IoT環境における知的情報処理技術

二宮 崇
愛媛大学
ninomiya@cs.ehime-u.ac.jp

本講座の概要

- 深層学習ツールであるPyTorchを用い、IoT環境においてリアルタイムに学習する深層学習の技術について学ぶ。

- 前半はPythonコードを基に深層学習について学び、後半は、深層学習ツールであるPyTorchをインストールし、実際にPCとエッジコンピュータ上で動作する深層学習器を作成することで、深層学習の技術を学ぶ。

スケジュール

- 10月11日(金) 9:00～18:00
  - 1. 導入【講義】
  - 2. 環境構築【実習】
  - 3. 深層学習(1): ニューラルネットワークの仕組み、活性化関数、推論について学ぶ【講義】
  - 4. Python: PythonとPyTorchの基礎を演習を通して学ぶ【実習】
- 10月12日(土) 9:00～18:00
  - 5. 深層学習(2): ニューラルネットワークの学習、損失関数、勾配法について学ぶ【講義】
  - 6. 深層学習(3): 誤差逆伝播法、計算グラフ、畳み込みニューラルネットワークについて学ぶ【講義】
  - 7. PyTorch演習: PyTorchを用いた深層学習の演習を行う。【実習】

教材

- 授業は配布資料をベースに進める
- ソースコード

http://aiweb.cs.ehime-u.ac.jp/~ninomiya/enpitpro/

- 参考書
  - 斎藤康毅, ゼロから作るDeep Learning――Pythonで学ぶディープラーニングの理論と実装, オライリージャパン, 2016.
  - Guido van Rossum, Pythonチュートリアル 第3版, オライリージャパン, 2016. (https://docs.python.jp/3/tutorial/)
  - 岡谷貴之, 深層学習 (機械学習プロフェッショナルシリーズ), 講談社, 2015.
  - 神嶌 敏弘, 機械学習のPythonとの出会い, 2017. (http://www.kamishima.net/mlmpyja/)
  - Pytorch公式サイト(https://pytorch.org/)

自己紹介

- **二宮 崇 (にのみや　たかし)**
- **愛媛大学 大学院理工学研究科　情報工学講座 教授**
- **言語処理学会理事、AAMT理事**
- **経歴:**
  - 1986～1989 愛大附属中
  - 1989～1992 松山東
  - 1992～1996 東京大学 理学部 情報科学科
  - 1996～2001 東京大学 大学院理学系研究科 情報科学専攻
  - 2001～2006 JST研究員
  - 2006～2010 東京大学 情報基盤センター 講師
  - 2010～ 愛媛大学 大学院理工学研究科 准教授、教授

辻井研

中川研

辻井潤一
産総研
元人工知能センター長

中川裕志
理研 チームリーダー5

導入: 人工知能とは？

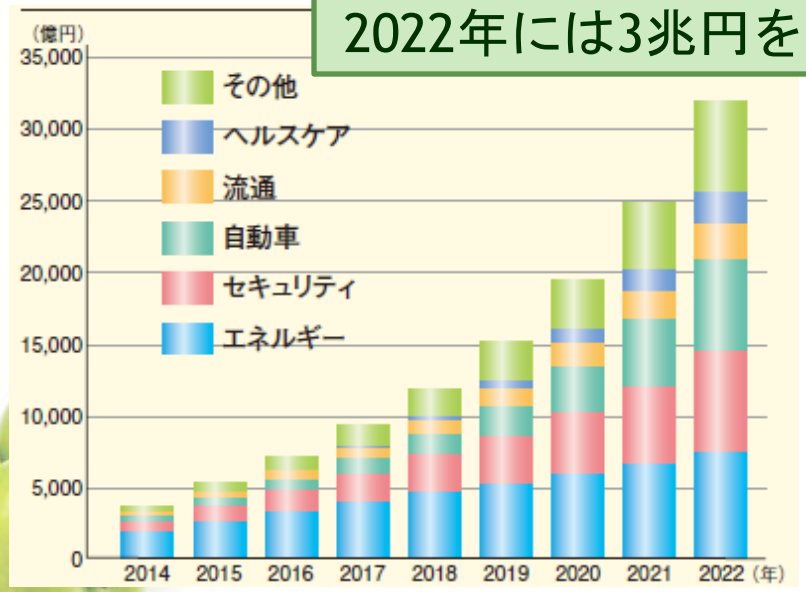
背景 (1/3)

- **IoT (Internet of Things; モノのインターネット)市場の急速な拡大**
  - ネットワークインフラ、インターネット技術の発展
  - センサー等のハードの低廉化、小型化

  など

IoT市場の分野別規模予測（野村総合研究所 2017）

2022年には3兆円を超える？

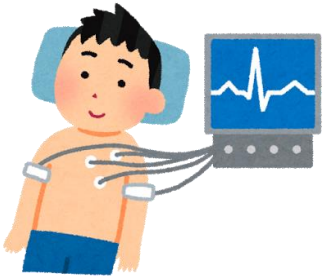


https://www.nri.com/~/media/PDF/jp/opinion/teiki/m_review/2017/nmr38-1.pdf

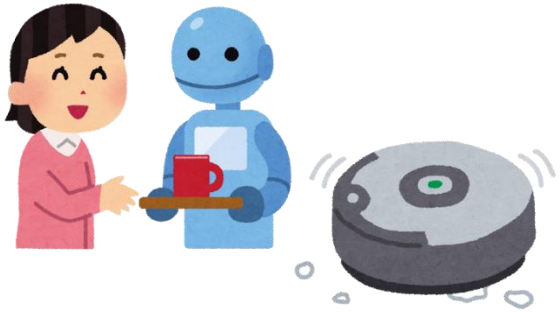
背景 (2/3)

- 人工知能(Artificial Intelligence; AI)が様々な場面で活躍しはじめている

将棋・囲碁



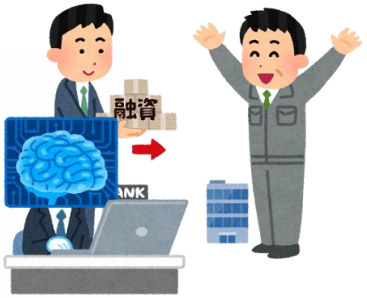
医療診断



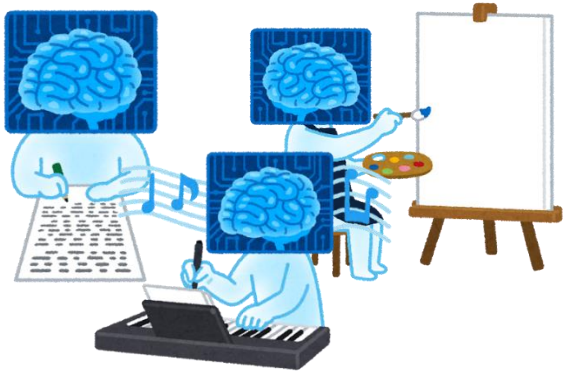
家事手伝いロボット

対話ロボット

証券取引



融資



芸術

自動運転

背景 (3/3)

- IoT＋AIが注目を集めている

AIで分析→
活用

エッジにAI搭載→
エッジが自動学習し知的動作が可能に

知的IoT環境

本講座では、

IoT+AI

↑理論と技術を学ぶ

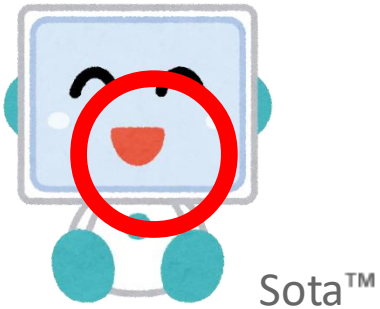
講義：AI（深層学習）の理論を学習
実習：PC上でAI（深層学習器）を作成
実習：Jetson Nano上でAI（深層学習器）を作成

※IoT環境下を想定し、低価格で超小型なシングルボードPCでAIを実現する（深層学習器を作成する）

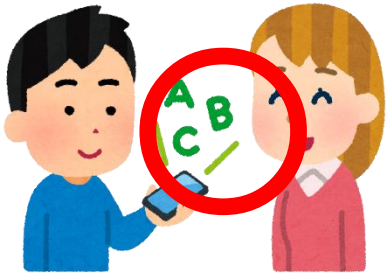
Jetson Nano
大きさ：100 x 80 x 29mm
価格：約15,000円

人工知能(Artificial Intelligence; AI)とは？

コミュニケーションロボット

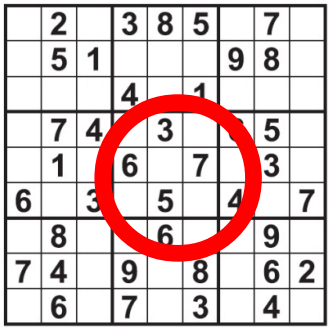


自動翻訳機／プログラム



形態は問わない

パズルを解くプログラム
（候補の全探索）



	2		3	8	5		7	
	5	1				9	8	
			4		1			
	7	4		3			5	
	1		6		7		3	
6		3		5		4		7
	8			6			9	
7	4		9		8		6	2
	6		7		3		4	

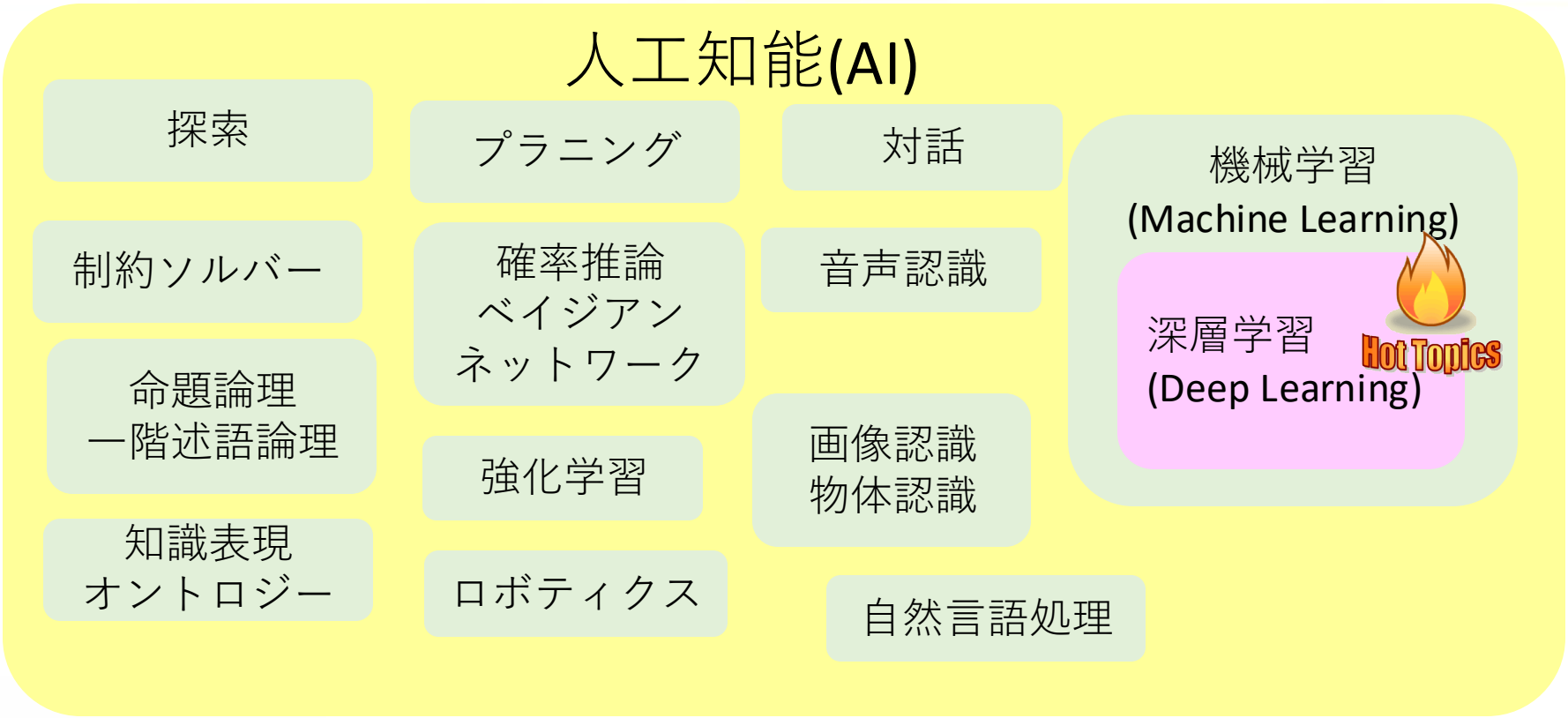
振る舞いが知的であればよい

電卓

「知的なモノ(機械、プログラム、技術)」

※ 研究者間で共有された明確な定義はない
※ 知性や知能の定義自体がない

人工知能分野



機械学習：データから法則を見つけ出す手法

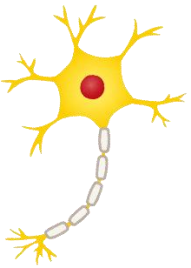
深層学習：ニューラルネットワーク(NN)を複数層重ねたモデル

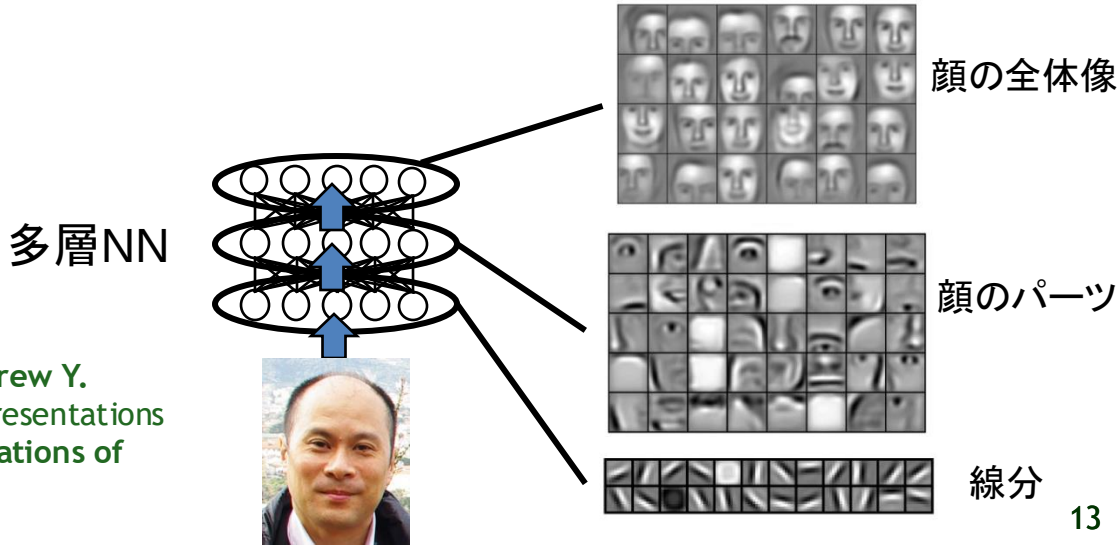
Stuart Russell, Peter Norvig (2010) **Artificial Intelligence: A Modern Approach, 3rd Edition.**

深層学習

- **深層学習 ＝　多層ニューラルネットワーク(NN)による学習**
  - ニューラルネットワーク
    - 人間の脳の神経細胞(ニューロン)の仕組みを模した計算モデル
      - パーセプトロン (Rosenblatt 1958)
      - 畳み込みNN (福島 1980)(LeCun+ 1989)
  - 特徴抽出、表現学習

深層学習では入力の特徴が自動的に学習されることがわかってきた





Honglak Lee, Roger Grosse, Rajesh Ranganath, Andrew Y. Ng (2011) Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks, **Communications of the ACM, Vol. 54 No. 10, Pages 95-103 より引用**

深層学習のインパクト(1/3)

- **深層学習の圧倒的精度**
  - 2012年のILSVRC (ImageNet Large Scale Visual Recognition Challenge)における画像分類の精度
    - 2010年 71.8% (NEC Labs America, Univ. of Illinois at Urbana-Champaign, Rutgers Univ.)
    - 2011年 74.2% (Xerox Research Center Europe, CIII)
    - 2012年 83.6% (Univ. of Toronto) … この年に深層学習が圧勝
    - 2013年 88.3% (Clarifai)
    - 2014年 93.3% (Google)
    - 2015年 96.4% (MSRA) … 人間の分類精度(95%)を超えたと言われる
    - 2016年 97.0% (公安調査庁第3研究所, 中国)
    - 2017年 97.75% (Momenta, Univ. of Oxford)

Olga Russakovsky+ (2015) ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, Volume 115, Issue3, pp 211-252
http://image-net.org/challenges/LSVRC/2017/index.php

AIの歴史

- **AIブームと冬の時代**

1950年代　1960年代　1970年代　1980年代　1990年代　2000年代　2010年代

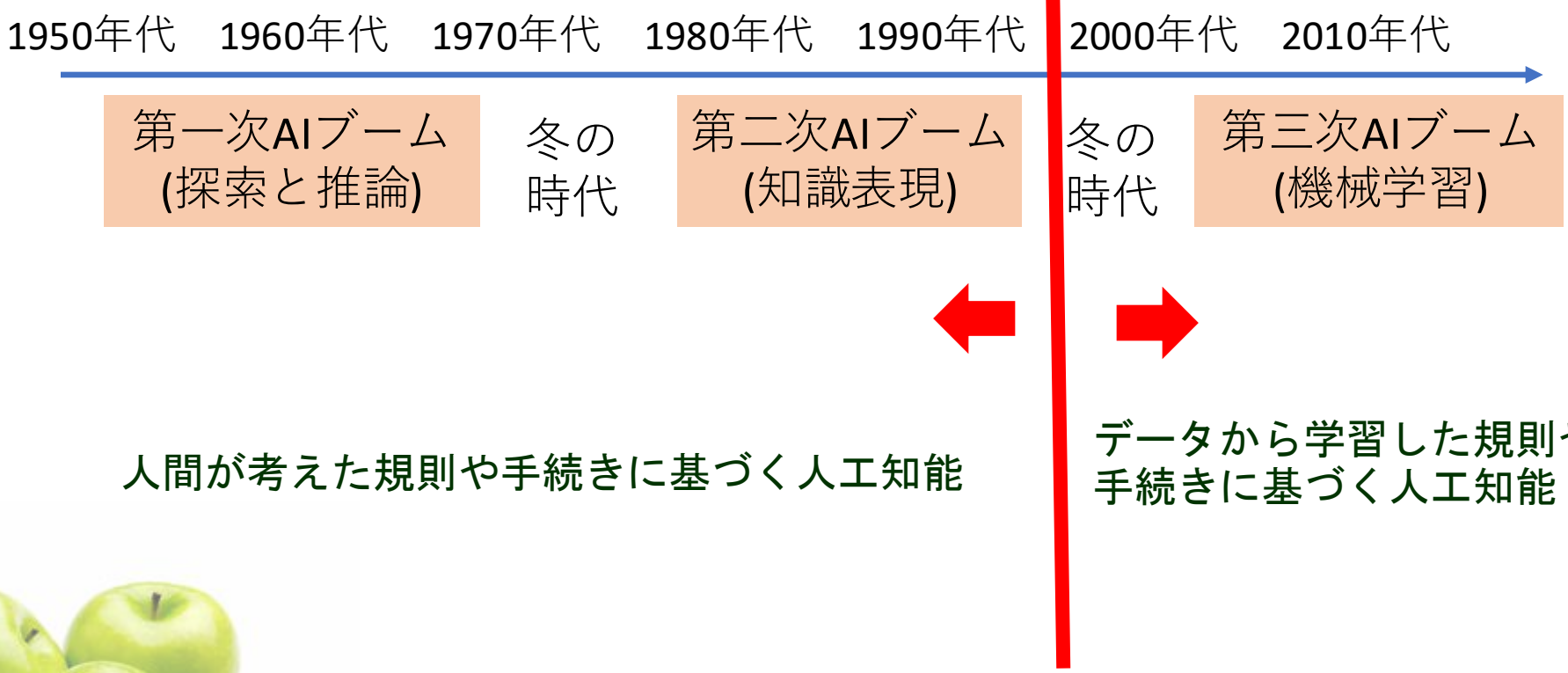
第一次AIブーム (探索と推論)　冬の時代　第二次AIブーム (知識表現)　冬の時代　第三次AIブーム (機械学習)

社会の期待 >> 技術的限界

冬の時代を迎えたくない…

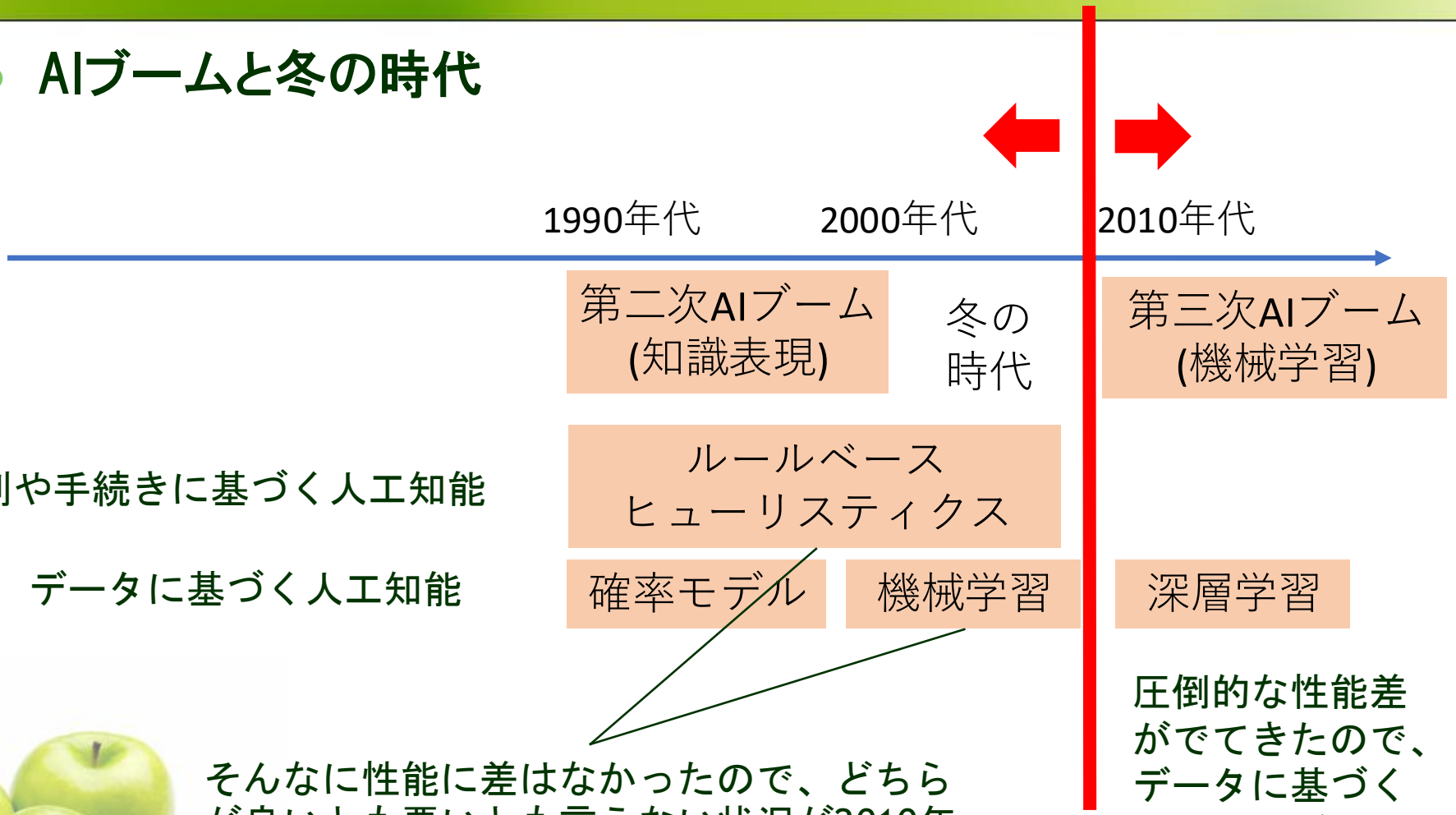
AIの歴史

- **AIブームと冬の時代**



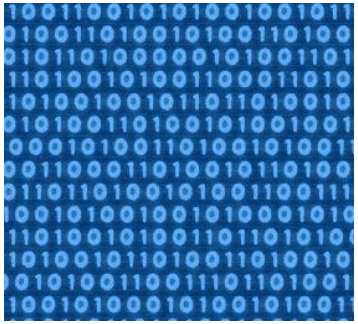
AIの歴史

- AIブームと冬の時代

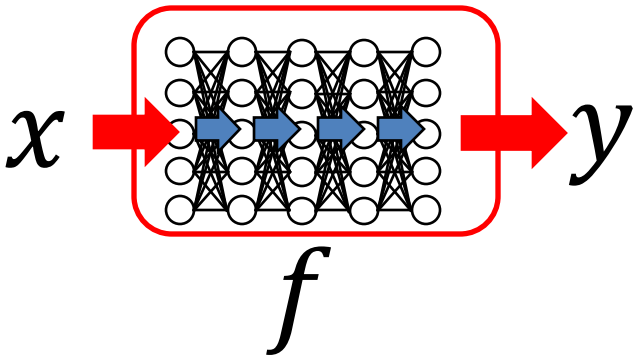


機械学習

- 機械学習 = データから関数を学習する学問分野

大量のデータ  $f(x)$

- 深層学習も機械学習の一種



機械学習

- **データから関数を学習**
  - 関数は、入力(x)と出力(y)の関係を表す

$$x \xrightarrow[f]{} y$$

  - 大量の入出力ペア(x, y)の集まり(データ)から関数fを自動的に学習！

データ

$$(x_1, y_1)$$
$$(x_2, y_2)$$
$$(x_3, y_3)$$
$$\vdots$$
$$(x_n, y_n)$$

学習 → $f(x)$

学習には、入力(x)と出力(y)のペアが揃ったデータが必要！(教師つき学習)

正解を必要としない学習(教師なし学習)もあるけど、そんなに高い精度を実現するわけではない

機械学習

例：犬猫判別機 $f(x)$

入力x

$\xrightarrow{f}$

出力y

犬

データ


：猫

：犬


：犬

：猫


：犬 …

大量のデータから関数f を学習する　f: 画像 → {犬/猫}

機械学習

- **データから学習**
  - 入力 $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$　←m次元ベクトル
    （画像の場合は、赤、青、緑の画像に対応する3つの行列）
  - データDは入力$\boldsymbol{x}$と出力yのペアの集合

  - 関数fはパラメータ(重み変数)の集合 $\boldsymbol{w} = (w_1, w_2, \cdots w_m)$から成る
    例: $f(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_m x_m$
    - fは$f_{\boldsymbol{w}}(\boldsymbol{x})$と書くとわかりやすい

  - 学習 = データDに対する誤差(損失)を最小にする重み変数wを求める

$$\text{データ全体の誤差(損失) } L(\boldsymbol{w}) = \sum_{(\boldsymbol{x},y)\in D} \left(y - f_{\boldsymbol{w}}(\boldsymbol{x})\right)^2$$

