  - if $y = 0 \to E(\boldsymbol{w}) = -\log(1 - f(\boldsymbol{w}^\top \boldsymbol{x}))$
    このとき，$f(\boldsymbol{w}^\top \boldsymbol{x})$ が小さいほど誤差 $E(\boldsymbol{w})$ は小
- エントロピーや対数尤度として解釈可能（今日は省略）

確率的勾配降下法

- ランダムに訓練データ集合 $\{(y_i, \boldsymbol{x}_i)\}$ を並べ替える
- 順番にデータ $y, \boldsymbol{x}$ を読み，誤差関数

$$E(\boldsymbol{w}) = -y \log f(\boldsymbol{w}^\top \boldsymbol{x}) - (1-y)\log(1 - f(\boldsymbol{w}^\top \boldsymbol{x}))$$

  が減少する方向に重み $w_0, w_1, \ldots, w_d$ を勾配降下法で更新する
  - 一度の更新ではランダムに並び替えて選ばれた一つの $\boldsymbol{x}$ と y だけ使うので「確率的」勾配
- 適当な初期値 $\boldsymbol{w}^0$ から始め，学習率を η_t として重みを更新 [2]（導出は次頁以降）

確率的勾配降下法

$$\boldsymbol{w}^{t+1} = \boldsymbol{w}^t - \eta_t \Big(f(\boldsymbol{w}^{t\top} \boldsymbol{x}) - y \Big) \boldsymbol{x}$$

[2] 誤差が減少する様子はエポック (データ全体を 1 回スキャンすること) ごとに確認する

誤差関数とその微分

誤差関数

$$E(\boldsymbol{w}) = -y \log f(\boldsymbol{w}^\top \boldsymbol{x}) - (1-y)\log(1 - f(\boldsymbol{w}^\top \boldsymbol{x}))$$

確率的勾配降下法

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} - \eta_t \nabla E(\boldsymbol{w}^{(t)})$$

- 誤差関数の微分:

$$\begin{aligned}\frac{\partial E(\boldsymbol{w})}{\partial \boldsymbol{w}} &= -y \underbrace{\frac{1}{f(\boldsymbol{w}^\top \boldsymbol{x})}\frac{\partial f(\boldsymbol{w}^\top \boldsymbol{x})}{\partial \boldsymbol{w}}}_{\text{対数関数の微分}} - (1-y)\underbrace{\frac{-1}{1-f(\boldsymbol{w}^\top \boldsymbol{x})}\frac{\partial f(\boldsymbol{w}^\top \boldsymbol{x})}{\partial \boldsymbol{w}}}_{\text{対数関数の微分}} \\ &= (f(\boldsymbol{w}^\top \boldsymbol{x}) - y)\boldsymbol{x}\end{aligned}$$

シグモイド関数とその微分

シグモイド関数

$$y = f(w_0 + w_1 x_1 + \cdots + w_d x_d) = f(\boldsymbol{w}^\top \boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}^\top \boldsymbol{x}}}$$

- 練習:

$$f(x) = \frac{1}{1+e^{-x}} \quad \Rightarrow \quad \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{e^{-x}}{(1+e^{-x})^2} = f(x)(1-f(x))$$

- シグモイド関数の微分[3]:

$$\frac{\partial f(\boldsymbol{w}^\top \boldsymbol{x})}{\partial \boldsymbol{w}} = \underbrace{\frac{\partial f(\boldsymbol{w}^\top \boldsymbol{x})}{\partial \boldsymbol{w}^\top \boldsymbol{x}} \frac{\partial \boldsymbol{w}^\top \boldsymbol{x}}{\partial \boldsymbol{w}}}_{\text{合成関数の微分}} = f(\boldsymbol{w}^\top \boldsymbol{x})(1 - f(\boldsymbol{w}^\top \boldsymbol{x}))\boldsymbol{x}$$

[3]偏微分を ∇ で表すこともある

まとめ

- ニューラルネットの最も基本的な構造: パーセプトロン
  - 入力の重み付き和を活性化関数に通し，出力を表現
- 確率勾配法によるパラメータの最適化
  - データを１点ランダムに選び，誤差関数の勾配方向に小さなステップ幅で重みを更新