他にも、マージン最大化、確率最大化による学習がある。
関数の最小値を求める方法は「最適化」という領域で研究されている

機械学習における学習と推論

- **学習 (learning)**
  - データから良いパラメータを推定すること
  - 誤差(損失)を最小化することによりパラメータが得られる
  - 推定 (estimation)、パラメータ推定 (parameter estimation) ともいう
  - 最尤推定、MAP推定、ベイズ推定、マージン最大化が有名

- **推論 (inference)**
  - 未知のデータに対し、学習した関数fを適用し、出力を予測すること
  - 予測ともいう

機械学習の教科書

- **機械学習**
  - Christopher M. Bishop, **Pattern Recognition and Machine Learning**
    - (邦訳) C. M. ビショップ他 パターン認識と機械学習 上・下
  - Nello Cristianini, John Shawe-Taylor, **An Introduction to Support Vector Machines and other kernel-based learning methods**
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman, **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**
    - **(邦訳)** Trevor Hastie, Robert Tibshirani, Jerome Friedman他,統計的学習の基礎 ―データマイニング・推論・予測―
- **数値最適化**
  - Jorge Nocedal, Stephen Wright, **Numerical Optimization**
  - Dimitri P. Bertsekas, **Nonlinear Programming**
  - 金森 敬文, 鈴木 大慈, 竹内 一郎, 佐藤 一誠, **機械学習のための連続最適化 (機械学習プロフェッショナルシリーズ)**

回帰問題と分類問題と構造予測

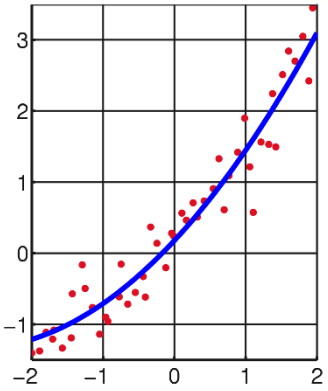
関数: $y = f(x)$

- **回帰問題 (regression)**
  - $y \in R$ (実数)

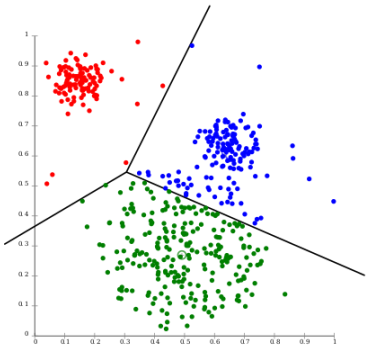
  例: 年齢予測、降水確率の予測、気温の予測
- **分類問題 (classification)**
  - $y \in \{C_1, C_2, \cdots, C_K\}$ (ラベル集合)

  例: 文書分類(政治、経済、スポーツ等)
- **構造予測 (structured prediction)**
  - $y \in G$ (グラフ集合)

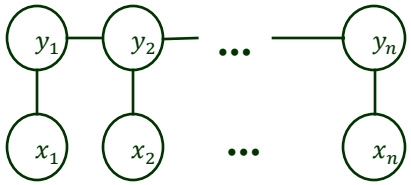
回帰

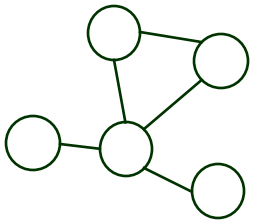


分類



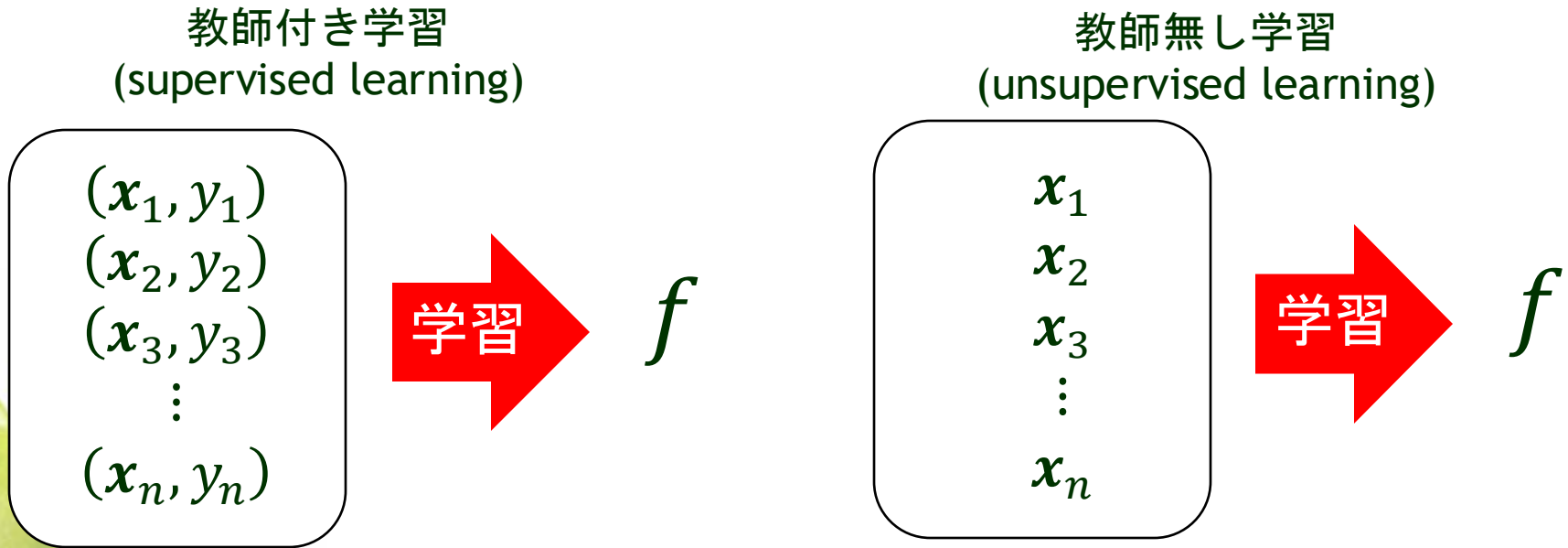
構造予測





教師付き学習と教師無し学習

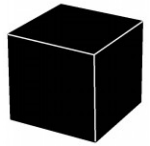
- **教師付き学習**
  - 入出力ペア$(\boldsymbol{x}, y)$の集まり(データ)から関数fを予測
- **教師無し学習**
  - 入力xの集まり(データ)から関数fを予測



機械学習のあれこれ

- 10年前、ニューラルネットワークはまったく見込みのない技術だった
- 機械学習は数学的に説明がつくエレガントな体系だったが、ニューラルネットワークは数学的に何をやっているのかよくわからない。

NN = ブラックボックス?

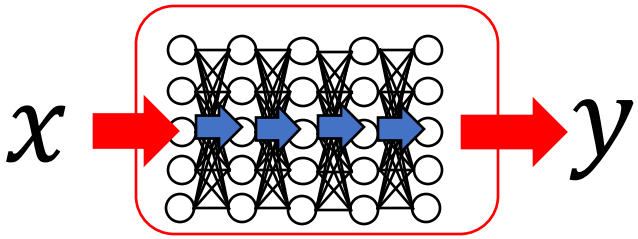


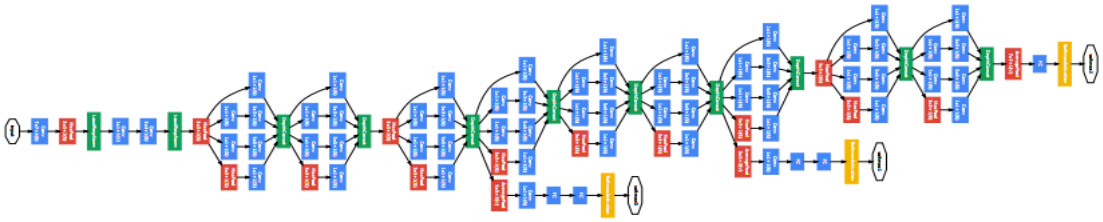
- 何か役に立つのか？
  - 10年前は顔認識ぐらいが唯一のアプリ？
  - 精度100%には(事実上)ならない。いつかかならず判断を間違える。(つまり、クリティカルな仕事には使えない。)

深層学習 (ディープラーニング)

深層学習

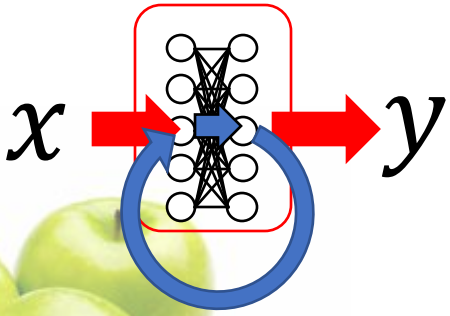
- 関数：重み付き線形和と非線形変換を多層化したもの
  - フィードフォワードニューラルネットワーク

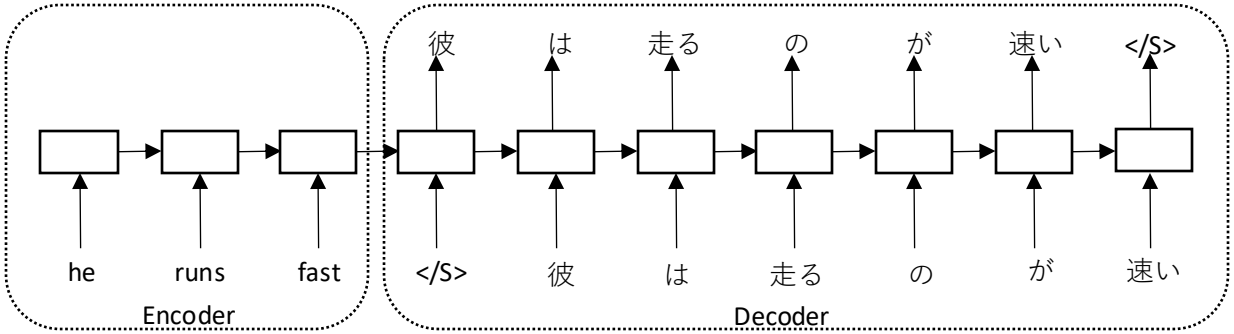




GoogLeNet (22層)　※画像分類に有効なネットワーク

  - リカレントニューラルネットワーク



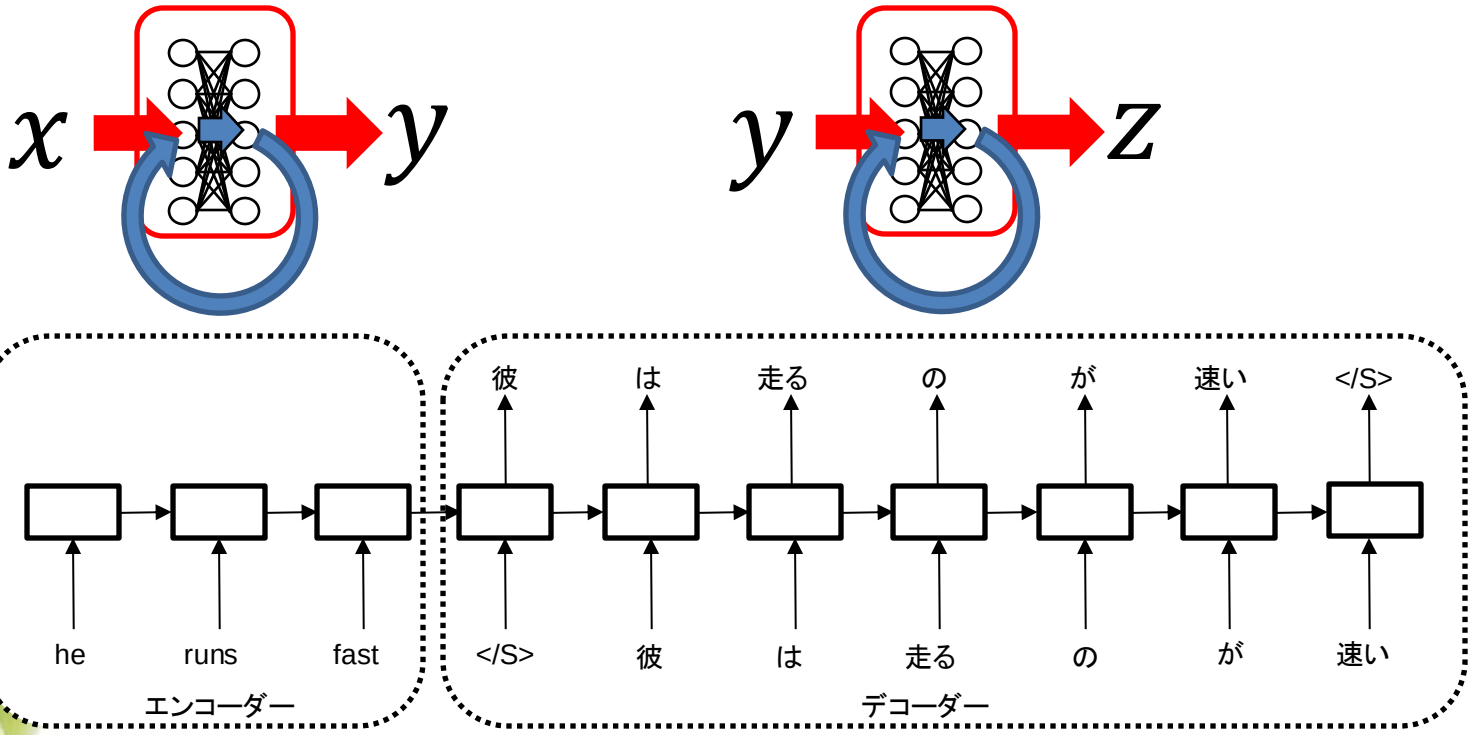


Encoder-Decoderによる機械翻訳

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich (2014) Going Deeper with Convolutions, CVPR 2015

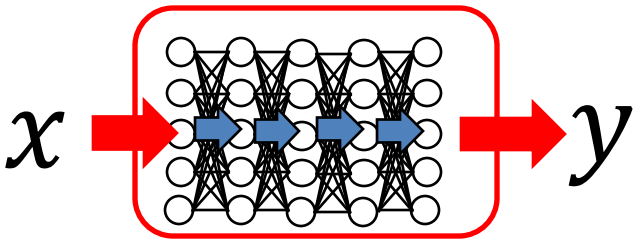
ニューラル機械翻訳

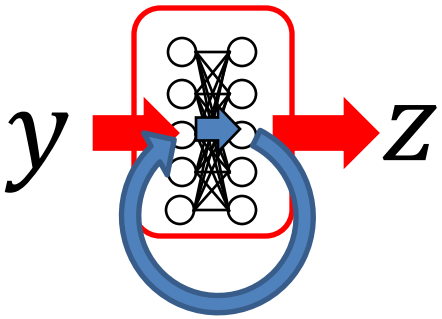
- **2つのリカレントニューラルネットワークを用いた機械翻訳**
  - エンコーダー: 入力文(英語)を中間表現(数百次元のベクトル)に変換
  - デコーダー: 中間表現から出力文(日本語)に変換

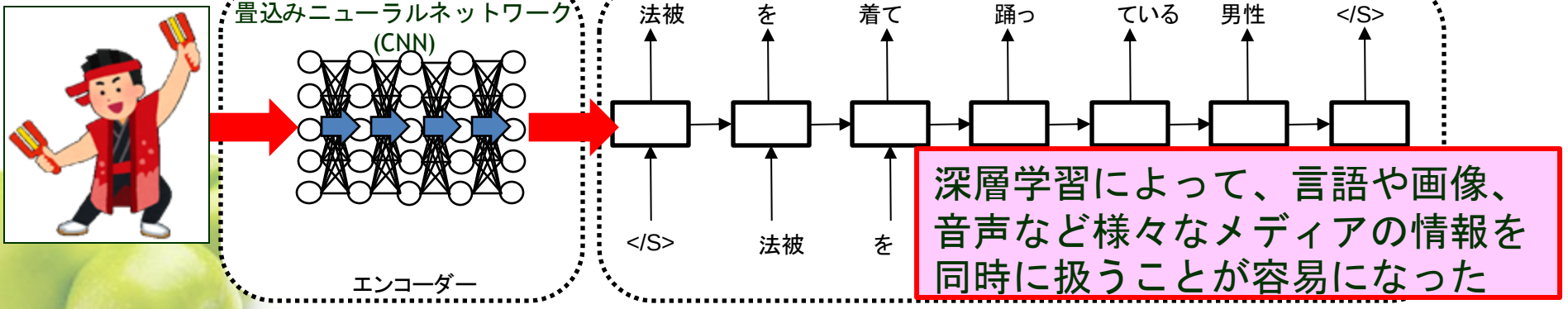


キャプション生成(説明文生成)

- **畳込みニューラルネットワーク(CNN)とリカレントニューラルネットワーク(RNN)を用いたキャプション生成**
  - エンコーダー: 画像を中間表現(数百次元のベクトル)に変換
  - デコーダー: 中間表現から出力文(日本語)に変換

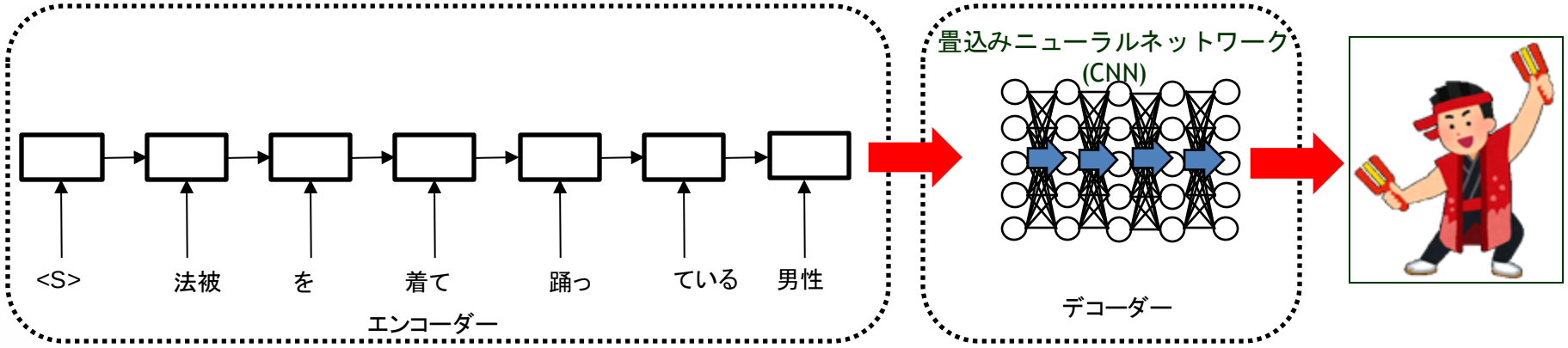






テキストからの画像生成 (text-to-image)

- **テキストエンコーダーと画像生成デコーダーを組み合わせれば、文書から画像を生成可能**
  - エンコーダー: テキストを中間表現(数百次元のベクトル)に変換
  - デコーダー: 中間表現から画像に変換



テキストからの画像生成 (text-to-image)

- Stable Diffusion

  https://stablediffusionweb.com/

プロンプト (prompt): a monkey eating a banana

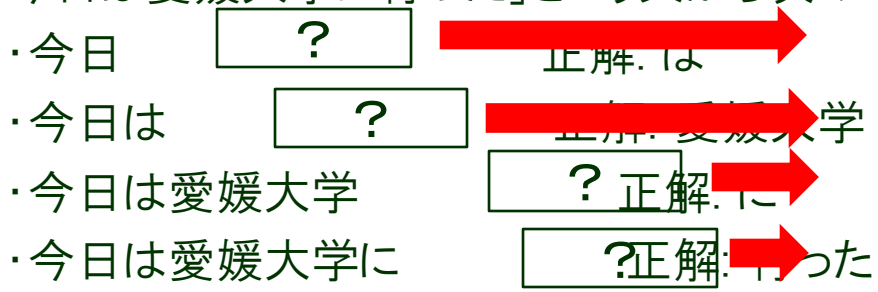
プロンプト: a monkey eating a banana in gogh style

プロンプト: Illustration of a beautiful girl in white dress sitting by the window, beautifull face, by Alphonse Mucha,body tilted

事前学習と言語モデル

- **言語モデル**
  - 次の単語やマスクされた単語を予測するモデル

    例:「今日は愛媛大学に行った」という文から次の単語の予測問題を作る

  - 自己教師付き学習、事前学習
    - 人手で正解ラベルを付ける必要はなく、ただのテキストがあればいくらでも単語予測問題を作ることができる
    - ウェブテキストなど大量のテキストから大規模言語モデルを学習
  - 最終的に解くべきタスクにおいて教師付き学習で再学習(ファインチューニング)
  - word2vec, ELMo, GloVe, Skip-thought, BERT, GPT, BARTなど

ChatGPT

- ChatGPT https://chat.openai.com/
  - 大規模言語モデルGPTをベースに、様々なタスクでファインチューニングし、さらに人手評価に基づく強化学習を行ったモデル
  - 単純なチャットボットに比べ、人間と対話しているかのような優れた応答能力があり、質問を抽象的に理解した上で回答することができている
  - 様々なタスクに対応
    - 会話
    - 機械翻訳
    - 自動要約
    - プログラミング
    - 原稿作成

深層学習が成功した要因

- **深層学習が成功した要因**
  1. 学習データの大規模化
  2. 高精度化
  3. 高速化
  4. End-to-End

学習データの大規模化

- **大昔(20年以上前)**
  - パラメータは手で調整、学習はしない
  - 学習データなし(専門家が作り込んだ少量(数千)の高品質な評価用データのみ)
- **昔(20年前～10年前)**
  - 専門家が作り込んだ少量(数万～数十万規模)の高品質データ
  - 大量(数百万、数億規模)の生データ
- **最近**
  - 素人が作った大量(数百万、数億規模)の中品質データ
    - クラウドソーシング(Amazon Mechanical Turk)など
  - 自動的に収集できる大量のデータ
    - 特許やEU議事録などの翻訳データ
  - 自動的に収集できる大量の生データ(数十テラバイト)
    - 大規模言語モデルの学習に用いられる

深層学習のキーテクノロジー：高精度化

- **解きたい仕事の特徴にあったNN**
  - 画像認識には畳み込みニューラルネットワーク(CNN)
  - 系列データ(テキストなど)には注意型長・短期記憶(Attention-based LSTM)やTransformer
- **事前学習**
  - BERT マスク付き言語モデル
  - GPT 大規模言語モデル
  - オートエンコーダ
- **正則化(Weight Decay, Dropout)**
- **正規化(Batch Normalization)**
- **種々の活性化関数(ReLU, Maxout)**
- **Residual Net**
- **アンサンブル**
- **潜在変数モデル(VAE, 拡散モデル)**

深層学習のキーテクノロジー：高速化

- **学習アルゴリズム**
  - 計算グラフによる誤差逆伝播法
  - オンライン学習(確率的勾配降下法、モーメンタム、AdaGrad、Adam)
- **GPU (CPU1コアよりも10倍以上速い)**
  - NVIDIA GeForce RTX 4080 16GB (約16万円)
  - NVIDIA A6000 Ada 48GB (約100万円)
  - NVIDIA H100 94GB (約500万円)
- **スーパーコンピューティング**
  - 大量のGPU(数百枚～数万枚)を使って並列分散計算

深層学習の特長：End-to-End

- **深層学習**
  - 解きたい仕事の入力（end）と出力（end）だけをNNに与えて全てまとめて学習する
    - 特徴もデータから自動的に学習
    - 複数のシステムをつなぐパイプライン処理を（あまり）しなくてよい

    例：日本語 → ニューラル機械翻訳 → 英語文

  - 開発がものすごく楽になった
    - 今まで敷居の高かった領域の研究がやりやすくなった
  - 専門家の技術が不要になりつつある

まとめ：AIの現状

- **機械学習、深層学習が現在のAIブームを牽引**
- **機械学習**
  - データから関数を学習する
  - 入力と出力の両方が揃った教師データが必要
- **深層学習**
  - 機械学習の一種、多層ニューラルネットワーク
  - 高精度
  - End-to-End

Pytorchを用いた深層学習

DEEP LEARNING WITH PYTORCH

環境設定

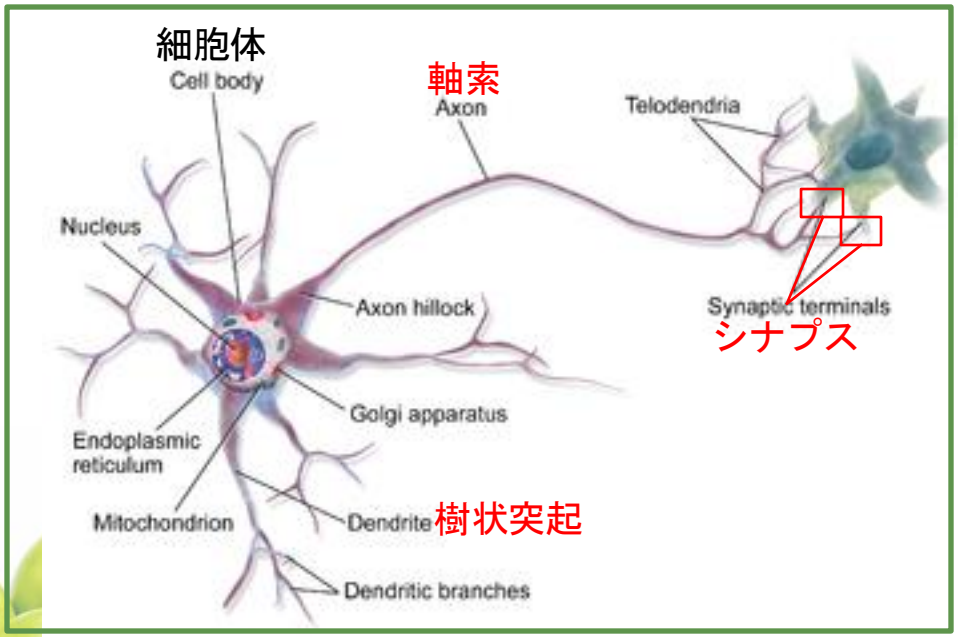
- **次のいずれかの環境設定を行いましょう（上から順に易しい設定となっています）**
  - 環境設定 Google Colaboratory（クラウド環境）Googleアカウントがあればすぐ
  - 環境設定 Anaconda Windows/Linux/MacOS（PC環境）
  - 環境設定 VirtualBox+Ubuntu（サーバー環境およびエッジ環境）
  - 環境設定 Jetson Nano編（エッジ環境）

深層学習 (1)
-ニューラルネットワークの仕組み、活性化関数、推論 -

単純パーセプトロン (1)

- ニューラルネットワーク（深層学習）の起源
- 人間の脳の神経細胞（ニューロン）の仕組みを模したモデル

ニューロン

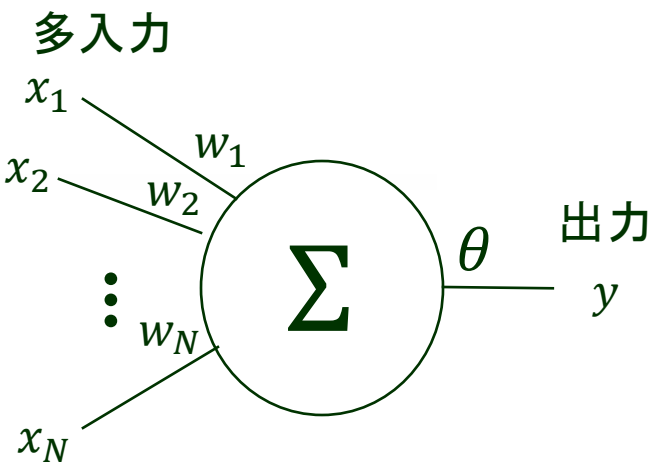


動作：

- シナプスを介して他のニューロンから電気信号を受け取る/他のニューロンへ電気信号を伝える
- 電気信号を受け取ったら電位が変化する
- 電位差が閾値を超えると出力信号を発生する（発火する）

単純パーセプトロン (2)

- ニューロンを模したもの
- 入力信号に対する固有の重みと閾値を持つ
- 動作
  - 多入力を受け付ける
  - 入力の重み付き和を計算する
  - 重み付き和が閾値を超える「1」、そうでなければ「0」を出力する



入力： $x_i \in \{0,1\}\ (i = 1 \cdots N)$

重み： $w_i \in R (i = 1 \cdots N)$

各入力の重要度をコントロール

閾値：$\theta \in R$

発火のしやすさをコントロール

出力： $y = \begin{cases} 1 & if\ \sum_{i=1}^{N} w_i x_i > \theta \\ 0 & otherwise \end{cases}$

パラメータは
w_i と θ

バイアスの導入

入力： $x_i \in \{0,1\}\ (i = 1 \cdots N)$

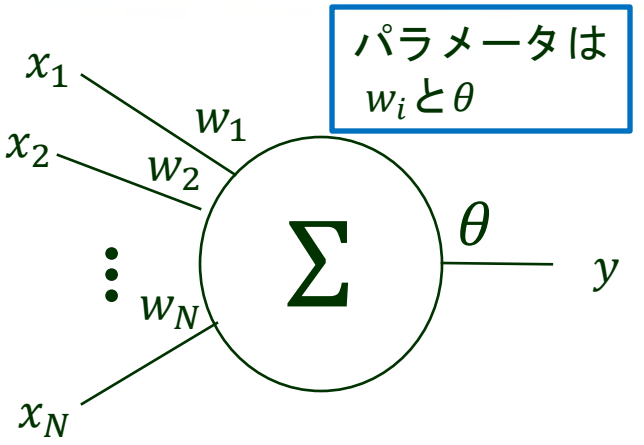
重み： $w_i \in R (i = 1 \cdots N)$

各入力の重要度をコントロール

閾値：$\theta \in R$

発火のしやすさをコントロール

出力： $y = \begin{cases} 1 & if\ \sum_{i=1}^{N} w_i x_i > \theta \\ 0 & otherwise \end{cases}$



$-\theta$をbと考える

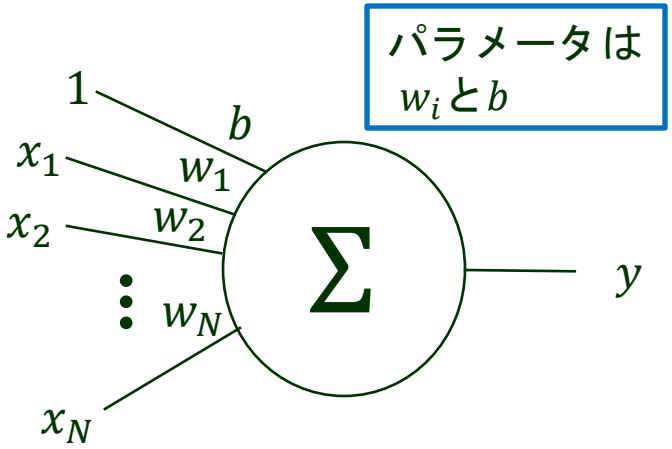
重み： $w_i \in R (i = 1 \cdots N)$

各入力の重要度をコントロール

バイアス：$b \in R$

発火のしやすさをコントロール

出力： $y = \begin{cases} 1 & if\ \sum_{i=1}^{N} w_i x_i + b > 0 \\ 0 & otherwise \end{cases}$



(w_1, w_2, θ)=(1.0, 1.0, 0.5)のパーセプトロンと
(w_1, w_2, b)=(1.0, 1.0, -0.5)のパーセプトロンは同じもの

単純パーセプトロンの学習

教師データを満たす
パラメータ（w_1、w_2、b）を設定する

教師データ

データ	x_1	x_2	y
A	0	0	0
B	1	0	0
C	0	1	0
D	1	1	1

例えば、w_1=1, w_2 = 1とすると…

重み付き和（データA）→ 0
重み付き和（データB）→ 1
重み付き和（データC）→ 1
重み付き和（データD）→ 2

$w_1x_1 + w_2x_2$

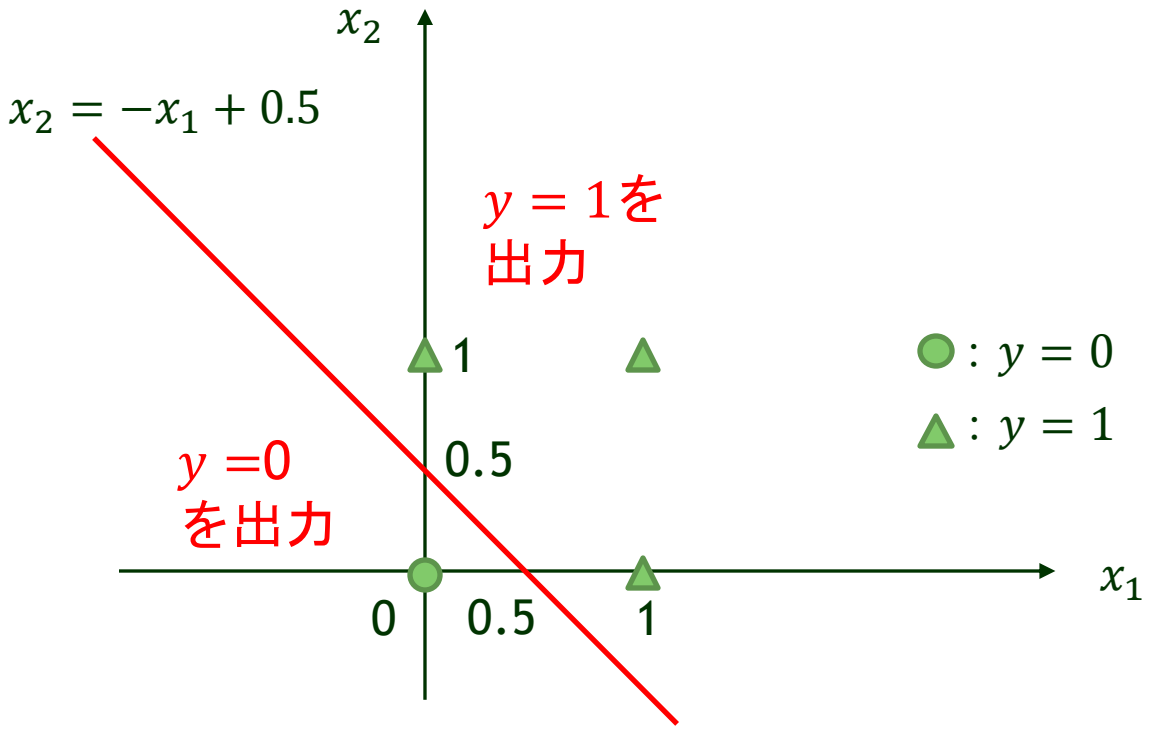
データDの時のみ発火させるように
bを決める
例えばb を-1.5にしてやればよい

(w_1, w_2, b)=(1.0, 1.0, -1.5)で
教師データの入出力を再現できる

単純パーセプトロンの動作の視覚的解釈(推論)

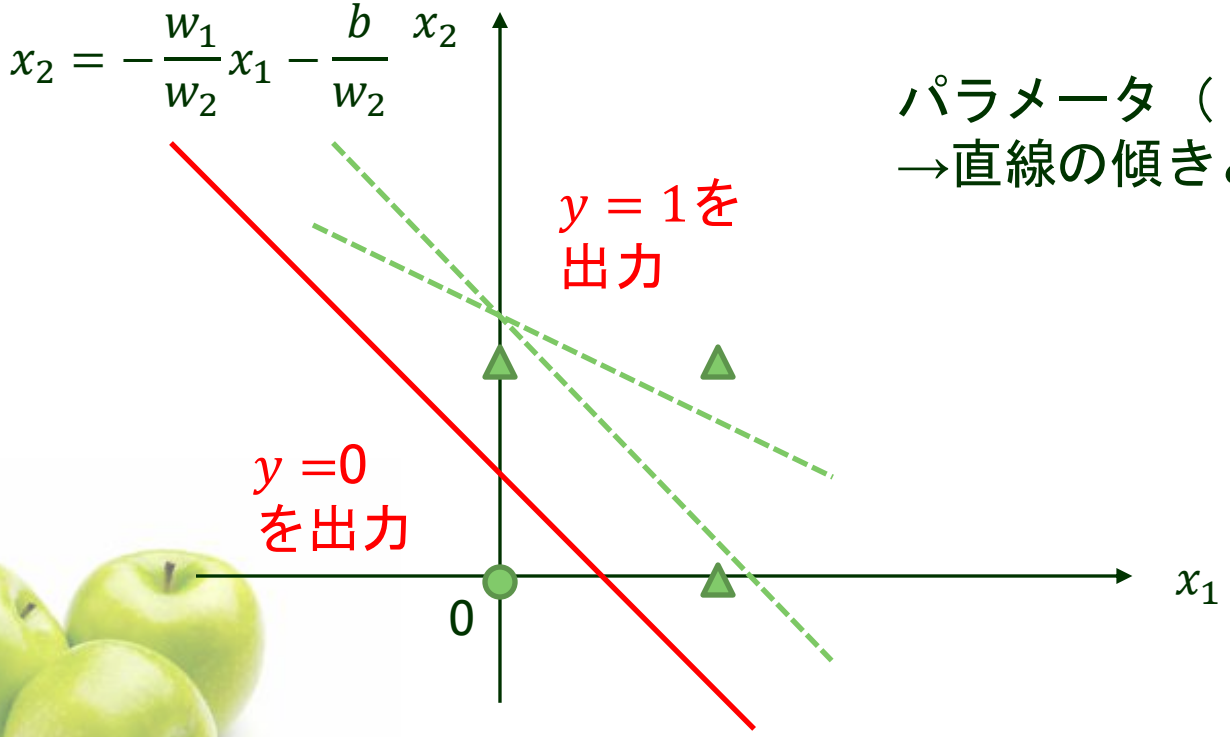
(w_1, w_2, b)=(1.0, 1.0, -0.5)の時

$$y = \begin{cases} 1 & (x_1 + x_2 - 0.5 > 0) \\ 0 & (x_1 + x_2 - 0.5 \leq 0) \end{cases} \Rightarrow y = \begin{cases} 1 & (x_2 > -x_1 + 0.5) \\ 0 & (x_2 \leq -x_1 + 0.5) \end{cases}$$



単純パーセプトロンの動作の視覚的解釈（学習）

$$y = \begin{cases} 1 & (w_1 x_1 + w_2 x_2 + b > 0) \\ 0 & (w_1 x_1 + w_2 x_2 + b \leq 0) \end{cases} \Rightarrow y = \begin{cases} 1 & (x_2 > -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}) \\ 0 & (x_2 \leq -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}) \end{cases}$$



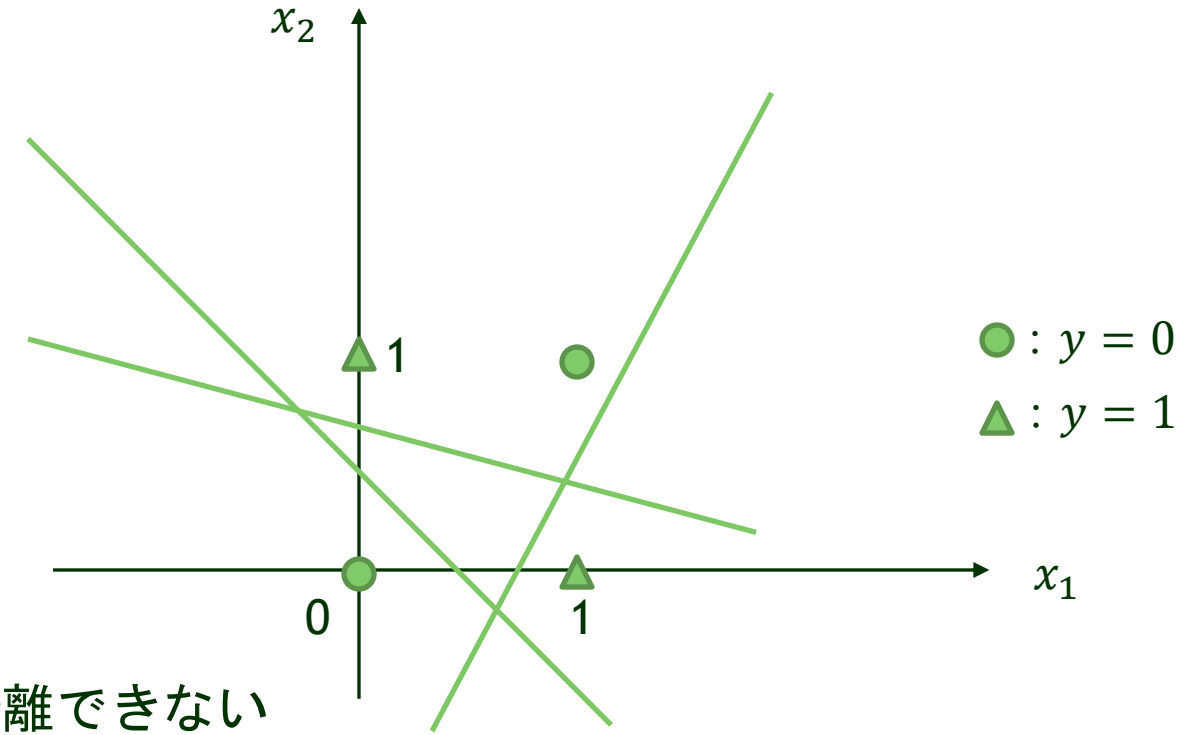
パラメータ（w_1、w_2、b）の調整
→直線の傾きと切片の調整

● : $y = 0$
▲ : $y = 1$

単純パーセプトロンの限界

教師データ

x_1	x_2	y
0	0	0
1	0	1
0	1	1
1	1	0



● と▲は直線で分離できない

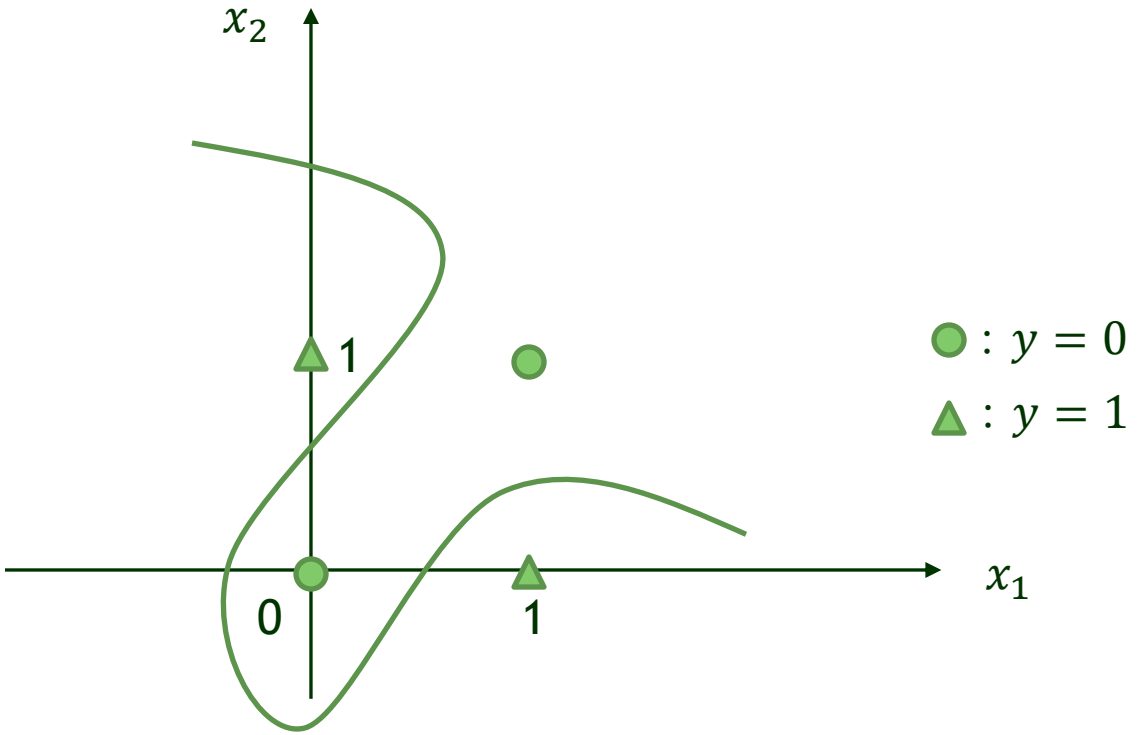
→ 単純パーセプトロンではどんなw_1、 w_2、bを設定しても再現できない

単純パーセプトロンは線形分離可能な問題しかとけない

ちなみに…

教師データ

x_1	x_2	y
0	0	0
1	0	1
0	1	1
1	1	0

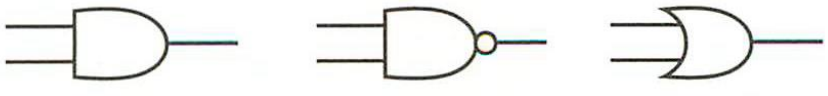


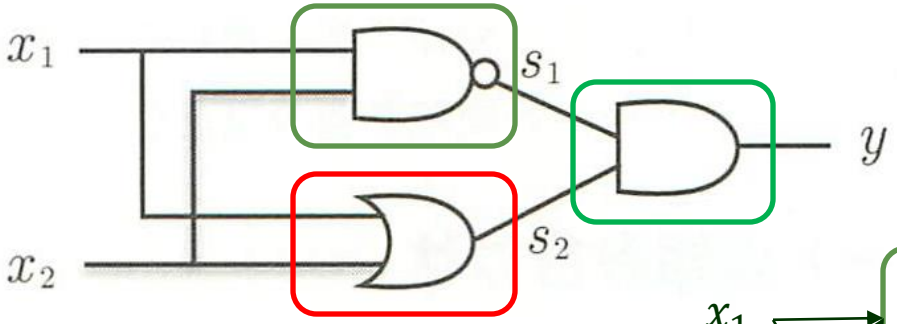
● と ▲ は曲線であれば分離できる

実は、
単純パーセプトロンでも層を重ねると非線形な分離が可能になる

単純パーセプトロンの多層化によるXORゲートの実現

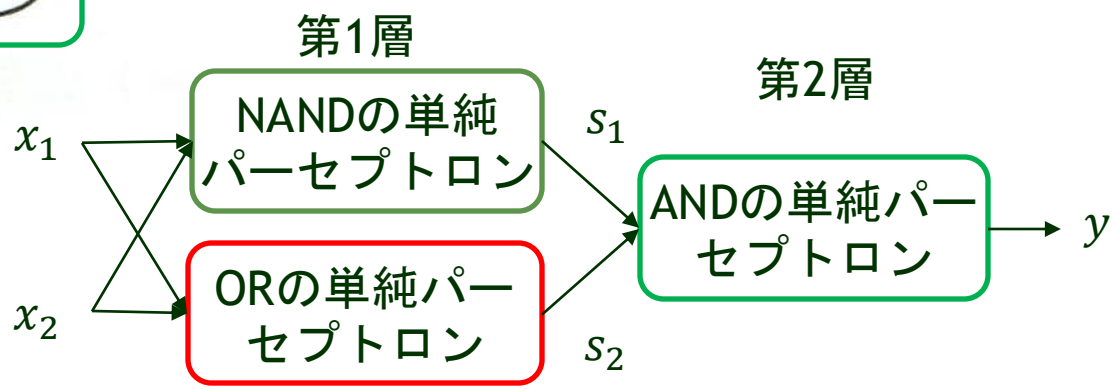
- XORゲートはAND、NAND、ORゲートの組み合わせで実現できる





x_1	x_2	s_1	s_2	y
0	0	1	0	0
1	0	1	1	1
0	1	1	1	1
1	1	0	1	0

AND　NAND　OR



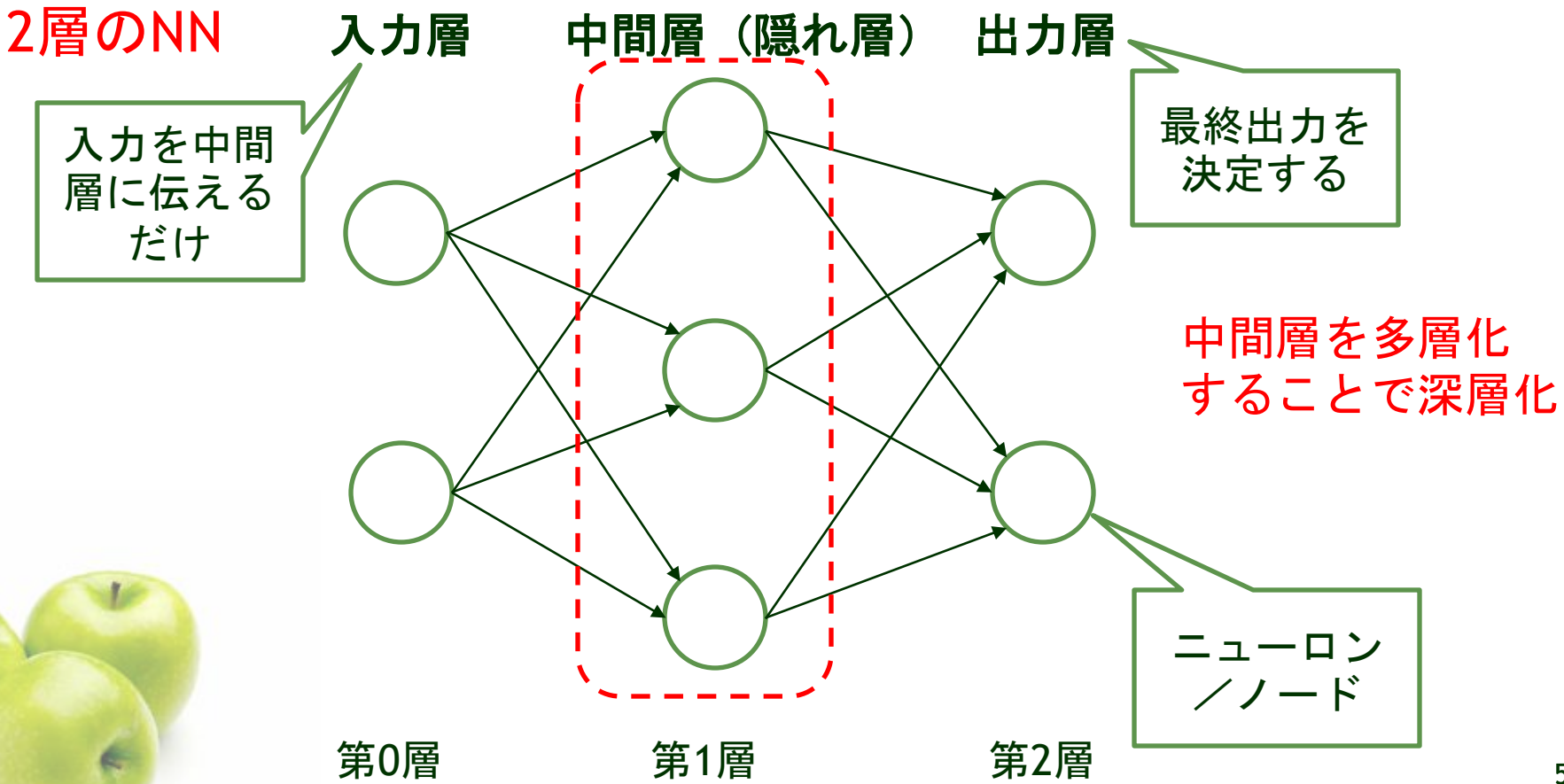
2層のパーセプトロン（多層パーセプトロン）

ニューラルネットワーク（NN）

- **パーセプトロンを発展させたもの**
- **大きな違いは、**
  - 中間層がある（多層）
  - 活性化関数が導入されている

標準的なNNの構造

- 順伝播型(フィードフォワード)NN:情報が入力側から出力側に一方向に伝わるNN



活性化関数の導入

- **パーセプトロン**

入力： $x_i \in \{0,1\}\,(i = 1 \cdots N)$
パラメータ： $w_i, b \in R (i = 1 \cdots N)$
出力： $y = \begin{cases} 1 & if \ \sum_{i=1}^{N} w_i x_i + b > 0 \\ 0 & otherwise \end{cases}$

$$a = \sum_{i=1}^{N} w_i x_i + b$$
$$y = h(a)$$
$$h(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$
ステップ関数

入力信号の総和を出力信号に変換する関数$h(x)$を
活性化関数と呼ぶ

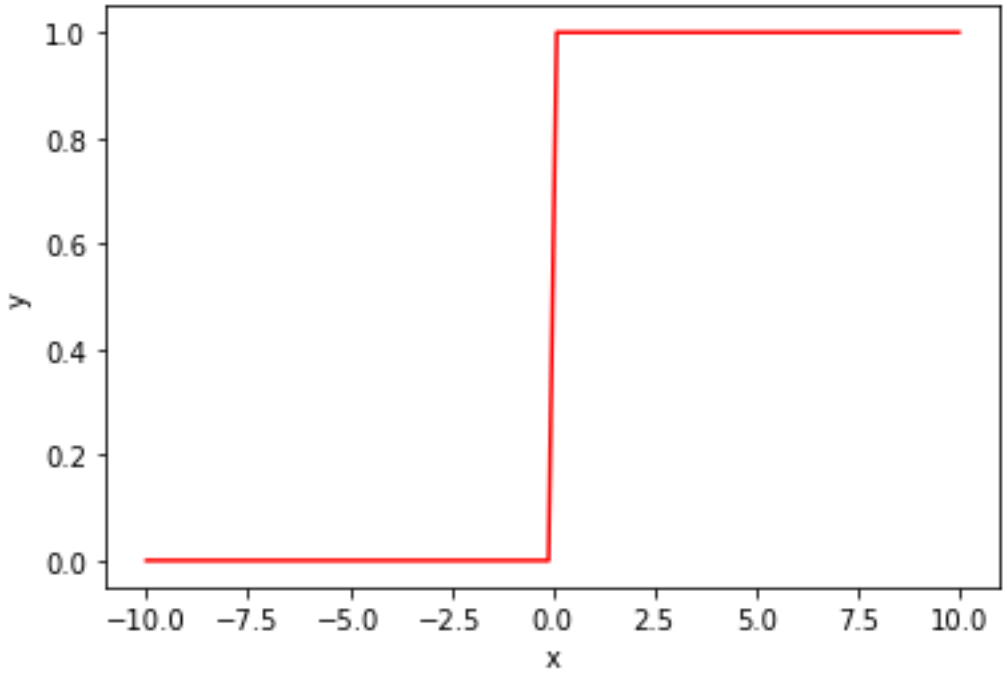
入力信号の総和がどのように発火（活性化）するかを決定する役割

NNの各ノードでは活性化関数としてステップ関数以外の関数も使う！

活性化関数の例

- ステップ関数(パーセプトロンの活性化関数)

$$h(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

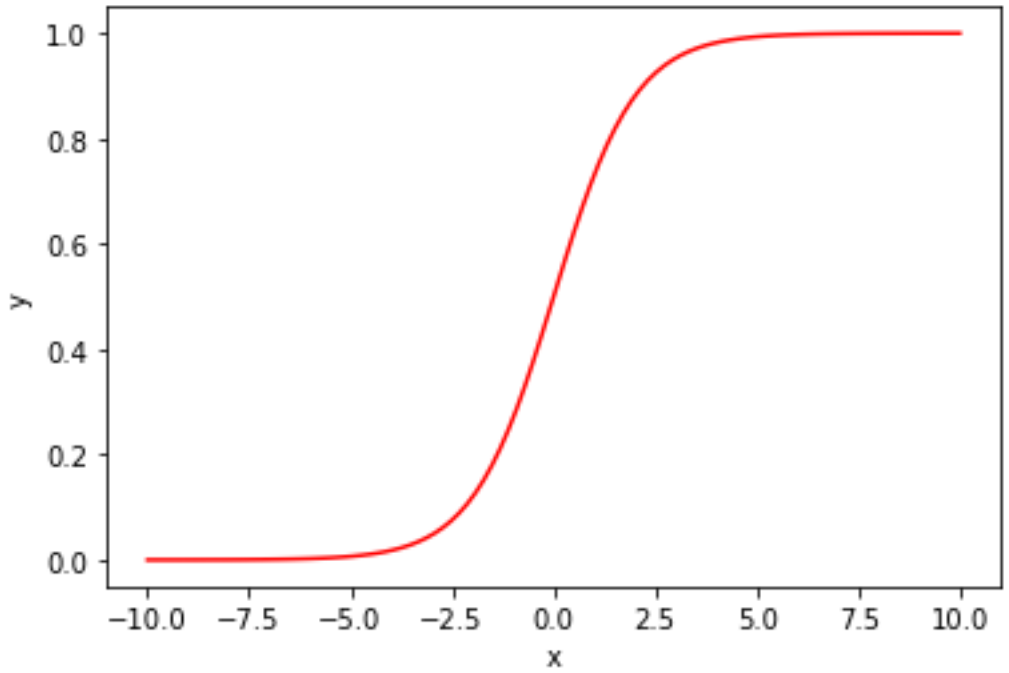


活性化関数の例

- シグモイド関数(ロジスティック関数とも呼ばれる)

$$h(x) = \frac{1}{1 + \exp(-x)}$$

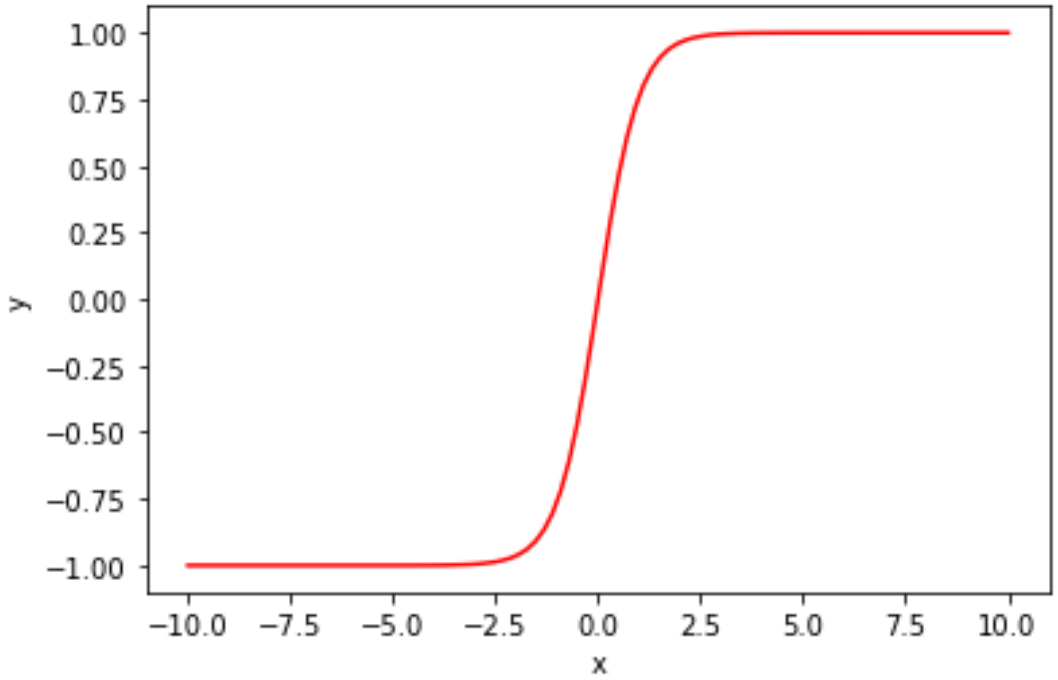
e^{-x}



活性化関数の例

- **ハイパボリックタンジェント（tanh）**
  - −1から1の間の値をとる

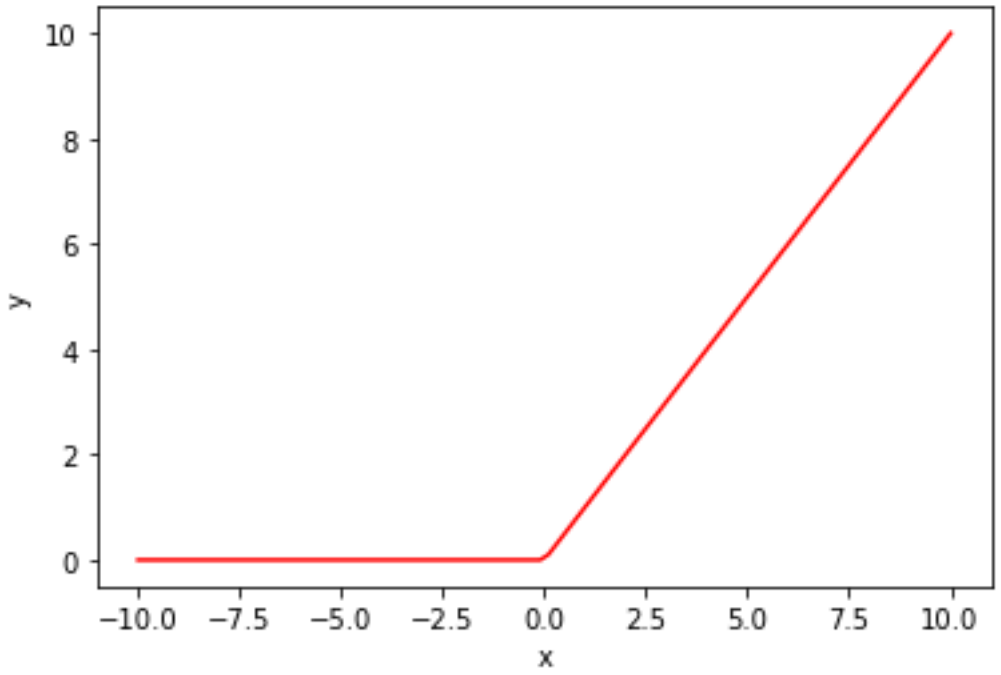
$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$



活性化関数の例

- ReLU関数

$$h(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$



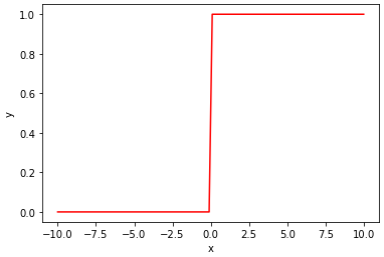
隠れ層の活性化関数の共通点

- **非線形関数**

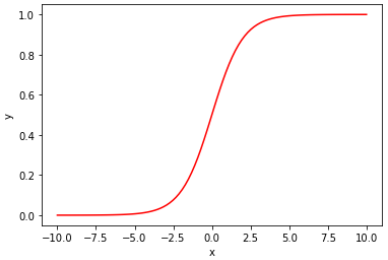
NNの隠れ層は非線形関数を活性化関数として用いないと深層化する意味がない！

仮に、線形関数$h(x) = cx$を活性化関数とすると、3層隠れ層を重ねた場合、その出力は、$y(x) = h\left(h\left(h(x)\right)\right) = c^3x$となる。
これは、$h(x) = ax$（ただし、$a = c^3$）を活性化関数とした1層の層と同等で、多層にするメリットなし

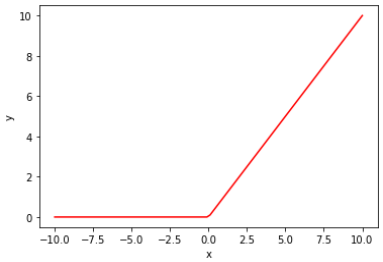
ステップ関数



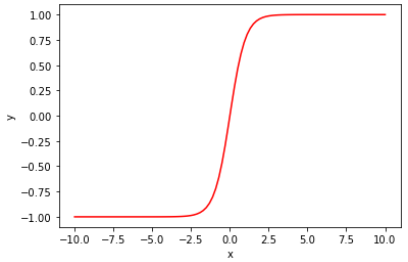
シグモイド関数



ReLU関数



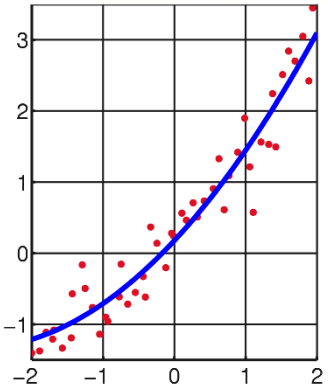
ハイパボリックタンジェント



出力層の活性化関数

関数: $y = f(\boldsymbol{x})$

回帰

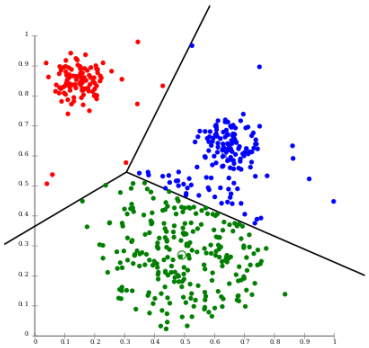


- **回帰問題 (regression)**
  - $y \in R$ (実数)

  例: 年齢予測、降水確率の予測、気温の予測
- **分類問題 (classification)**
  - $y \in \{C_1, C_2, \cdots, C_K\}$ (ラベル集合)

  例: 文書分類(政治、経済、スポーツ等)

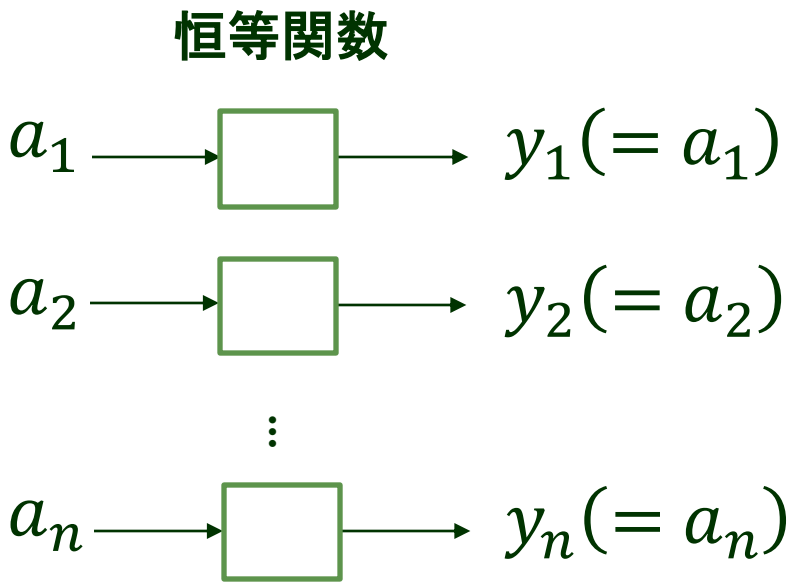
分類



回帰問題 → 恒等関数
分類問題 → ソフトマックス関数

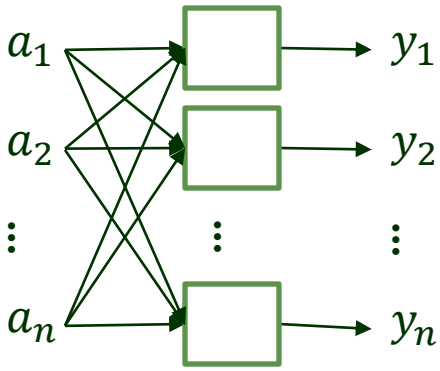
回帰問題の活性化関数

- **恒等関数**
  - 入力をそのまま出力



分類問題の活性化関数

- **ソフトマックス関数**
  - 自分以外のノードの出力も使って正規化



$$y_k = \frac{\exp(a_k)}{\sum_{i=1}^{n} \exp(a_i)}$$

$$0 \leq y_k \leq 1 \ (1 \leq k \leq n)$$

$$\sum_{k=1}^{n} y_k = 1$$

出力層のノード数を「分類クラス数」にする
→ 各ノードの出力はそのクラスの分類確率として解釈可能
→ 分類確率の最も高いクラスに分類

例：犬猫鳥判別機

入力x		a	正のスコア		ソフトマックス	
	$a_{犬}$ →	9	→ e^{9}	→	$e^{9}/(e^{9}+e^{2}+e^{-1})$	総和は1
	$a_{猫}$ →	2	→ e^{2}	→	$e^{2}/(e^{9}+e^{2}+e^{-1})$	
	$a_{鳥}$ →	-1	→ e^{-1}	→	$e^{-1}/(e^{9}+e^{2}+e^{-1})$	

本講義でのノードの書き方

入力： $x_i \in \{0,1\}\ (i = 1 \cdots N)$

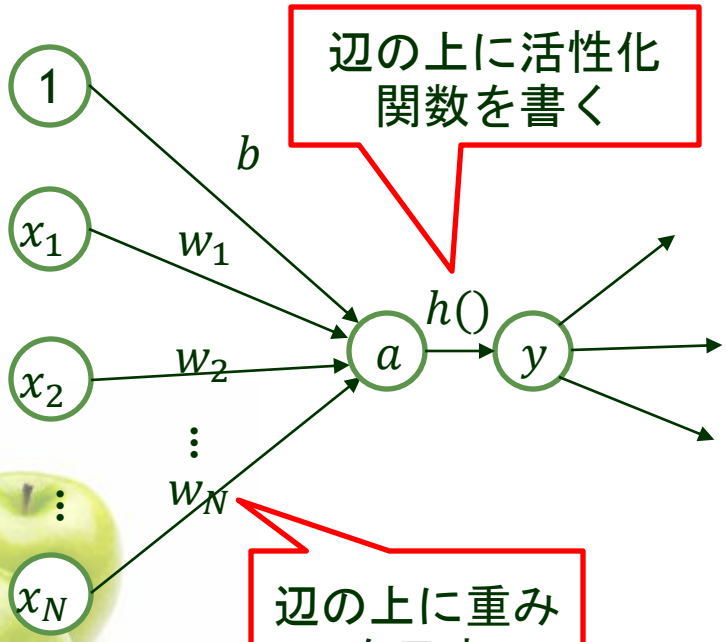
パラメータ： $w_i, b \in R (i = 1 \cdots N)$

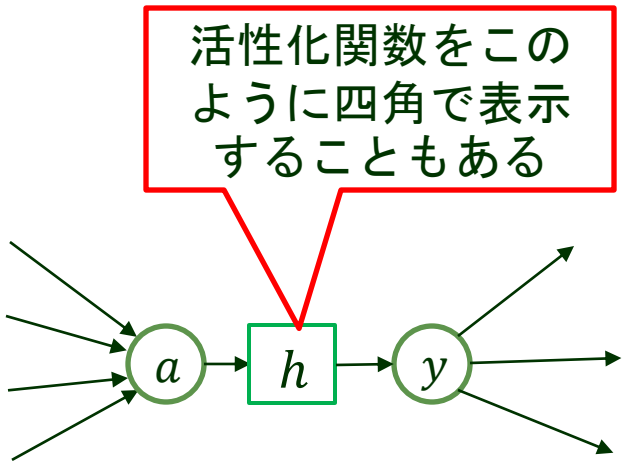
出力：

$$a = \sum_{i=1}^{N} w_i x_i + b$$

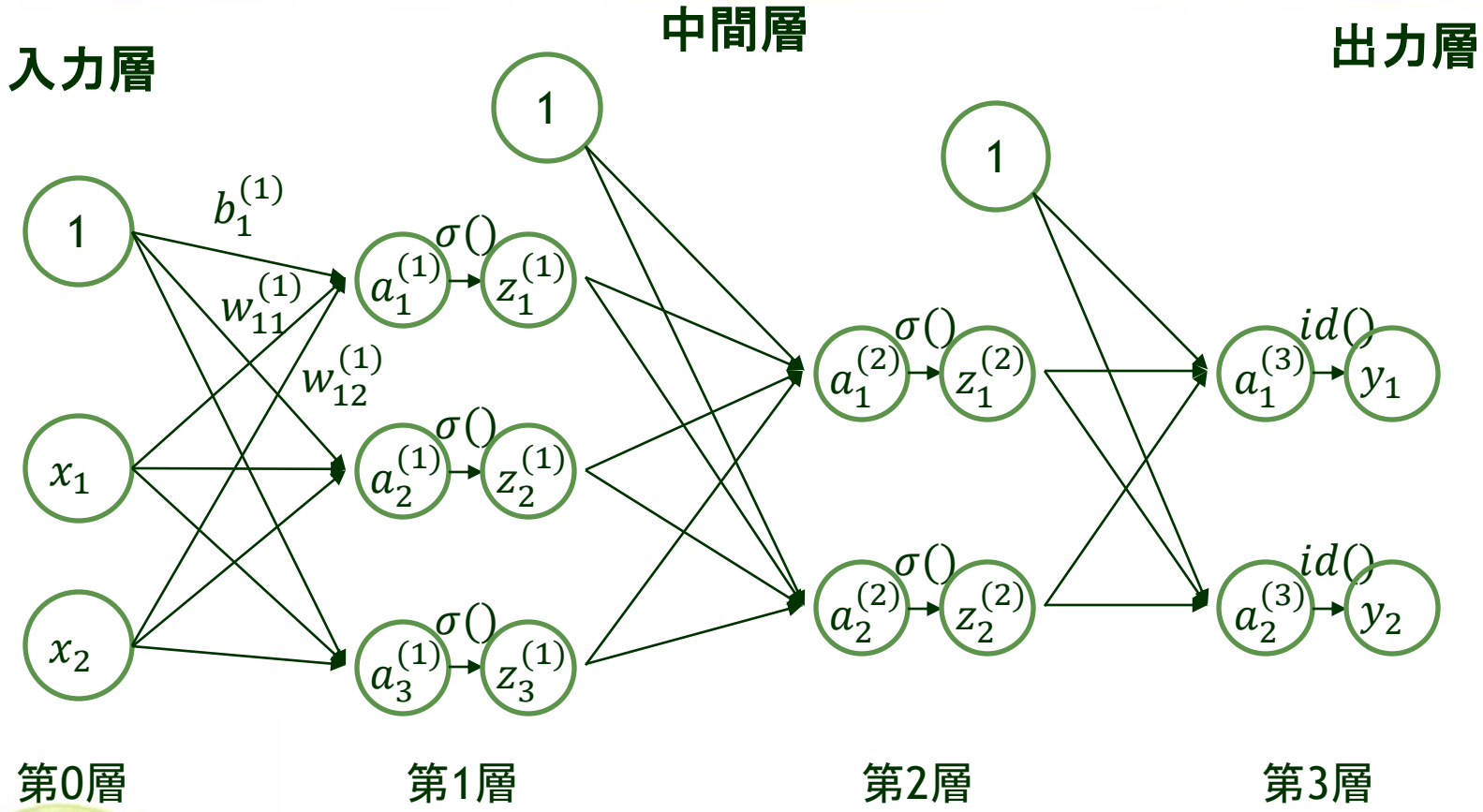
$$y = h(a)$$

※パーセプトロンの場合、ステップ関数





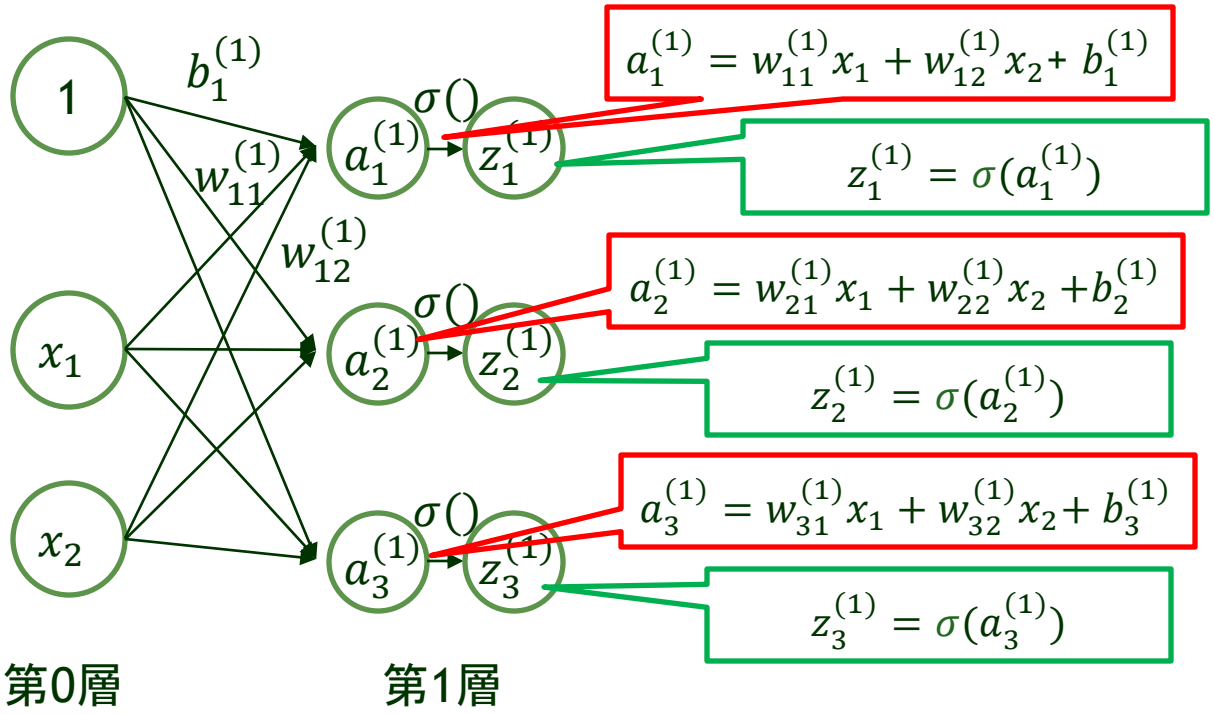
3層NNの推論



$\sigma()$：シグモイド関数
$id()$：恒等関数

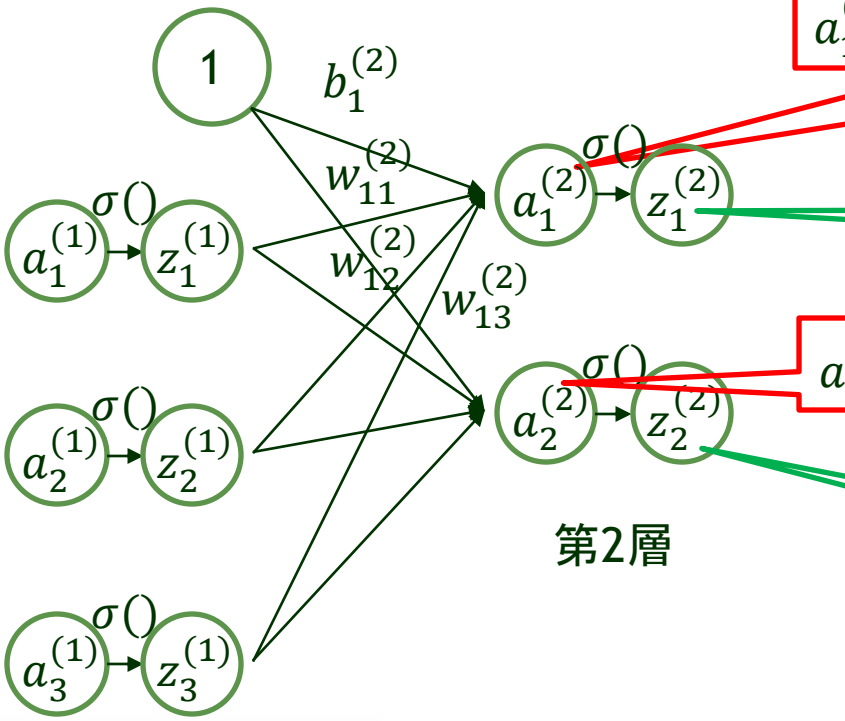
信号は低層から順に伝わる
→低層の演算から順に行う

1層目の動作



$\sigma()$：シグモイド関数

2層目の動作



$$a_1^{(2)} = w_{11}^{(2)} z_1^{(1)} + w_{12}^{(2)} z_2^{(1)} + w_{13}^{(2)} z_3^{(1)} + b_1^{(2)}$$

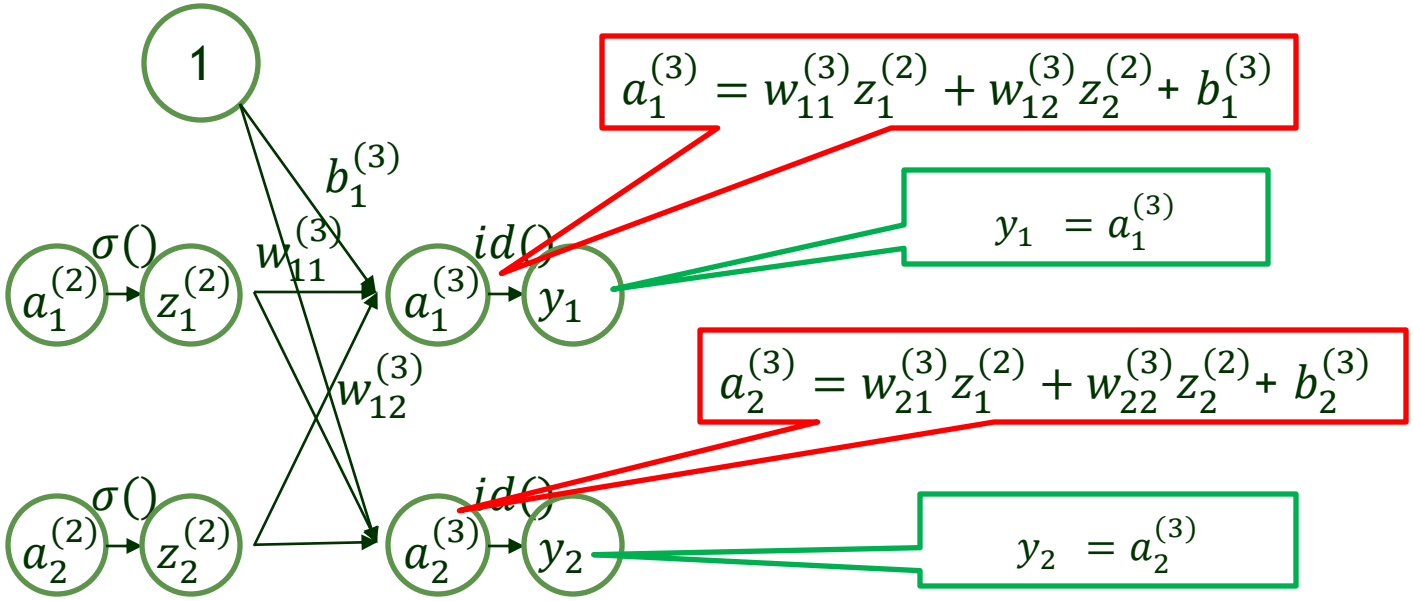
$$z_1^{(2)} = \sigma(a_1^{(2)})$$

$$a_2^{(2)} = w_{21}^{(2)} z_1^{(1)} + w_{22}^{(2)} z_2^{(1)} + w_{23}^{(2)} z_3^{(1)} + b_2^{(2)}$$

$$z_2^{(2)} = \sigma(a_2^{(2)})$$

$\sigma()$：シグモイド関数

3層目（出力層）の動作



$\sigma()$：シグモイド関数
$id()$：恒等関数

多層パーセプトロンの行列計算

- **多層パーセプトロンにおける各層の計算は行列計算**
  - 線形変換(アフィン層、全結合層)+活性化関数
  - i層(n次元)からj層(m次元)への計算

$$\boldsymbol{a}^{(j)} = \boldsymbol{W}^{(j)}\boldsymbol{z}^{(i)} + \boldsymbol{b}^{(j)}$$
$$\boldsymbol{z}^{(j)} = h\left(\boldsymbol{a}^{(j)}\right)$$

まとめて書くと、

$$\boldsymbol{z}^{(j)} = h\left(\boldsymbol{W}^{(j)}\boldsymbol{z}^{(i)} + \boldsymbol{b}^{(j)}\right)$$

$$\begin{pmatrix} a_1^{(j)} \\ a_2^{(j)} \\ \vdots \\ a_m^{(j)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(j)} & w_{12}^{(j)} & \cdots & w_{1n}^{(j)} \\ w_{21}^{(j)} & w_{22}^{(j)} & \cdots & w_{2n}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1}^{(j)} & w_{m2}^{(j)} & \cdots & w_{mn}^{(j)} \end{pmatrix} \begin{pmatrix} z_1^{(i)} \\ z_2^{(i)} \\ \vdots \\ z_n^{(i)} \end{pmatrix} + \begin{pmatrix} b_1^{(j)} \\ b_2^{(j)} \\ \vdots \\ b_m^{(j)} \end{pmatrix}$$

$$\begin{pmatrix} z_1^{(j)} \\ z_2^{(j)} \\ \vdots \\ z_m^{(j)} \end{pmatrix} = h\begin{pmatrix} a_1^{(j)} \\ a_2^{(j)} \\ \vdots \\ a_m^{(j)} \end{pmatrix}$$

多層パーセプトロンの行列計算

- **ベクトルに対する活性化関数の適用**
  - ステップ関数、シグモイド関数、ReLU関数の場合
    - ベクトルの各要素に関数を適用

$$h\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} h(a_1) \\ h(a_2) \\ \vdots \\ h(a_m) \end{pmatrix}$$

  - ソフトマックス関数の場合

$$softmax\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \frac{\exp(a_1)}{\sum_{i=1}^{m}\exp(a_i)} \\ \frac{\exp(a_2)}{\sum_{i=1}^{m}\exp(a_i)} \\ \vdots \\ \frac{\exp(a_m)}{\sum_{i=1}^{m}\exp(a_i)} \end{pmatrix}$$

まとめ

- **NNの推論**
  - NNは入力層、中間層、出力層の多層構造
  - 信号は入力層、中間層、出力層へと階層的に伝わる
  - 中間層、出力層には活性化関数を用いる

実習編: PYTHON

Python演習

- Python https://www.python.jp/
  - 最もよく使われているプログラミング言語の一つ
    - 多くのプログラミング言語ランキングで1位～2位。
  - ライブラリが充実しており、データ処理、データ解析、機械学習のためによく使われる。特に深層学習のプログラムはPythonで書くことが多い。
  - ドキュメント https://docs.python.jp/3/

※Pythonには2系と3系があり、構文や同じ関数でも処理が違う場合があるので注意。本講義ではPython 3系を想定。

Python環境設定

- **深層学習(ディープラーニング)を行うには？**
  - クラウド環境を使う方法
    - Google Colaboratory (略称 Colab)
      - https://colab.research.google.com/
      - Googleが提供するクラウド環境の深層学習プラットフォーム
      - Googleアカウントがあれば無料で制限付きながら使える
      - Googleアカウントをもっていればおすすめ
    - Microsoft Azure Machine Learning (通称 Azure)
      - Microsoftアカウント+サブスクライブで使える。
      - 無料で1年間だけ制限付きながら使える。

Python環境設定

- **深層学習(ディープラーニング)を行うには？**
  - 自分が持っているサーバー/PCを使う方法
    - Ubuntuに自力でいろんなライブラリを追加する
      - バージョンの違いでうまくインストールできなかったり動かなかったりで結構大変。
    - Anacondaを使う（オススメ！）
      - Python + 科学技術計算(データサイエンス)ライブラリ
      - データサイエンス向けライブラリが全部入りで最初から入っている
      - 最初から入っているライブラリ
        - NumPy ベクトル行列演算ライブラリ
        - SciPy 数値演算、最適化、信号処理、関数
        - Pandas 統計処理、データ分析、時系列解析
        - scikit-learn 機械学習

Python演習

PYTHONプログラミング

Python 算術演算

- Pythonでの算術演算
  - 足し算 +
  - 引き算 -
  - 掛け算 *
  - 割り算 /
  - 整数の割り算 //
  - 余りの計算 %
  - べき乗 **
- 右のプログラムコードを実行してみよう

```
[1] 1+3
4

[2] 3-1
2

[3] 5*2
10

[4] 2**4
16

[5] 7/5
1.4

[6] 7//5
1

[7] 7%5
2
```

Python 変数

- アルファベットで定義
- 整数型(int)や実数型(float)といった型は宣言せず自動的に決定される
- 変数への代入は「=」
- 変数の型や計算結果の型はtype関数で調べることができる
- 右のプログラムを実行してみよう

```
[8]  x=100

[9]  x

     100

[10] type(x)

     int

[11] y=3.14

[12] type(y)

     float

[13] x*y

     314.0

[14] type(x*y)

     float
```

変数の中身をみることができる

intは整数型という意味

floatは実数型という意味

計算結果にも型がついている

Python 変数

- **次のプログラムを実行してみよう**

```
[15] a=1

[16] a
     1

[17] type(a)
     int

[18] b=1.0

[19] type(b)
     float
```

小数点をつけない数字は整数型と判断される

実数型にするには小数点をつけて入力する

```
[20] c=1.

[21] c
     1.0

[22] type(c)
     float
```

実数型をこのように入力しても良い

ブール型

- **ブール型(bool): TrueかFalseの値**
- **ブール演算**
  - and演算: $x\ and\ y = \begin{cases} True & xとyの両方ともTrueのとき \\ False & それ以外 \end{cases}$
  - or演算: $x\ or\ y = \begin{cases} True & xとyのどちらかもしくは両方がTrueのとき \\ False & それ以外 \end{cases}$
  - not演算: $not\ x = \begin{cases} True & xがFalse \\ False & xがTrue \end{cases}$ (xのTrue/Falseの反転)

ブール型

- コンソールで次のPythonプログラムを書いてみよう

```
[23] hungry=True

[24] sleepy=False

[25] type(hungry)

     bool

[26] not hungry

     False

[27] hungry and sleepy

     False

[28] hungry or sleepy

     True
```

文字列型

- **文字列(string)**
  - ’(シングルクォート)で囲む
  - ”(ダブルクォート)で囲む
  - ”””(ダブルクォート3つ)で囲む
  - どの方法でも同じ文字列が得られる
- **文字列の連結は「+」で実行できる**
- **右のプログラムを実行してみよう**

```
[29] a = "こんにちは"

[30] b = 'こんにちは'

[31] c = """こんにちは"""

[32] a

     'こんにちは'

[33] b

     'こんにちは'

[34] c

     'こんにちは'

[35] a+b+c

     'こんにちはこんにちはこんにちは'
```

文字列の連結

リスト型

- リスト型(list)
  - 複数のデータをまとめる
  - コンマ区切りの値の並びを角括弧で囲む

```
[36] a = [1, 2, 3, 4, 5]

[37] a
     [1, 2, 3, 4, 5]

[38] len(a)
     5

[39] a[0]
     1

[40] a[1]
     2
```

角括弧の中にリストの各要素を並べる。各要素はカンマ(,)で区切る

リストの長さは**len関数**で計算できる

a[n]と書くことでn番目の要素を得ることができる。ただし、リストの最初の要素は0番目と数える

```
[41] a[4]
     5

[42] a[4]=99

[43] a
     [1, 2, 3, 4, 99]
```

a[n]=xと書くことでn番目の要素にxを代入することができる

リスト型

- **リストの連結**
  - +でリストを連結することができる（文字列の連結と似ている）

リストの連結

```
[44] a=[1,2,3]

[45] b=[4,5,6]

[46] a+b

     [1, 2, 3, 4, 5, 6]
```

- **リストの生成**
  - [x]*n でxを要素とする長さnのリストを生成できる

要素1の長さ10のリスト

```
[47] a=[1]*10

[48] a

     [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

辞書型

- **辞書型(dictionary): 一般には連想配列と呼ばれるデータ構造**
  - キーと値のペアをたくさん格納したリスト
  - キーに対応する値をすぐに取り出すことができる
  - 「キー:値」を並べて中括弧で囲むことで生成する

```
[49] height={"sato":180, "yamada":170}

[50] height["sato"]

     180

[51] height["kato"]=175

[52] height

     {'kato': 175, 'sato': 180, 'yamada': 170}
```

”sato”は180
”yamada”は170

d[x]は辞書dにおけるxの値

d[x]=yで、辞書dにおけるxの値をyにする

”kato”:175が追加されている

if文

- **条件分岐**
  - ブロック: 一連の処理の範囲のことをブロックという。
  - CやProcessingなど普通の言語では中括弧{}を用いてブロックを指定する。
    - Processingの例

```
int a = 150;
if( a > 100) {
  background(255,0,0);
  ellipse(50, 50, 50, 50);
}
```

If文の条件式が真であったときのブロック

  - Pythonのブロックはイ
    - Pythonの例

```
[53] a = 150

[54] if a > 100:
       a = 0
       print("1")
     else:
       print("0")

     1
```

If文の条件式が真であったときのブロック

空白4個かタブが一般的

If文の条件式が偽であったときのブロック

if文

- 次のコードを実行してみよう

```
[55] a=150
     if a > 100:
       print("aは100より大きい")
     elif a > 0:
       print("aは0より大きくて100以下です")
     else:
       print("aは0以下です")
```

aの値をいろいろと変えてどのように結果が変わるのかみてみよう

```
aは100より大きい
```

if文の構文

条件式の後にコロン(:)が必要

```
if 条件式1:
        条件式1が真のときの処理
elif 条件式2:
        条件式1が偽で条件式2が真だったときの処理
…
else:
        上記の条件式がすべて偽だったときの処理
```

上から下に向かって実行

if文の条件式

- **条件式には次のようなものが使えます**
  - 等号(等しい) ==
  - 不等号(より大きい) >
  - 不等号(より小さい) <
  - 不等号(以上) >=
  - 不等号(以下) <=
- **また、ブール演算を使って、複雑な条件を書くこともできます**
  - 例

```
[56] x=150
     if 100 < x and x <= 200:
       print("xは100より大きく200以下です")

     xは100より大きく200以下です
```

for文

- Pythonのfor文はリストに対する繰り返し処理

```
[57] words = ["cat", "dog", "lion"]
     a=1
     for w in words:
       print(w+": "+str(a))
       a = a + 1
```

```
cat: 1
dog: 2
lion: 3
```

for文の構文

```
for 変数 in リスト:
        繰り返し処理のブロック
        (変数を参照しながら処理することができる)
```

for文

- Pythonのfor文はリストに対する繰り返し処理

```
[57]  words = ["cat", "dog", "lion"]
      a=1
      for w in words:
          print(w+": "+str(a))
          a = a + 1
```

①　②　③

繰り返し処理ブロック

文字列と数値の足し算はできない。
str関数を用いると数値を文字列に変換できる。

①に対する処理　②に対する処理　③に対する処理

リストの先頭から順に各要素が変数に代入されて繰り返し処理ブロックの処理が行われる

①w = ”cat”として繰り返し処理ブロックを実行
②w = ”dog”として繰り返し処理ブロックを実行
③w = ”lion”として繰り返し処理ブロックを実行

for文とrange関数

- 一定回数繰り返したいときは？→range関数を使う

- range関数
  - range(n)で[0, 1, 2, …, n−1]の(仮想的な)リストを作る
  - range(i, j)で[i, i+1, i+2, …, j−1]の(仮想的な)リストを作る

```
[58]  list(range(20))
```

```
      [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]
```

```
[59]  list(range(10, 20))
```

```
      [10, 11, 12, 13, 14, 15, 16, 17, 18, 19]
```

for文とrange関数

- for文とrange関数を用いた繰り返し処理

```
[60] sum=0
     for i in range(100):
       x = i + 1
       sum = sum + x
     print(sum)
```

[0, 1, 2, …, 99]のリストがあると思えば良い

x = i + 1とすることで、xは1, 2, …, 100の値になる

5050

i = 0 としてx = 1, sum = sum + x
i = 1 としてx = 2, sum = sum + x
i = 2 としてx = 3, sum = sum + x
…
i = 99としてx = 100, sum = sum + x

これは$\sum_{x=1}^{100} x$ =1 + 2 + 3 + … + 100=5050の計算をしている

関数

- lenなど組み込みの関数を用いてきたが自分で新しく関数を定義することができる
- defで関数を定義

```
[61] def add(x, y):
        ans = x + y
        return ans

     print(add(10,30))
```

40

関数定義の構文

```
def 関数名(引数の列):
        関数内での処理
        return 返り値
        (返り値は関数が返す値)
```

Pythonプログラミング

- **対話モードは便利ですが、プログラムを書くには不向きです。**
- **テキストエディタでプログラムを書きましょう！**
  - gedit（初級者向け）ubuntuに最初から入っています
  - visual studio code（上級者向け）
  - emacs（上級者向け）セットアップ編でインストールしました

どれを使ってもかまいません。visual studio codeやemacsを使ったことがない人はgeditを使いましょう。左下のアプリボタンから起動するか、下記のとおり端末から起動することができます。

```
$ gedit &    (geditを起動したいとき)
$ emacs &     (emacsを起動したいとき)
$ code &      (visual studio codeを起動したいとき)
```

※ 後ろに&をつけると、別ウィンドウで起動します

(参考) Visual Studio Code (vscode)のインストール

- Visual Studio Codeのホームページにアクセス
  https://azure.microsoft.com/ja-jp/products/visual-studio-code/
- 今すぐダウンロードを選び、「↓.deb」(ubuntu用)を選択します。
- 「プログラムで開く」を選ぶとインストールが始まります。
- 左下のアプリボタンから「Visual Studio Code」を選択するか、端末から「code」で実行できます。

※Jetson Nanoではこの方法ではインストールできません。(アーキテクチャが違うため)

Python演習

- 次のプログラムを入力して実行してみよう

```
def sign(x):
  if x < 0:
    print("negative")
    return -1
  elif x == 0:
    print("zero")
    return 0
  else:
    print("positive")
    return 1

x = sign(10)
y = sign(0)
z = sign(-5.3)
print(x, y, z)
```

```
positive
zero
negative
1 0 -1
```

このような結果がでていたらうまく動いているということ

Python演習

- **数値のリストを受け取ったとき、受け取ったリストの各要素を10倍にして返す関数を書きましょう。**
  - まず、「ファイル」→「ノートブックを新規作成」を選んで新しいファイルを作成しましょう。
  - 続いて、新しく作ったファイル名を変更しましょう。（例えば、ai5_tentimes.ipynbなど）

```
def tentimes(numlist):
  #ここにプログラムを書く


print(tentimes([1,2,3,4,5]))
```

```
[10, 20, 30, 40, 50]
```

うまくプログラムが書けたらこのような結果が返ってくる

Python演習 解答例

- **解答例**

```
def tentimes(numlist):
  r = []
  for x in numlist:
    r = r + [10*x]
  return r

print(tentimes([1,2,3,4,5]))
```

最初にrに空リストをいれる

rにnumlistの各要素を10倍にした値を追加していく

最後にrを返す

リストの連結はリスト同士でないとできないので、要素1個のリスト[10*x]を作ってから連結している

Python演習 解答例

- 別解

```
def tentimes(numlist):
  r = [0]*len(numlist)
  for i in range(len(numlist)):
    r[i] = 10 * numlist[i]
  return r

print(tentimes([1,2,3,4,5]))
```

最初にnumlistと同じ長さのダミーのリストを作っておく

i=0, .., (numlistの大きさ-1)まで繰り返す

r[i]をnumlist[i]の10倍の値にする

最後にrを返す

クラス

- **独自のデータ型の定義**
  - メソッド(クラス用の関数)や属性(クラス用の変数)の定義が可能

クラスの定義

```
class Man:
  def __init__(self, name):
    self.name = name
    print('Initialized!')

  def hello(self):
    print('hello ' + self.name + '!')

  def goodbye(self):
    print('good bye '+ self.name + '!')

m = Man('Taro')
m.hello()
m.goodbye()
```

__init__：コンストラクタ（初期化用メソッド）

.でメソッドや属性にアクセス（クラス内ではself.）

メソッドの第一引数にはselfを明示的に書く

Man型オブジェクトの生成

実行結果

```
Initialized!
hello Taro!
good bye Taro!
```

クラスの継承

- 親クラス(基底クラス)に新しい機能を追加したクラス(派生クラス)を作る
  - 基底クラスの属性やメソッドを継承
  - 属性やメソッドの追加、上書き(オーバーライド)ができる

派生クラスの定義

```
class YoungMan(Man):
  def __init__(self, name):
    super(YoungMan, self).__init__(name)
    self.name = 'Mr. ' + self.name

  def hello(self):
    print('hi ' + self.name + '!')

m = YoungMan('Taro')
m.hello()
m.goodbye()
```

class 派生クラス名(基底クラス名):

super(派生クラス名, self).メソッド名で基底クラスのメソッドを呼び出せる

Helloメソッドを上書き

派生クラスのhelloが呼ばれる

基底クラスのgoodbyeが呼ばれる

実行結果

```
Initialized!
hi Mr. Taro!
good bye Mr. Taro!
```

(参考) Python: リストのメソッド

- **sort**
  - リストの各要素を並べ替えて昇順にソートする
  - x.sort(reverse=True)で、降順にソート
  - x.sort(key=関数)と書くことで各要素を関数に適用した値の順に並べ替える
    - 例

      x = [[1,2,3], [4,5], [6]]

      x.sort(key=len)

      →[[6], [4,5], [1,2,3]]
- **append**
  - リストの末尾に要素を追加する
  - 例 x.append(10)

(参考) Python: ラムダ式

- **ラムダ式**

lambda 変数x: 式e

- 名前無し関数を生成
- 変数xを受け取って、式eの計算結果を返す関数
- 例 f = lambda x: x+10

- 次の2つは同じ定義

```
f=lambda x: x+10
```

```
def f(x):
    return x + 10
```

(参考) Python: 内包表記

- **内包表記 (comprehension)**

  [式e for 変数x in リストl]

  - for文を用いたリストの生成を簡便に行うための構文
  - リストlの各要素を変数xに代入し、式eを実行する
  - 式eの実行結果をリストとして順に並べて出力する

  - 例

    [10*x for x in [1,2,3,4]]

    →[10, 20, 30, 40]

(参考) Python: 3項演算 (if関数)

- **Pythonにおいて通常用いられるifは制御構造文(返値がない)であったが、if関数(3項演算子)も用意されている**

式e if 条件式c else 式f

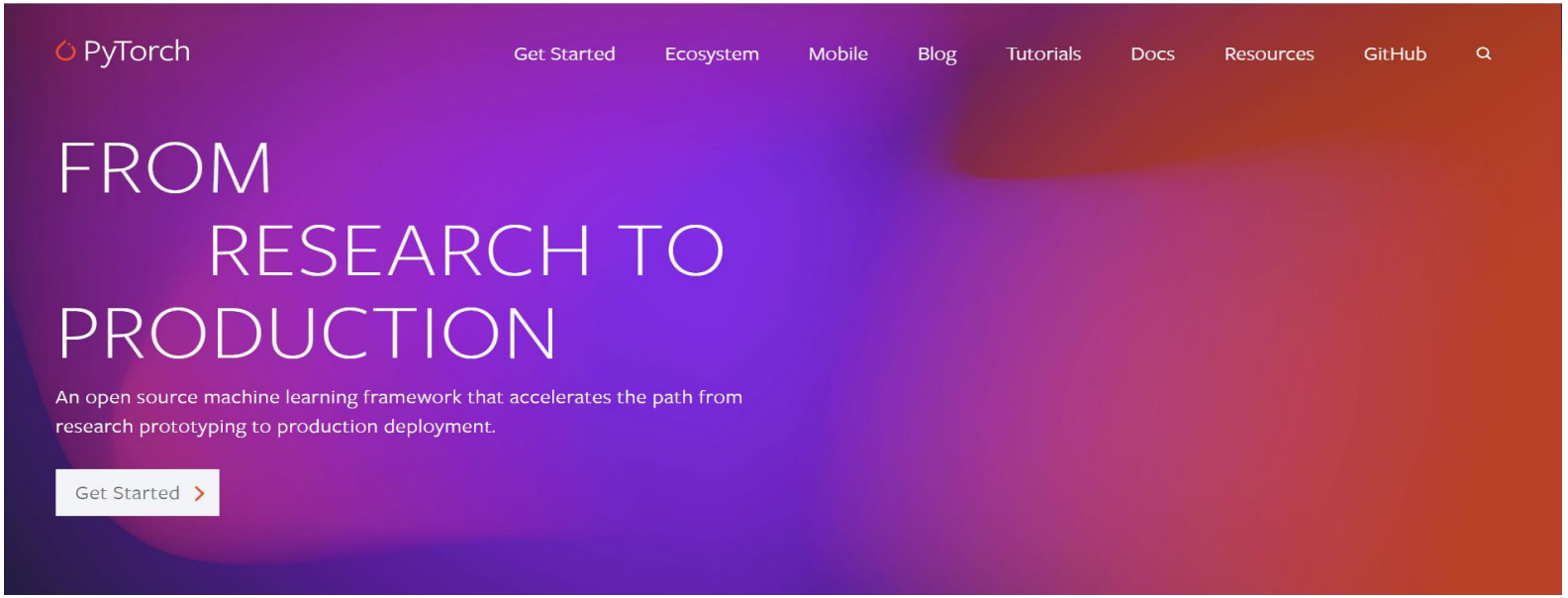
  - 条件式cを満たす時、式eの計算結果を返す。そうでなければ、式fの計算結果を返す

  - 例 x = 'odd' if y % 2 == 1 else 'even'

PYTORCH

PyTorch

- https://pytorch.org/



PyTorch

- **Metaによって開発されたニューラルネットワーク学習フレームワーク**
  - Define-by-Runによるネットワーク定義
    - データごとに構成が変化するネットワークにも柔軟に対応
    - ⇔Define-and-Run
  - Pythonさえ知っていれば記述できる

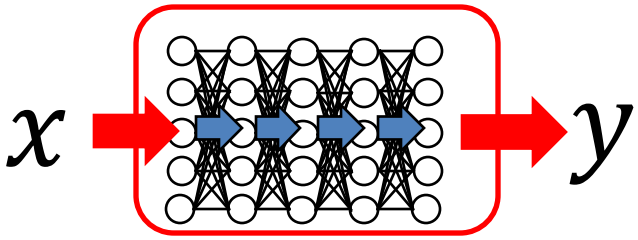
PyTorch

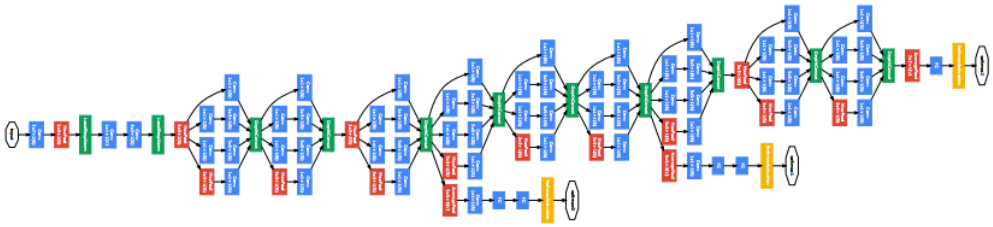
- Define-and-Run (Caffe, TensorFlow)
  - ネットワークを先につくってから、データをネットワークに流し込む
  - 一度作ったネットワークを変更することができない
  - 計算が速い

- Define-by-Run (PyTorch, Chainer)
  - データを流し込むと同時にネットワークを作る
  - データ入力時にデータにあわせてネットワークを作ることができる
    - 文のように可変サイズのデータを扱うときに便利
  - データの入力ごとにネットワークを作っているため遅い

Define-by-Runの利点

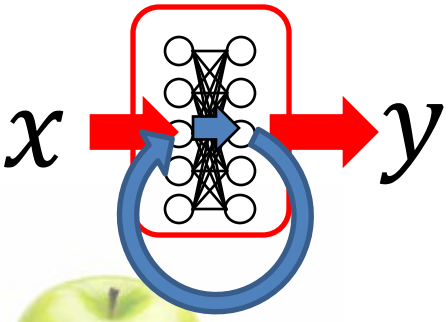
- フィードフォワードニューラルネットワーク

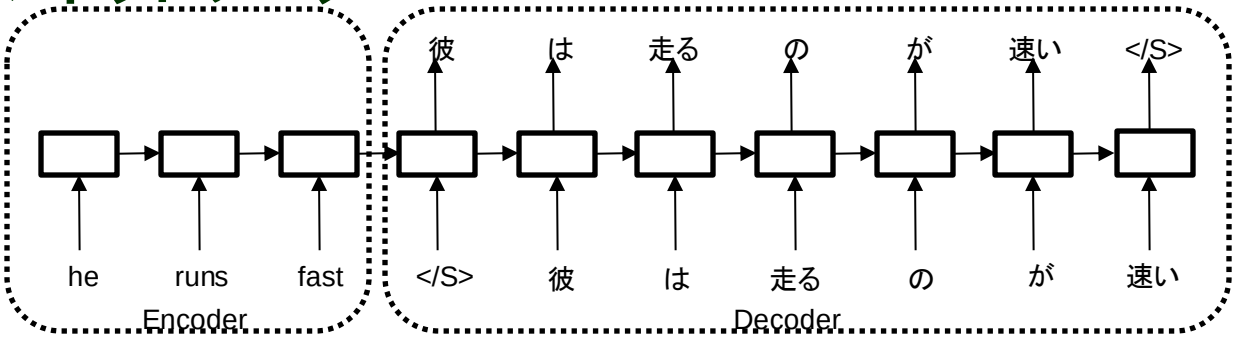




GoogLeNet (22層)

- リカレントニューラルネットワーク





Encoder-Decoderによる機械翻訳

入力によってネットワークが変化する時系列解析でも、
Define-by-Runなら問題なく適用できる

PYTORCHのテンソル型

テンソル型 (torch.tensor)

- **Pythonで高速に行列計算を行うためのデータ型**
  - 行列演算は単純な足し算掛け算を繰り返すがpythonはこういった計算が遅くて苦手
  - Pythonの便利さと高速計算を同時に実現する
- **PyTorchのテンソル型(torch.tensor)には多次元配列を操作するための便利なメソッドが多数用意されている**
- **PyTorchをインポートして利用する**

```
>>> import torch
```

torch.tensor：テンソル型の生成

- **テンソル型の生成**
  - メソッドtensor()を使用する。引数はPythonのリスト。
  - 多次元配列も生成可能。2次元配列は行列と捉えられる。
  - メソッドsize()で配列の形状が得られる
  - 属性dtypeに要素のデータ型が保存されている

2×3の行列生成

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

```
>>> import torch
>>> x = torch.tensor([1.0, 2.0, 3.0])
>>> x
tensor([ 1.,  2.,  3. ])
>> type(x)
<class 'torch.Tensor'>
>>> A = torch.tensor([[1, 2, 3], [4,5,6]])
>>> A
tensor([[1, 2, 3],
        [4, 5, 6]])
>>> A.size()
torch.Size([2, 3])
>>> A.dtype
torch.int64
```

torch.tensor：テンソル型の演算 (1)

- 四則演算子による演算は**要素毎**の演算

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} -1 & 1 \\ 5 & 6 \end{pmatrix}$$

```
>>> x = torch.tensor([1., 2., 3.])
>>> y = torch.tensor([2., 4., 6.])
>>> x + y
tensor([3., 6., 9.])
>>> x - y
tensor([-1., -2., -3.])
>>> x * y
tensor([2., 8.,18.])
>>> x / y
tensor([0.5000,0.5000,0.5000])
```

要素毎の積

行列の積（ドット積）は matmulメソッド

```
>>> A = torch.tensor([[1, 2], [3, 4]])
>>> B = torch.tensor([[-1, 1], [5, 6]])
>>> A + B
tensor([[ 0,  3],
        [ 8, 10]])
>>> A * B
tensor([[-1,  2],
        [15, 24]])
>>> torch.matmul(A,B)
tensor([[ 9, 13],
        [17, 27]])
```

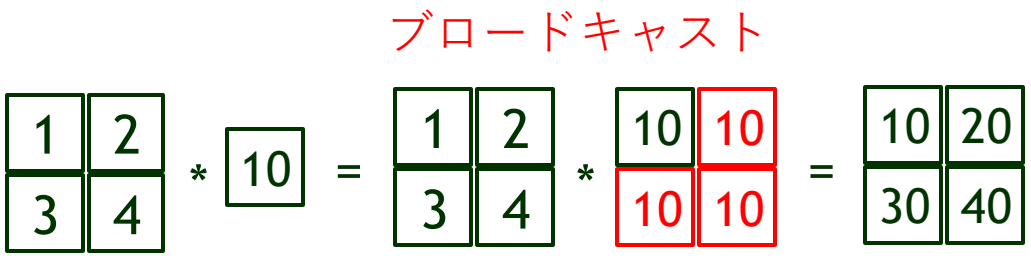
$$\begin{pmatrix} 1\times -1 & 2\times 1 \\ 3\times 5 & 4\times 6 \end{pmatrix}$$

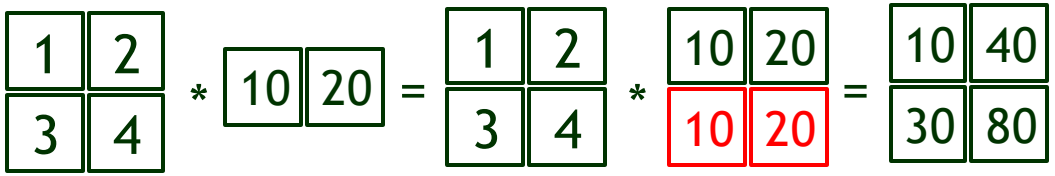
$$\begin{pmatrix} 1\times -1+2\times 5 & 1\times 1+2\times 6 \\ 3\times -1+4\times 5 & 3\times 1+4\times 6 \end{pmatrix}$$

torch.tensor：テンソル型の演算 (2)

- 配列の形状が異なる場合は、形状を揃えて要素毎の演算がされる(ブロードキャスト)。

```
>>> A = torch.tensor([[1, 2], [3, 4]])
>>> A*10
tensor([[10, 20],
        [30, 40]])
>>> A+10
tensor([[11, 12],
        [13, 14]])
>>> B = torch.tensor([10, 20])
>>> A*B
tensor([[10, 40],
        [30, 80]])
```





torch.tensor: 要素へのアクセス

- **Pythonリストの要素へのアクセスと同様にインデックスによりアクセス可能**
- **加えて、tensor型配列によるアクセスも可能**
  - インデックスの配列により指定番目の要素だけ取得
  - ブーリアンの配列によりTrueに対応する要素を取得

```
>>> X = torch.tensor([[10, 40], [15, 30], [1, 0]])
>>> X
tensor([[10, 40],
        [15, 30],
        [ 1 , 0]])
>>> X[0] #0行目
tensor([10, 40])
>>> X[0][1] #(0,1)の要素
tensor(40)
>>> X = X.flatten()
>>> X
tensor([10, 40, 15, 30,  1,  0])
>>> X[torch.tensor([0,2,4])]
tensor([10, 15,  1])
>>> X>20
tensor([False,  True, False,  True, False, False])
>>> X[X>20]
tensor([40, 30])
```

Xを1次元配列に変換

tensor配列に対して不等号演算を行うと配列の各要素に対して不等号演算が行われた結果のブーリアンの配列となる

まとめ

- Pythonプログラミングの基礎について学んだ
- PyTorch
  - torch.tensor型について学んだ

深層学習 (2)
-ニューラルネットワークの学習、損失関数、勾配法 -

機械学習

例：犬猫判別機 $f(x)$

入力x

出力y

$\xrightarrow{f}$ 犬

データ


：猫

：犬


：犬

：猫


：犬 …

大量のデータから関数f を学習する　f: 画像 → {犬/猫}

機械学習

- **データから学習**
  - 入力 $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$　←m次元ベクトル
    （画像の場合は、赤、青、緑の画像に対応する3つの行列）
  - データDは入力$\boldsymbol{x}$と出力yのペアの集合
  - 関数fはパラメータ(重み変数)の集合 $\boldsymbol{w} = (w_1, w_2, \cdots)$から成る計算式
    - NNも巨大な一つの関数
  - 学習 = データDに対する誤差(損失)を最小にする重み変数wを求める

データ全体の誤差(損失) $L(\boldsymbol{w}) = \sum_{(\boldsymbol{x},y)\in D} \left(y - f_{\boldsymbol{w}}(\boldsymbol{x})\right)^2$ ← 損失関数

NNの学習：教師データに対する誤差(損失)が
最小になる重み変数を探すこと

NNの学習

教師データに対する損失が最小になる各ノードの重みを探すこと

$\mathrm{argmin}_w L(w)$　　$L(w)$：データ全体の損失

平均2乗誤差　$L(w) = \frac{1}{2N}\sum_n\sum_k (y_{nk} - t_{nk})^2$

交差エントロピー損失　$L(w) = -\frac{1}{N}\sum_n\sum_k t_{nk} log_e\, y_{nk}$

y_{nk}：n個目のデータのk次元目のNNの出力、
t_{nk}：教師データ
N　：教師データの個数

2乗和誤差

$$\frac{1}{2}\sum_k (y_k - t_k)^2$$

y_k ：NNの出力、
t_k ：教師データ

例) $y = (13.5, 20.1, 0.7)$

$t = (12.1, 22.8, 0.3)$

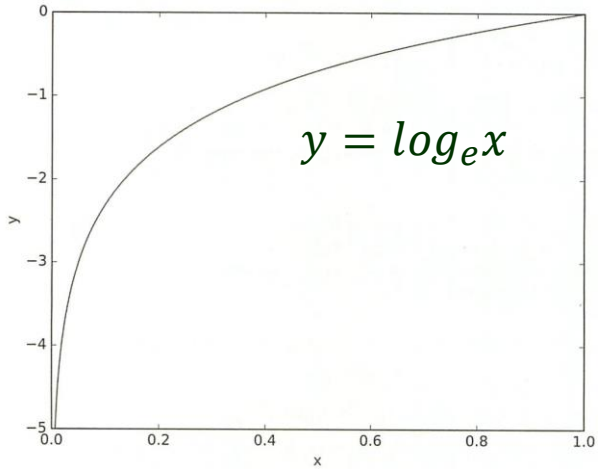
損失$= \frac{1}{2} \times (1.4^2 + 2.7^2 + 0.4^2) = 4.705$

回帰問題の場合によく用いられる

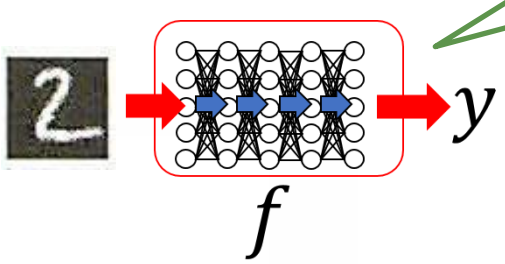
交差エントロピー損失

$$-\sum_k t_k log_e y_k$$

y_k ：NNの出力、
t_k ：教師データ



（例）手書き数字認識



出力層の活性化関数：ソフトマックス関数
出力層のノード数：10

$y = [0.1, 0.05, 0.6, 0.0, 0.05, 0.1, 0.0, 0.1, 0.0, 0.0]$

$t = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]$

損失$= -log_e 0.6 = 0.51$

分類問題の場合によく用いられる

データ集合に対する損失への拡張

- N個の教師データの集合に対する損失

1個のデータに対する誤差

N個のデータに対する誤差

平均2乗誤差

2乗和誤差

$$\frac{1}{2}\sum_{k}(y_k - t_k)^2 \quad \Rightarrow \quad L(w) = \frac{1}{2N}\sum_{n}\sum_{k}(y_{nk} - t_{nk})^2$$

y_{nk}：n個目のデータのk次元目NNの出力、
t_{nk}：教師データ

交差エントロピー損失

$$-\sum_{k} t_k \log_e y_k \quad \Rightarrow \quad L(w) = -\frac{1}{N}\sum_{n}\sum_{k} t_{nk} \log_e y_{nk}$$

ミニバッチ学習

平均2乗誤差 $L = \frac{1}{2N}\sum_n\sum_k (y_{nk} - t_{nk})^2$

交差エントロピー損失 $L = -\frac{1}{N}\sum_n\sum_k t_{nk} log_e y_{nk}$

y_{nk}：n個目のデータのk次元目のNNの出力、
t_{nk}：教師データ、
N　：教師データの個数

- **バッチ学習：教師データ中の全データに対する損失を利用**
  - $N = |D|$：教師データ中の全データ数

データが増えると計算大変

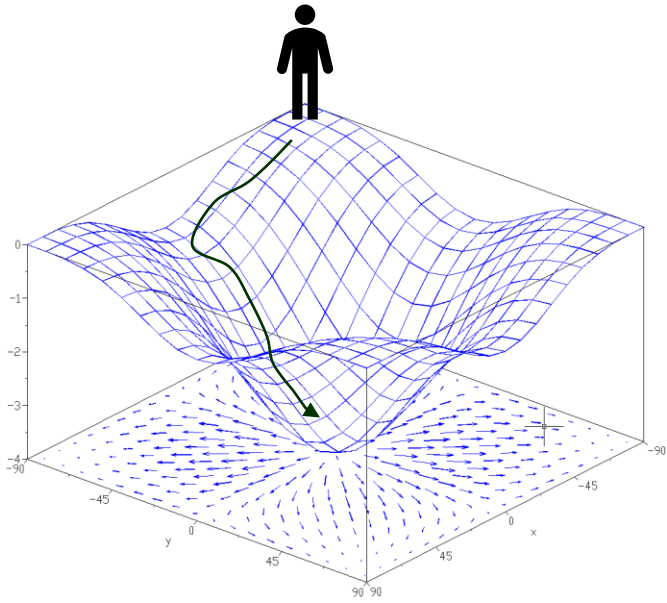
- **ミニバッチ学習：教師データ中の一部のデータに対する損失を利用**
  - $N = M(M < |D|)$：バッチサイズ

損失の最小化

- **損失を最小化するには数値最適化の手法を用いる**
  - ニュートン法、準ニュートン法
  - 最急降下法
  - オンライン学習(確率的勾配法(SGD), AdaGrad, Adamなど)

- **勾配を用いて計算する手法が多い**
  - 勾配

$$\nabla L = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \cdots, \frac{\partial L}{\partial w_m} \right)$$



微分

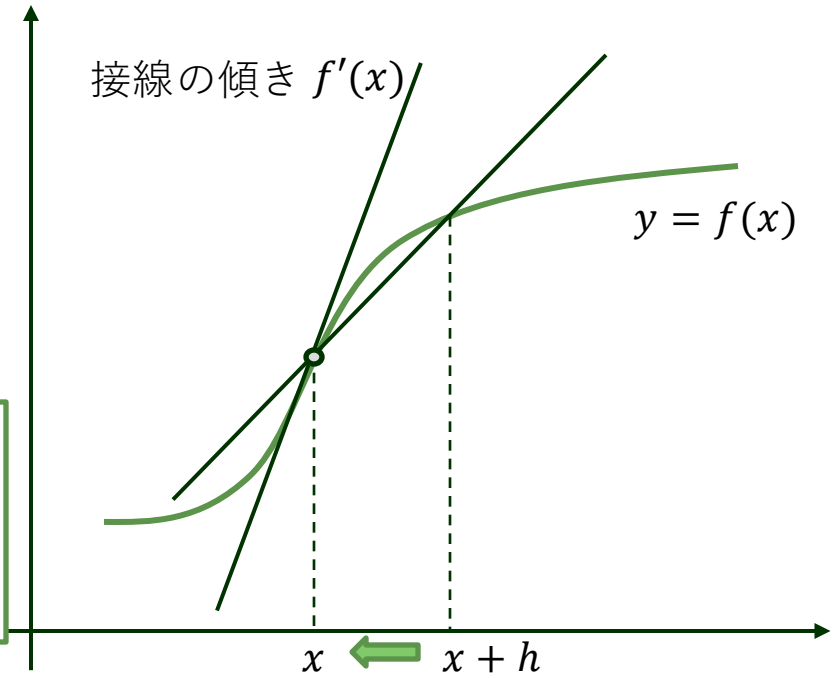
- xに対するf(x) の変化（接線の傾き）

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

符号：変化の方向
大きさ：変化の度合い

実装

```
def numerical_diff(f, x):
    h = 1e-4
    return (f(x+h) - f(x)) / h
```



数値微分

- **数値微分**
  - 微分の定義により微分を計算する

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- **解析的に解く**
  - 数式の展開により微分を計算する

例：$f(x) = x^2$、$x = 2$での微分の計算

数値微分

$f'(x)$を陽に求めなくてよい

$$h = 0.0000001$$
$$f'(2) = \frac{(2+h)^2 - 2^2}{h} = 4.00000009153311$$

解析的に解く

$$f'(x) = 2x$$
$$f'(2) = 4$$

真の微分が求まる

偏微分 (1/2)

- **変数が複数ある場合の微分**
  - 特定の変数に着目して微分を計算（その他の変数は定数とみなす）

例：$f(x_0, x_1) = {x_0}^2 + {x_1}^2$、$x_0 = 3$、 $x_1 = 4$のときの各変数に対する偏微分

解析的に解く

$$\frac{\partial f}{\partial x_0} = 2x_0 \qquad \frac{\partial f}{\partial x_1} = 2x_1$$

$$\frac{\partial f}{\partial x_0} = 6 \text{（} x_0 = 3 \text{のとき）} \qquad \frac{\partial f}{\partial x_1} = 8 \text{（} x_1 = 4 \text{のとき）}$$

勾配（gradient）(1/2)

- **全ての変数の偏微分をベクトルとしてまとめたもの**

例：　関数$L(w_0, w_1)$の勾配

$$\nabla L = \left(\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1} \right)$$

$L(w_0, w_1) = {w_0}^2 + {w_1}^2$の$w_0 = 3$、 $w_1 = 4$における勾配は$(6, 8)$

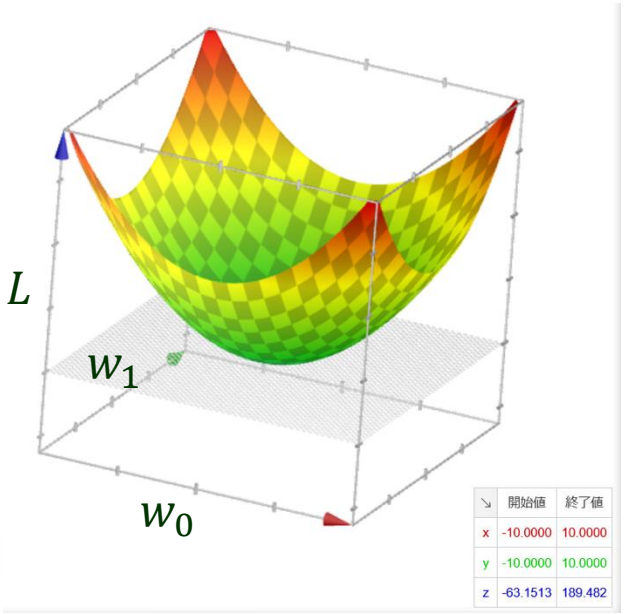
$$\nabla L = \left(\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1} \right) = (2w_0, 2w_1)$$

$$\nabla L(3,4) = (6,8)$$

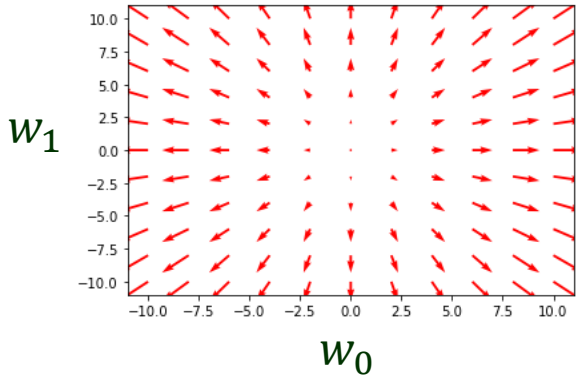
勾配(gradient)(2/2)

- **その点における傾斜方向を表す**

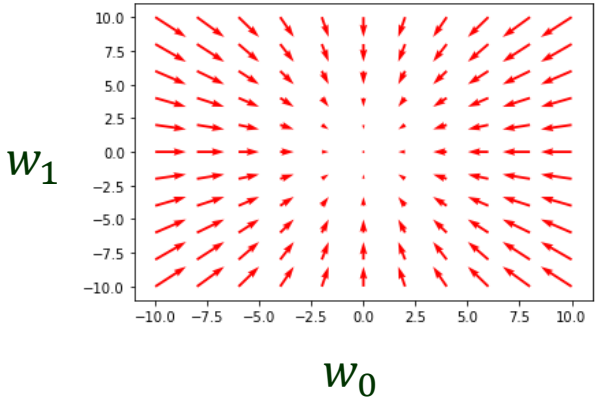
$$L(w_0, w_1) = {w_0}^2 + {w_1}^2$$



各点における勾配



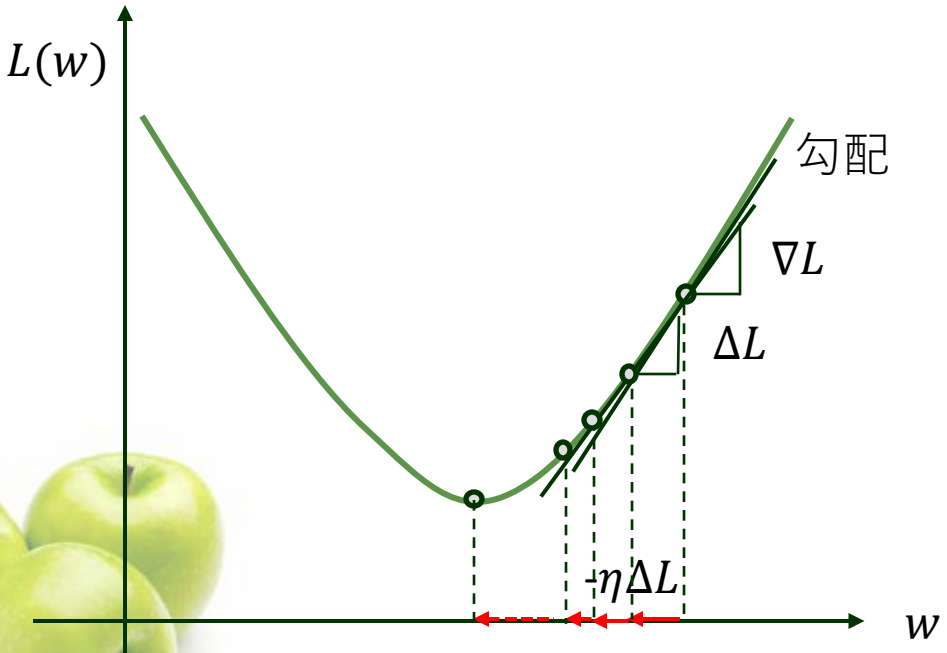
各点における勾配に-1をかけたベクトル



勾配に-1をかけたベクトルは各場所で関数の値を最も減らす方向を示す！！

勾配降下法(Gradient Descent)(1/4)

- **関数の勾配を利用して最小値となる所を見つける方法**
  - 「現在の場所から勾配方向に一定の距進む→移動先で勾配を求める→勾配方向に一定の距離移動進む」この動作を繰り返して関数の値を徐々に減らす



1変数の時の勾配

$$w \leftarrow w - \eta \frac{dL(w)}{dw}$$ ←この計算を繰り返す

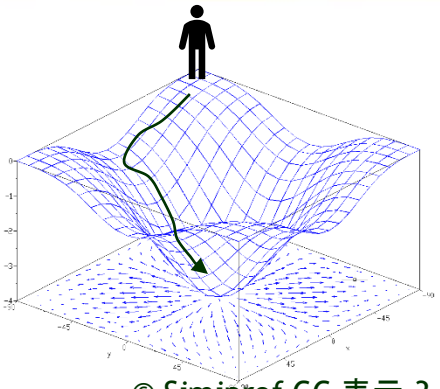
η：**学習率**

勾配降下法(Gradient Descent)(2/4)

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla L(\boldsymbol{w})$$

←この計算を繰り返す

η：学習率



ステップ0：初期値の設定
　各パラメータの初期値を設定

$\nabla L(\boldsymbol{w})$

ステップ1：勾配の算出
　最小化する関数に基づき、パラメータの勾配を求める

ステップ2：パラメータの更新
　パラメータを勾配方向とは逆向きに微小量更新する

$-\eta \nabla L(\boldsymbol{w})$

ステップ3：繰り返す
　ステップ1に戻る

勾配降下法(Gradient Descent)(3/4)

例：$L(w_0, w_1) = w_0^2 + w_1^2$を最小にする$w_0$と$w_1$を求める（学習率$\eta$：0.1）

ステップ0　　$w_0 = 1, w_1 = 0.9$

ステップ1（1回目）

$$\nabla L(w_0, w_1) = \left(\frac{\partial L(w_0, w_1)}{\partial w_0}, \frac{\partial L(w_0, w_1)}{\partial w_1}\right) = (2w_0, 2w_1)$$

$$\nabla L(1, 0.9) = (2, 1.8)$$

ステップ2（1回目）　$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \nabla L(\boldsymbol{w})$

$$w_0 = 1 - 0.1 \times 2 = 0.8, w_1 = 0.9 - 0.1 \times 1.8 = 0.72$$

ステップ1（2回目）

$$\nabla L(0.8, 0.72) = (1.6, 1.44)$$

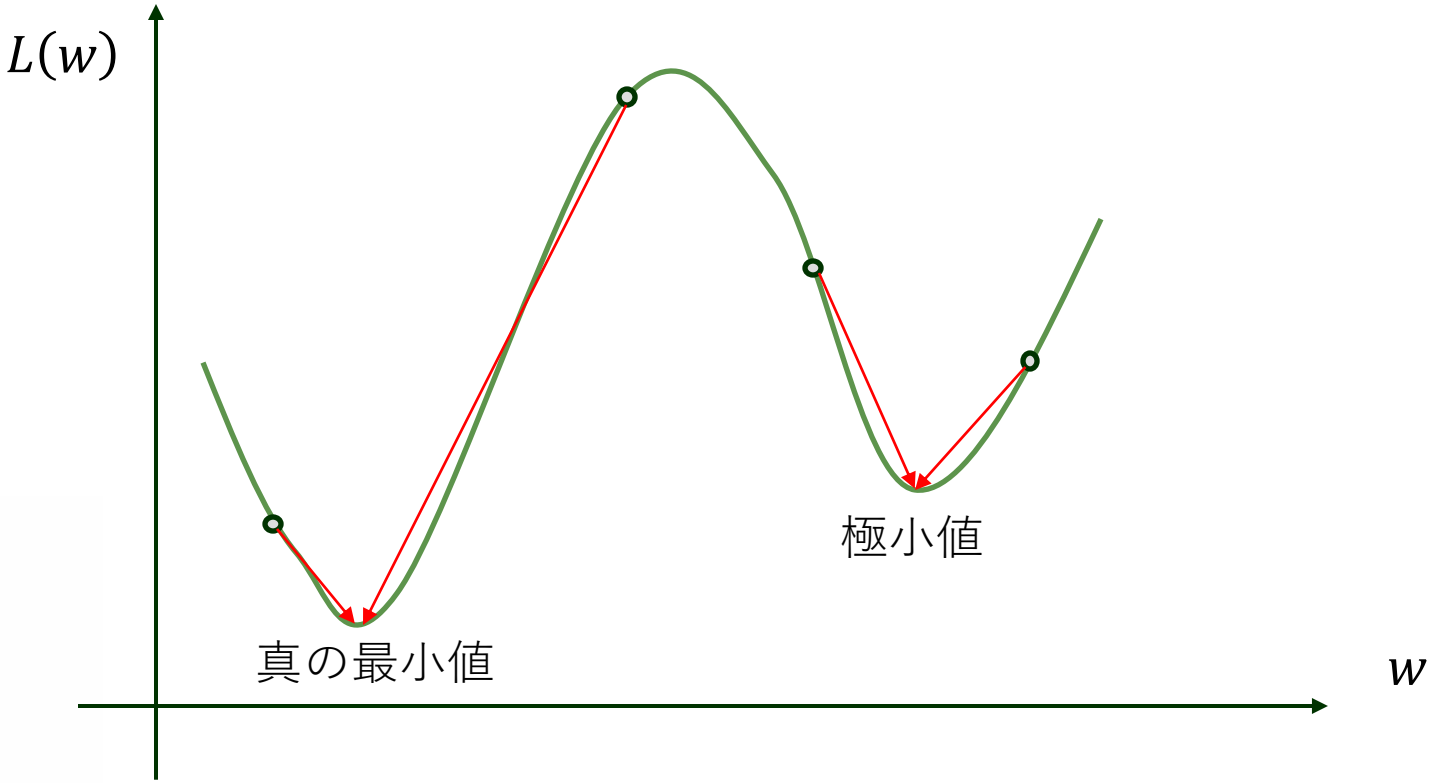
ステップ2（2回目）

$$w_0 = 0.8 - 0.1 \times 1.6 = 0.64, w_1 = 0.72 - 0.1 \times 1.44 = 0.576$$

⋮

勾配降下法(Gradient Descent)(4/4)

- 注意点：勾配のさし先が**本当の最小値とは限らない**。その場合、極小値（局所的な最小値）となる。



【参考】勾配降下法の実装 (1/2)

Pythonによる実装

```
def gradient_descent(f, init_x, lr,
step_num):
    x = init_x

    for i in range(step_num):
        grad = numerical_grad(f, x)
        x -= lr * grad

    return x
```

数値微分で勾配を算出

f：最適化したい関数（損失）
init_x：パラメータの初期値
lr：学習率
step_num：繰り返しの数

$$\boldsymbol{x} = \boldsymbol{x} - \eta \cdot \mathrm{gradient}(f, \boldsymbol{x})$$

↑この計算を
繰り返す

例：$f(x_0, x_1) = x_0^2 + x_1^2$の最小値

```
>>> def function(x):
. . .       return x[0]**2+x[1]**2
. . .
>>> init_x = torch.tensor([-3.0, 2.0])
>>> lr, step_num = 0.1, 100
>>> gradient_descent(function, init_x, lr, step_num)
tensor([-6.1111e-10,  4.0741e-10])←ほとんど(0,0)
```

確率的勾配降下法(Stochastic Gradient Descent; SGD)

- **ミニバッチ学習版の勾配降下法**
  - 損失の計算に教師データの一部を使用する

- **NNのSGDによる学習**
  1. ミニバッチ:教師データの中からランダムにミニバッチサイズ分のデータを選ぶ
  2. 勾配の算出:ミニバッチの損失に対する各重みパラメータの勾配を求める
  3. パラメータの更新:重みパラメータを勾配方向とは逆方向に微小量更新する
  4. 繰り返す:ステップ1に戻る

$$\boldsymbol{w} = \boldsymbol{w} - \eta \frac{\partial L}{\partial \boldsymbol{w}}$$

【参考】SGDにおける繰り返し

- **エポック数：全ての教師データを何回繰り返したかを表す表現**
  - 教師データが1,000個で、バッチサイズを100とした場合、ミニバッチSGDを10回繰り返すと全ての教師データを見たことに相当する

    → 10回の繰り返し=1エポック
- **「教師データD個、ミニバッチサイズMのSGDでNエポック学習した」**

→「 (D/M)×N回パラメータを更新した」ということ

学習した関数の評価

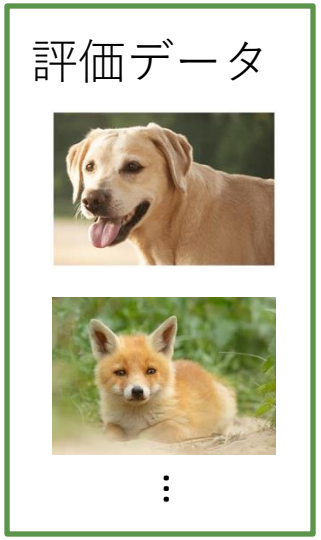
- **未知のデータに対する推論性能（汎化能力、汎化性能）が重要**

→ 教師データとは違う評価データ（未知のデータ）で性能評価

- 過学習の問題があるため

教師データに対する性能で評価してはダメ！

学習 → f^*



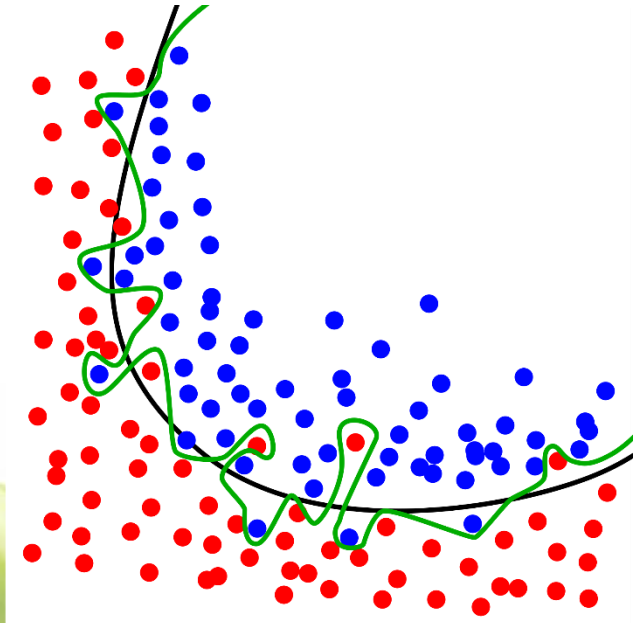
推論 → f^*

出力		正解
犬	○	犬
犬	×	キツネ
⋮	⋮	⋮

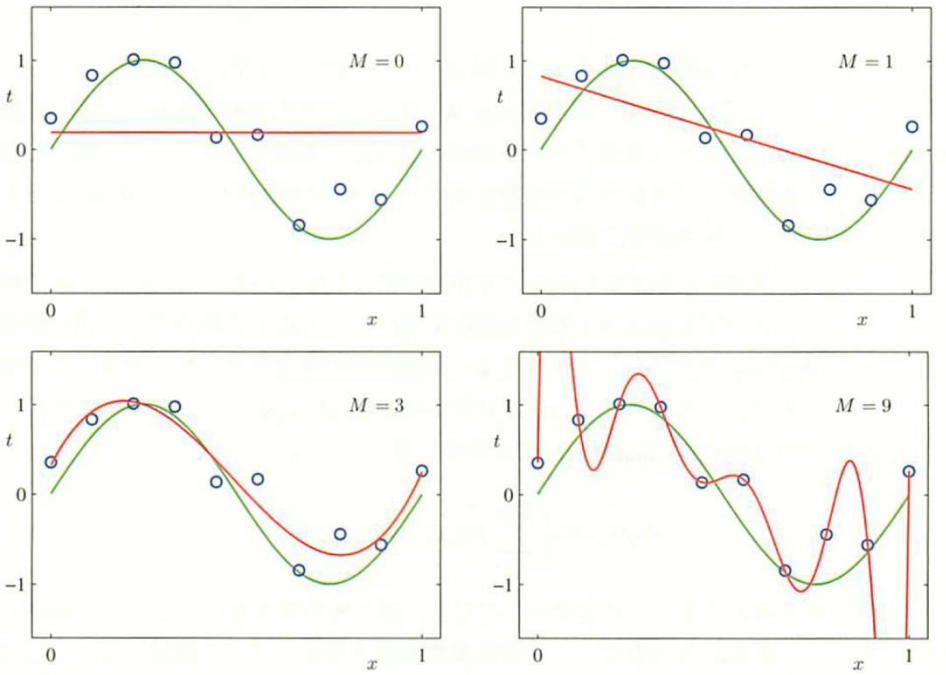
正解率：80%

過学習の問題

- **過学習の問題**
  - 関数が複雑すぎると過剰に適合してしまいうまく学習できない



$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$



Bishop, Pattern Recognition and Machine Learningより引用

まとめ

- **機械学習の基礎**
  - データから入力と出力をつなぐ関数を学習する手法
  - 問題の種類
    - 回帰問題、分類問題、構造予測
  - データ全体の損失を最小化することで重み変数(パラメータ)を学習
    - 回帰問題→平均2乗誤差
    - 分類問題→交差エントロピー損失
  - 過学習
    - 訓練データに過剰に適合してしまうこと
  - 学習の方法
    - 損失に対する勾配を計算して勾配方向に重みベクトルを更新
    - 勾配降下法、SGD

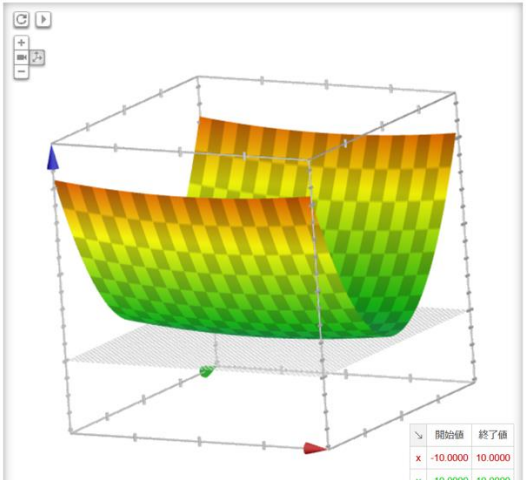
(参考) SGDの欠点

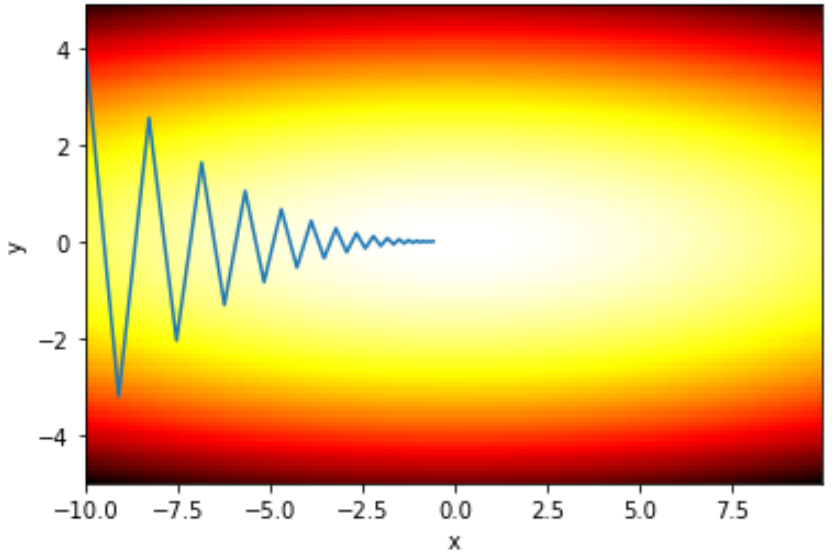
- 勾配の方向が本来の最小値ではない方向を指しているときは非効率な場合がある

例：

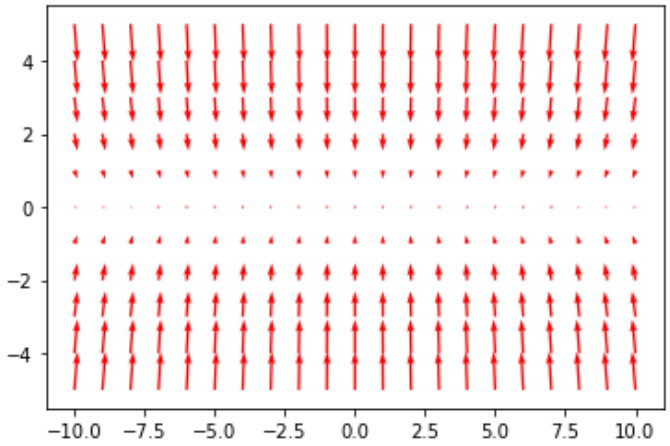
$$f(x,y) = \frac{1}{20}x^2 + y^2$$

最小値(0,0)





勾配



$$\boldsymbol{w} = \boldsymbol{w} - \eta \cdot \frac{\partial L}{\partial \boldsymbol{w}}$$

ジグザグで最小値に近づく
→非効率

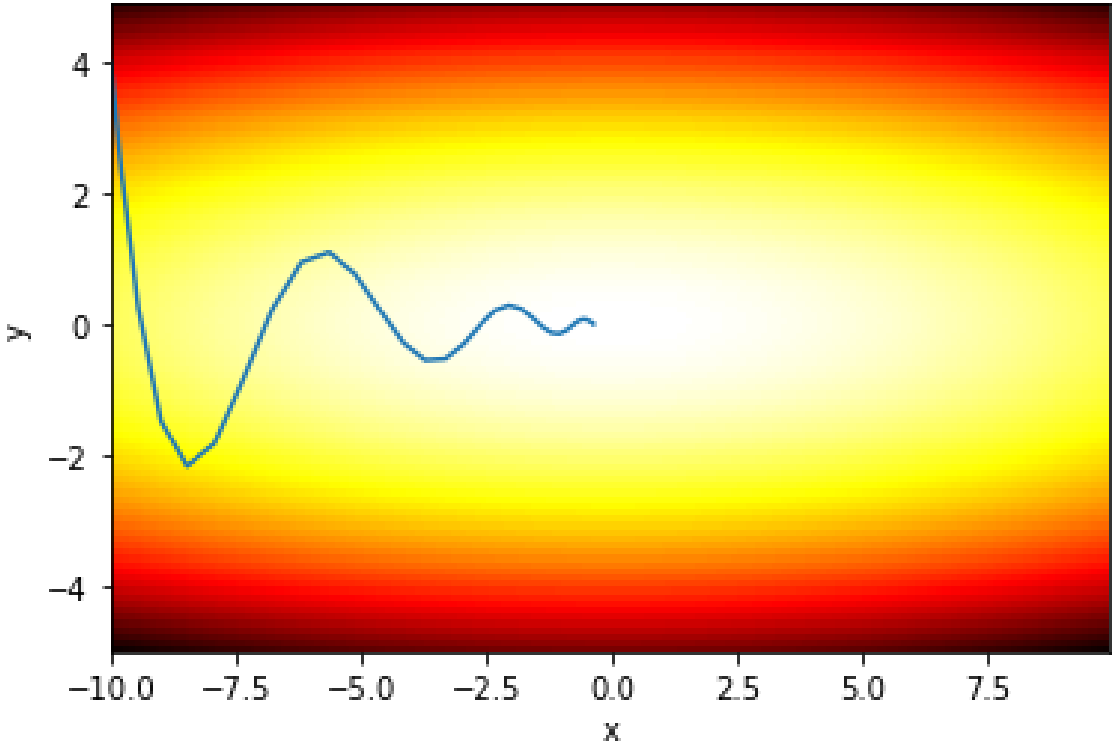
(参考) 最適化手法:Momentum

$$\boldsymbol{v} = \alpha \boldsymbol{v} - \eta \cdot \frac{\partial L}{\partial \boldsymbol{w}}$$

$$\boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$$

$\boldsymbol{v}$：速度の役割
（勾配方向に速度が加算される）

α：ハイパーパラメータ(0.9等)



・x軸の勾配の方向は一定
→x軸方向には加速される
・y軸の勾配の方向は交互
→y軸方向への動きは抑制

⇒ジグザグの動きが軽減

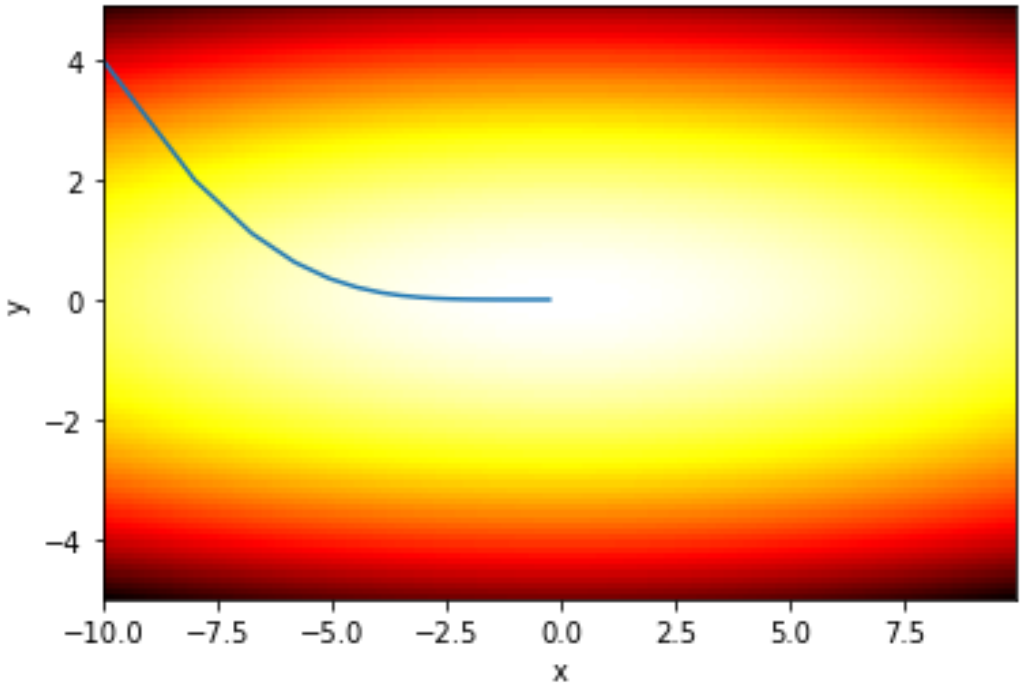
(参考) 最適化手法: AdaGrad

- **パラメータ毎に学習率をコントロール**
  - 最初は大きく次第に小さく
  - 大きく更新されたパラメータほど小さくする

$$\boldsymbol{h} = \boldsymbol{h} + \frac{\partial L}{\partial \boldsymbol{w}} \otimes \frac{\partial L}{\partial \boldsymbol{w}}$$

$$\boldsymbol{w} = \boldsymbol{w} - \eta \frac{1}{\sqrt{\boldsymbol{h}}} \otimes \frac{\partial L}{\partial \boldsymbol{w}}$$

$\otimes$: 要素毎の積

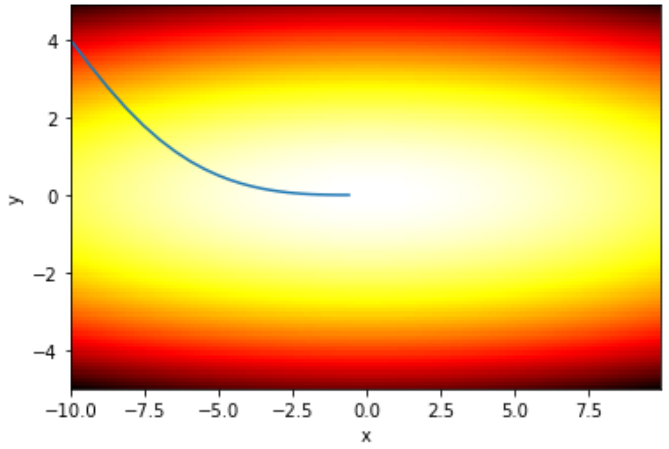


Y軸方向の勾配は大きい
→最初は大きく更新されるが
　更新が進むほど更新度合いが弱まる
⇒ジグザグの動きが軽減

(参考) その他の最適化手法

- Adadelta
  - AdaGradの発展形

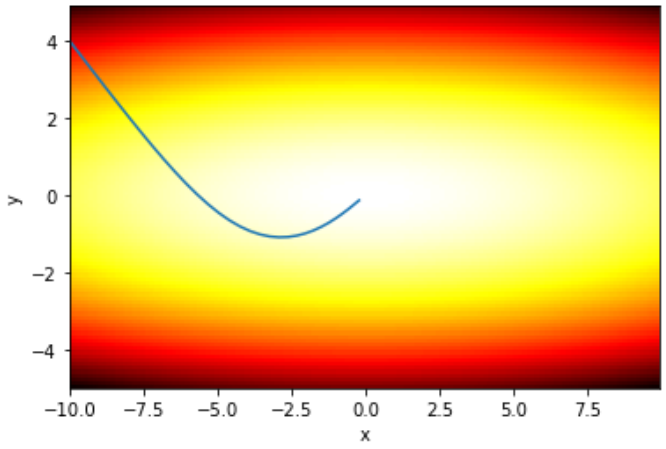
※ Matthew Zeiler, ADADELTA: An Adaptive Learning Rate Method, 2012.



- Adam
  - AdaGradのアイデア＋Momentumのアイデアの融合

※ Diederik Kingma and Jimmy Ba, Adam: a Method for Stochastic Optimization, 2015.

など



(参考) 重みの初期値

- **活性化関数がシグモイド関数の場合**
  - Xavierの初期値：$\frac{1}{\sqrt{n}}$を標準偏差とするガウス分布で各ノードの重みを初期化（n：前層のノード数）

  Xavier Glorot and Yoshua Bengio, Understanding the difficulty of training deep feedforward neural networks, 2010.

- **活性化関数がReLUの場合**
  - Heの初期値：$\frac{2}{\sqrt{n}}$を標準偏差とするガウス分布で各ノードの重み初期化（n：前層のノード数）

  Kaiming He, Xiangyu Zhang, Shaoqing Re, and Jian Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, 2015.

(参考) 過学習への対応法：荷重減衰 (Weight Decay)

- **重みのL2ノルムを正則化項として加えた損失に基づき学習**
  - Weight Decayを用いた損失

$$L(\boldsymbol{w}) + \frac{1}{2}\lambda\|\boldsymbol{w}\|_2^2$$

λ：ハイパーパラメータ

$\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$に対するL2ノルムは

$$\|\boldsymbol{w}\|_2 = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$

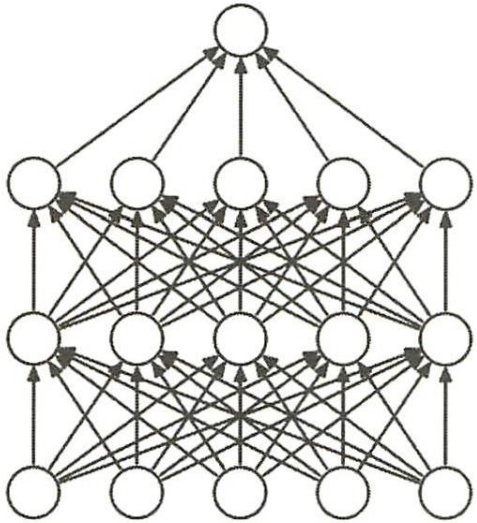
従って、

$$\|\boldsymbol{w}\|_2^2 = w_1^2 + w_2^2 + \cdots + w_n^2$$

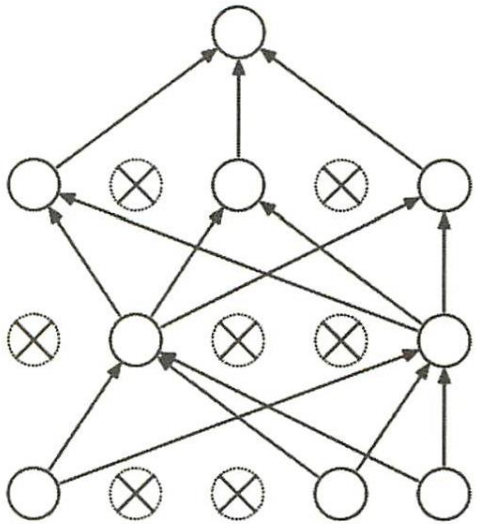
(参考) 過学習への対応法：Dropout

- 学習時にはデータが流れるたびに隠れ層のノードをある割合でランダムに選択し消去して学習
- テスト時には各層の出力に対して訓練時に消去した割合を乗算した値を出力

N. Srivastava, G. Hinton, A. Krizhevsky, I.sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, 2014.



Dropoutなし



Dropoutあり

(参考) Batch Normalization

- **ミニバッチ毎に入力データを正規化**

ミニバッチ：$\{x_1, x_2, \ldots x_m\}$
ε：小さな値(10e-7等)

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu)^2$$

$$\widehat{x_i} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}}$$

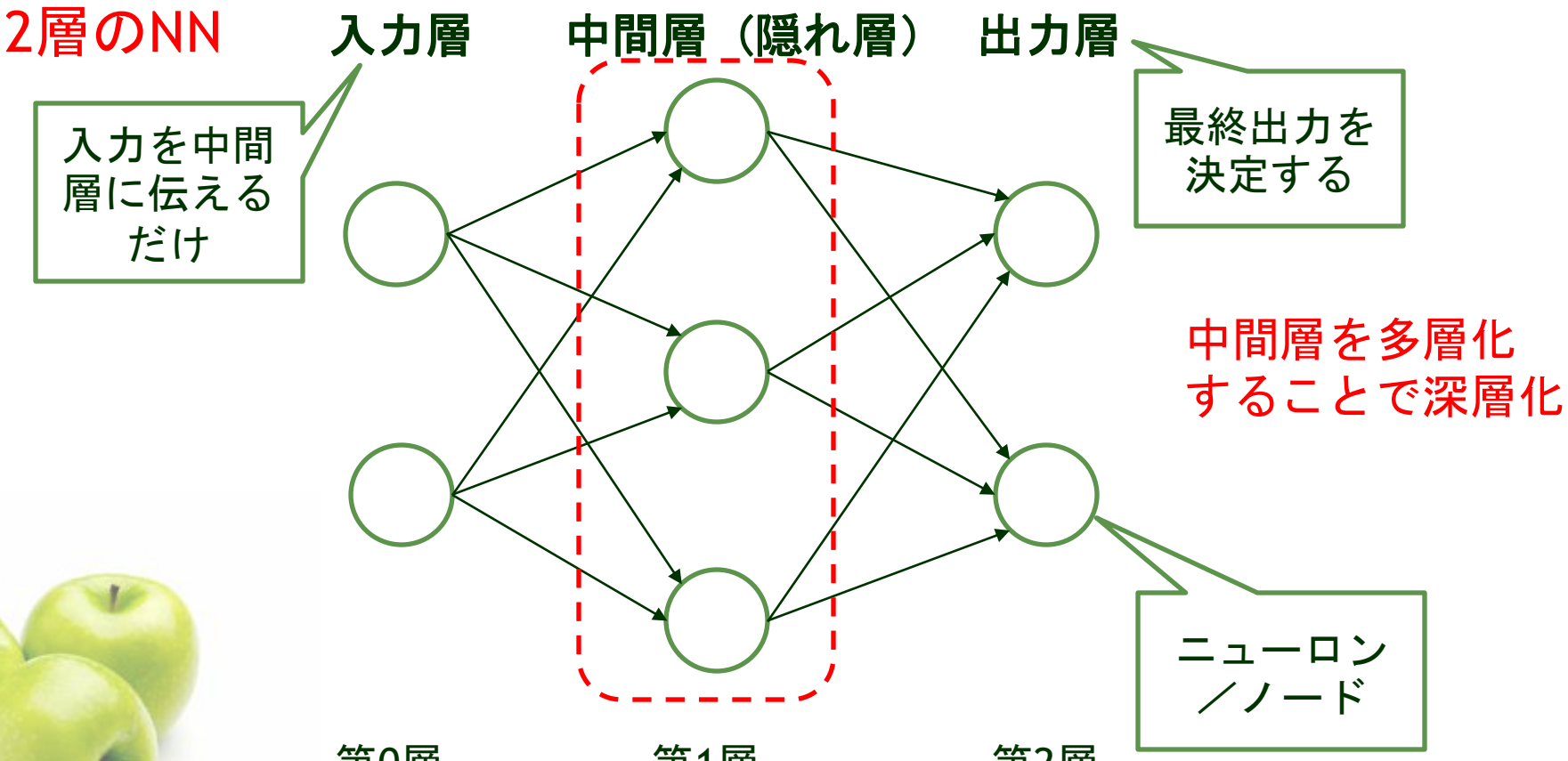
- **効果**
  - 学習が速くなる
  - 初期値への依存度を減らせる
  - 過学習を抑制する

Sergey Ioffe and Christian Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015.

深層学習 (3)
-誤差逆伝播法、計算グラフ、畳込みニューラルネットワーク -

【復習】標準的なNNの構造

- 順伝播型(フィードフォワード)NN:情報が入力側から出力側に一方向に伝わるNN



ここからは、

重みパラメータの勾配$\frac{\partial L}{\partial w_i}$の計算を効率良く行う「誤差逆伝播法(Back Propagation)」を学びます

連鎖律 (chain rule)

- **合成関数（複数の関数によって構成される関数）の微分に関するルール：合成関数の微分は、構成するそれぞれの関数の微分の積で計算できる**

例：　$z = t^2$
　　$t = x + 2y$

$z = (x + 2y)^2$のこと

連鎖律

打ち消し

$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x}$ と理解

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x} = 2t \cdot 1 = 2(x + 2y)$$

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial y} = 2t \cdot 2 = 4(x + 2y)$$

多変数の連鎖律

- **多変数から成る合成関数の微分に関するルール**

$z = f(u, v, w)$**のとき**

関数fに対しu, v, wについてのみ偏微分すれば良い。

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial z}{\partial w}\frac{\partial w}{\partial x}$$

u, v, wに対して再帰的にxについて微分すれば良い。

例：
$z = uv$
$u = x^2 + y^2$
$v = \sin w$
$w = xy$

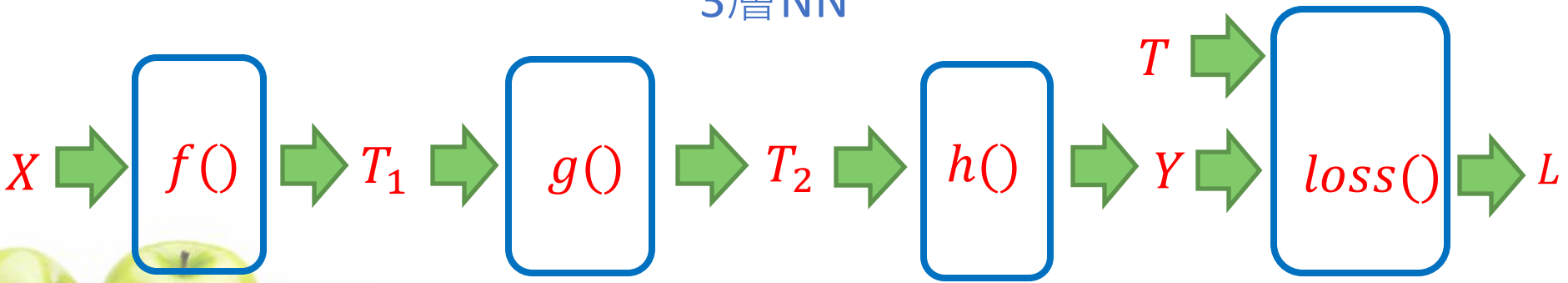
$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x} = v\frac{\partial u}{\partial x} + u\frac{\partial v}{\partial x} = v \cdot 2x + u\left(\frac{\partial v}{\partial w}\frac{\partial w}{\partial x}\right)$$

$$= v \cdot 2x + u \cdot \cos w \cdot \frac{\partial w}{\partial x} = v \cdot 2x + u \cdot \cos w \cdot y$$

なぜ連鎖律はNNの学習に関係する？

- **巨大で複雑な関数の自動微分を実現したいから**
  - NNは単純な関数を組み合わせた複雑な関数とみなせる
  - 複雑な関数$L(w)$の微分$\frac{\partial L}{\partial w}$を計算したい
  - 連鎖律で基本的な関数(四則演算や、三角関数、活性化関数など)の微分に分解できる
  - 連鎖律で分解された微分を効率的に計算できる(誤差逆伝播法)

3層NN



$T_1 = f\left(X, W^{(1)}\right), T_2 = g\left(T_1, W^{(2)}\right), Y = h\left(T_2, W^{(3)}\right), L = loss(Y, T)$ 合成関数

連鎖律を用いた微分

- 例: 変数x_1, x_2, x_3からyを計算

$$s_1 = x_1 x_2 \quad t_1 = s_1 s_2$$
$$s_2 = x_1 x_3 \quad t_2 = s_1 s_3 \quad y = t_1^2 + t_2^2 + t_3^2$$
$$s_3 = x_2 x_3 \quad t_3 = s_2 s_3$$

局所的に微分可能

つまり、
$y = (x_1^2 x_2 x_3)^2 + (x_1 x_2^2 x_3)^2 + (x_1 x_2 x_3^2)^2$

$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial t_1}\frac{\partial t_1}{\partial x_1} + \frac{\partial y}{\partial t_2}\frac{\partial t_2}{\partial x_1} + \frac{\partial y}{\partial t_3}\frac{\partial t_3}{\partial x_1}$$
$$= 2t_1\frac{\partial t_1}{\partial x_1} + 2t_2\frac{\partial t_2}{\partial x_1} + 2t_3\frac{\partial t_3}{\partial x_1}$$

再帰的微分

$$= 2t_1\left(\frac{\partial t_1}{\partial s_1}\frac{\partial s_1}{\partial x_1} + \frac{\partial t_1}{\partial s_2}\frac{\partial s_2}{\partial x_1}\right) + 2t_2\left(\frac{\partial t_2}{\partial s_1}\frac{\partial s_1}{\partial x_1} + \frac{\partial t_2}{\partial s_3}\frac{\partial s_3}{\partial x_1}\right) + 2t_3\left(\frac{\partial t_3}{\partial s_2}\frac{\partial s_2}{\partial x_1} + \frac{\partial t_3}{\partial s_3}\frac{\partial s_3}{\partial x_1}\right)$$
$$= 2t_1\left(s_2\frac{\partial s_1}{\partial x_1} + s_1\frac{\partial s_2}{\partial x_1}\right) + 2t_2\left(s_3\frac{\partial s_1}{\partial x_1} + s_1\frac{\partial s_3}{\partial x_1}\right) + 2t_3\left(s_3\frac{\partial s_2}{\partial x_1} + s_2\frac{\partial s_3}{\partial x_1}\right)$$
$$= 2t_1(s_2x_2 + s_1x_3) + 2t_2(s_3x_2) + 2t_3(s_3x_3)$$
$$= 2x_1^2x_2x_3(2x_1x_2x_3) + 2x_1x_2^2x_3(x_2^2x_3) + 2x_1x_2x_3^2(x_2x_3^2)$$

しかし、$\frac{\partial s_1}{\partial x_1}$と$\frac{\partial s_2}{\partial x_1}$と$\frac{\partial s_3}{\partial x_1}$は同じ計算を2回繰り返している

誤差逆伝播法(Back Propagation)

- **損失に対する各パラメータの微分を計算する方法**
  - 次の計算ステップで全てのパラメータの微分を計算

1. 計算グラフを作る
2. 入力から出力に向かって順方向に計算(順伝播)
3. 出力から入力に向かって逆方向に微分を計算(逆伝播)

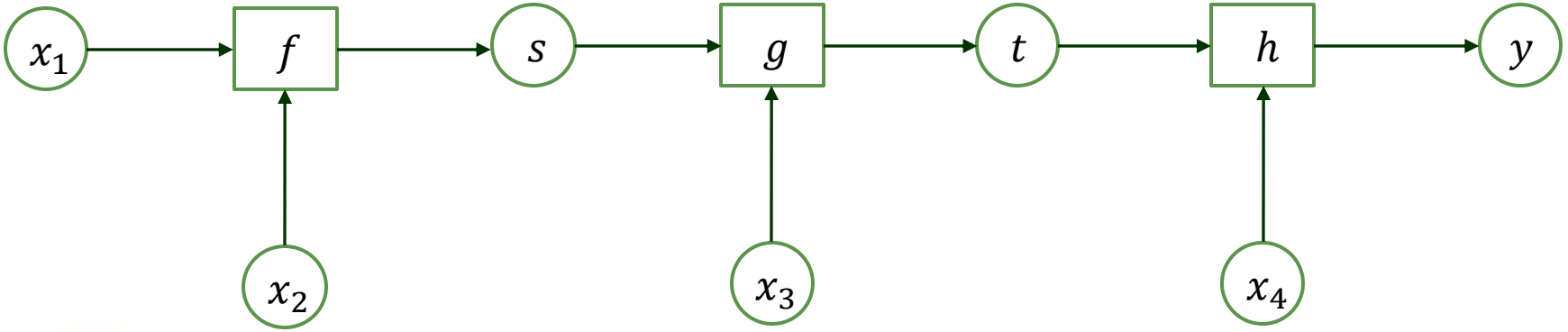
  - 同じ微分の計算を繰り返さない(動的計画法の一種)

計算グラフによる解釈

- **計算グラフ：計算の過程をグラフ（ノードとエッジ）で表したもの**

○：変数
□：関数（演算の中身）

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ

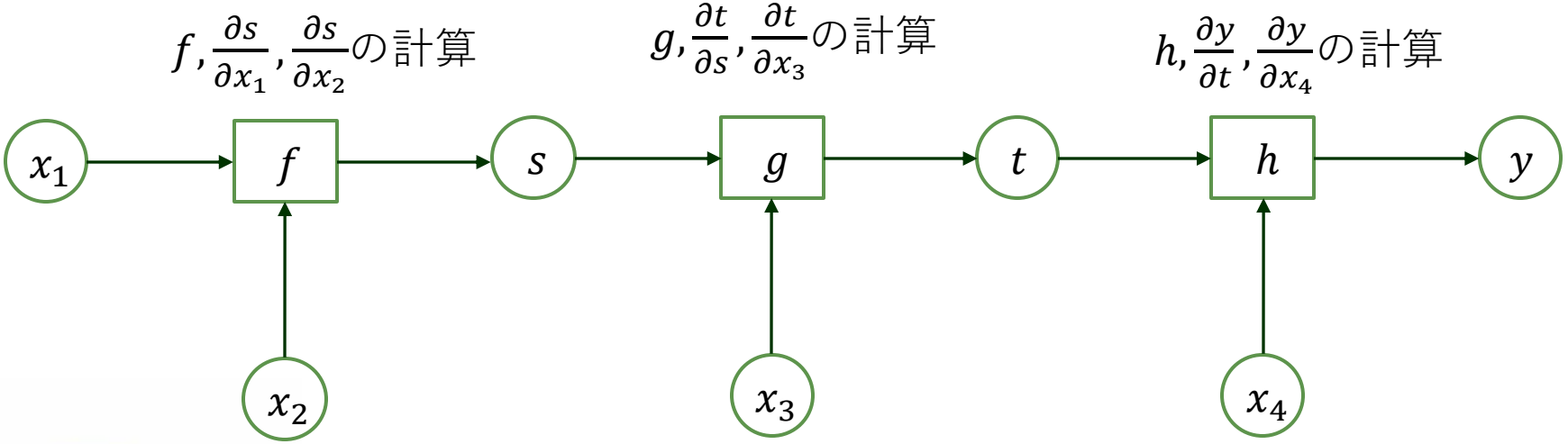


計算グラフによる解釈

- **計算グラフ：計算の過程をグラフ（ノードとエッジ）で表したもの**

○：変数
□：関数（演算の中身）

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ



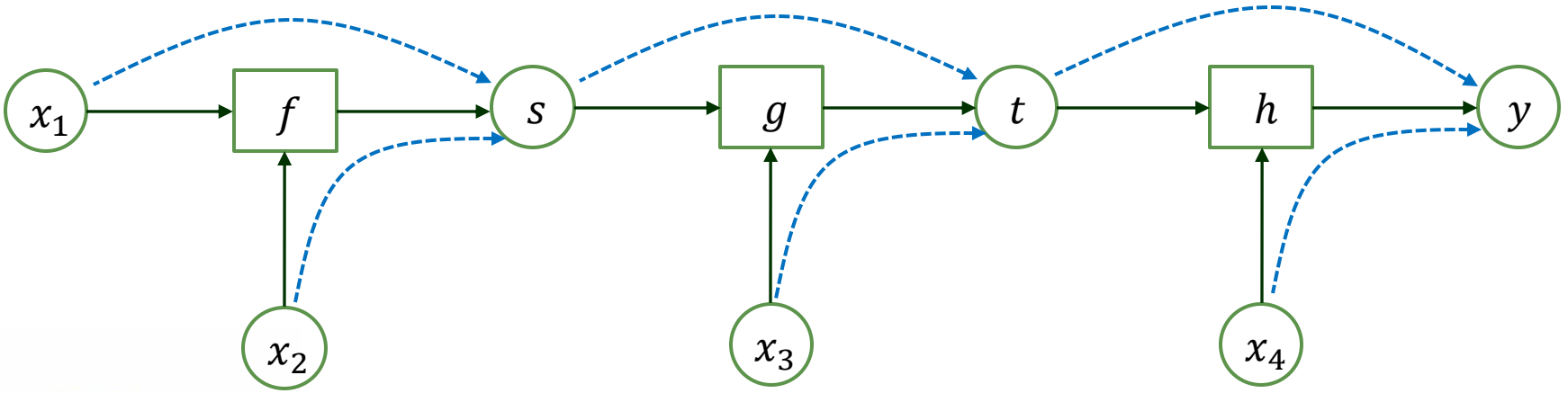
各関数は
・入力を受け取ったときの出力の計算
・入力となる各変数に対する偏微分の計算
ができれば良い

計算グラフによる解釈

- **順伝播の計算**

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ

○：変数
□：関数（演算の中身）



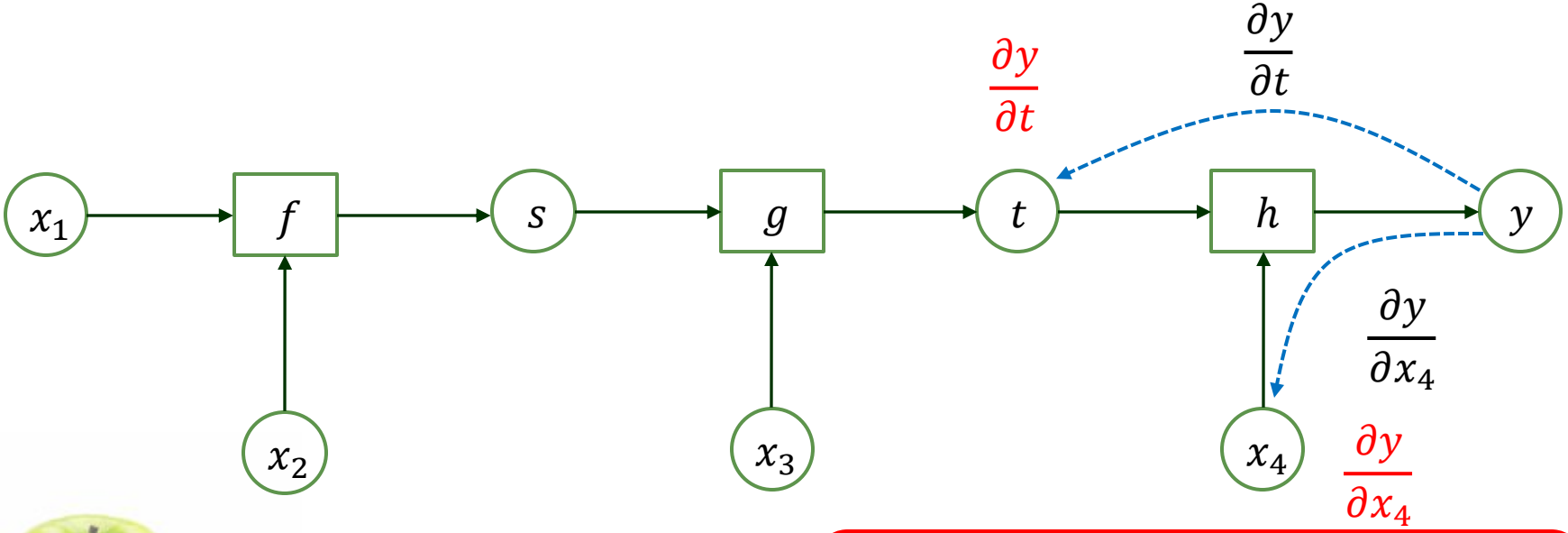
全体の入力$(x_1、\ x_2、\ x_3、x_4)$を与えて、すべてのノードの値(s, t, y)を計算する。

計算グラフによる解釈

- **逆伝播の計算**

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ

○：変数
□：関数（演算の中身）



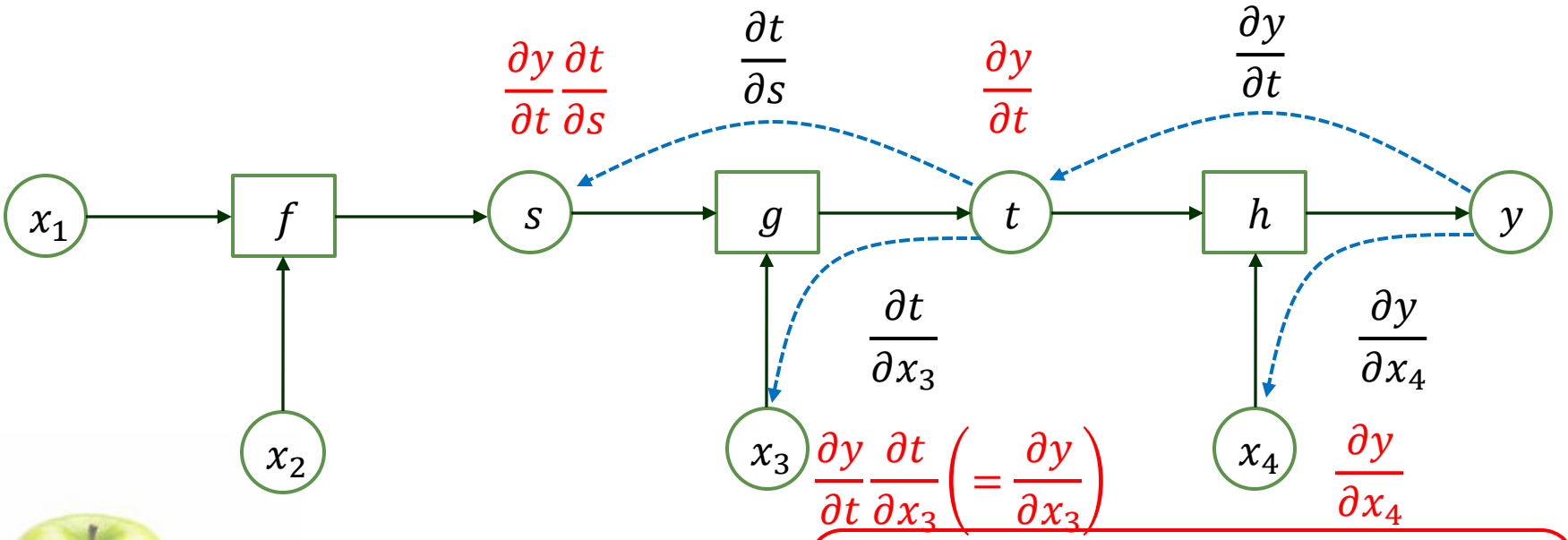
誤差逆伝播法：上流の関数から伝達された値に局所的な微分を乗算して前のノードへ渡していく

計算グラフによる解釈

- **逆伝播の計算**

○：変数
□：関数（演算の中身）

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ



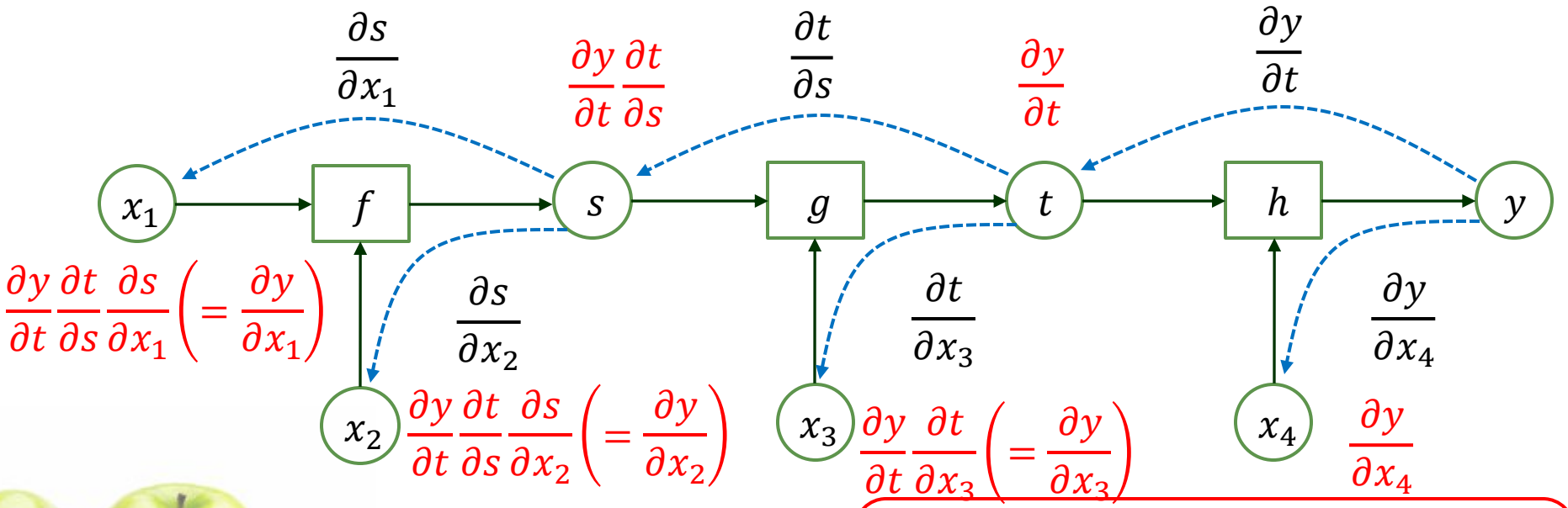
誤差逆伝播法：上流の関数から伝達された値に局所的な微分を乗算して前のノードへ渡していく

計算グラフによる解釈

- **逆伝播の計算**

○：変数
□：関数（演算の中身）

$s = f(x_1, x_2), t = g(s, x_3), y = h(t, x_4)$の計算グラフ



yの勾配$\left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial y}{\partial x_3}, \frac{\partial y}{\partial x_4}\right)$の計算できた!

誤差逆伝播法：上流の関数から伝達された値に局所的な微分を乗算して前のノードへ渡していく

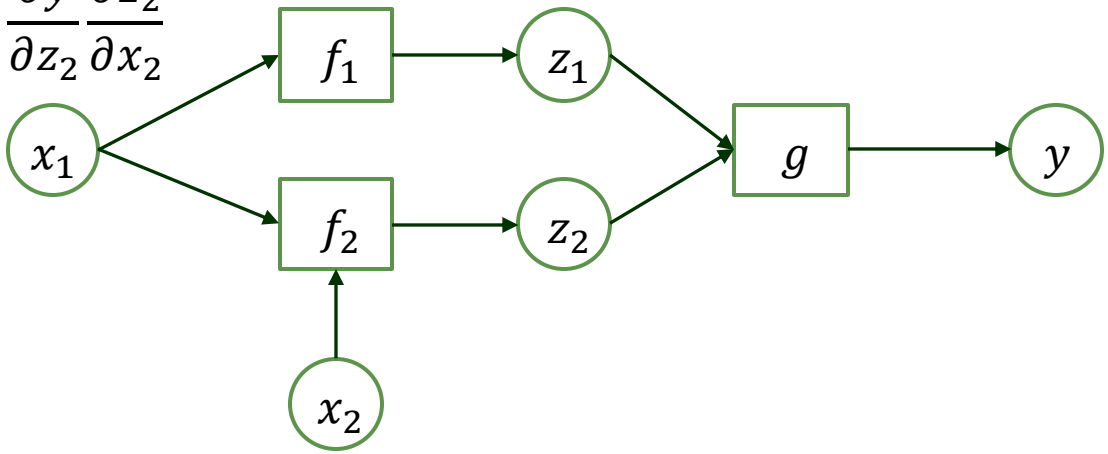
連鎖律を用いた微分2

- 一般の数式 (順伝播で再合流する場合)
  - x_1, x_2からyを計算する過程

$$z_1 = f_1(x_1)$$
$$z_2 = f_2(x_1, x_2)$$
$$y = g(z_1, z_2)$$

$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

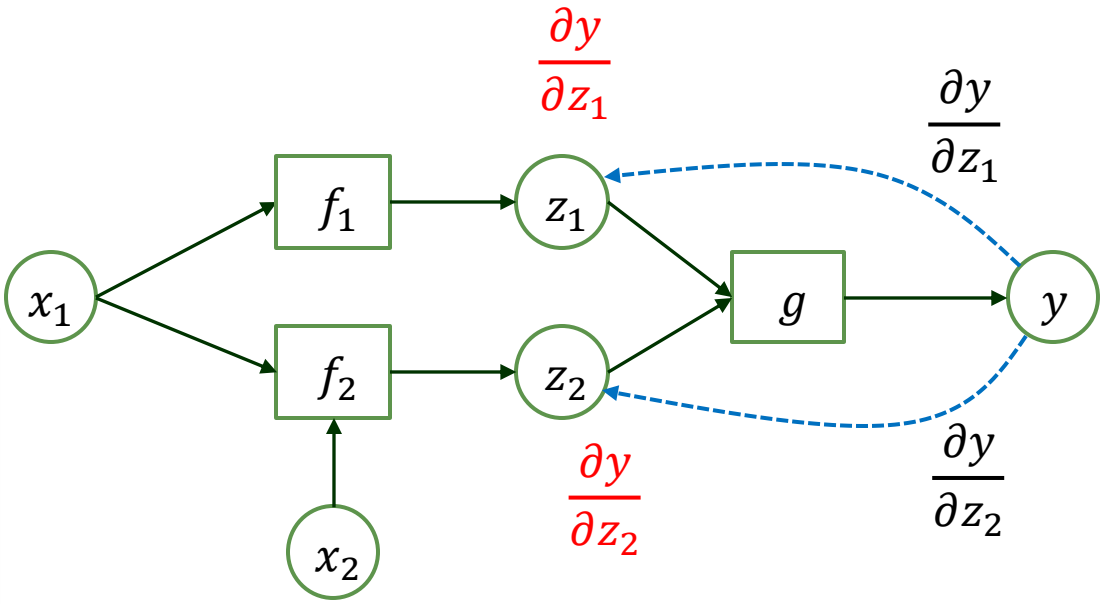
$$\frac{\partial y}{\partial x_2} = \underbrace{\frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_2}}_{=0} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_2} = \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



計算グラフによる解釈2

$$z_1 = f_1(x_1), z_2 = f_2(x_1, x_2), y = g(z_1, z_2)$$

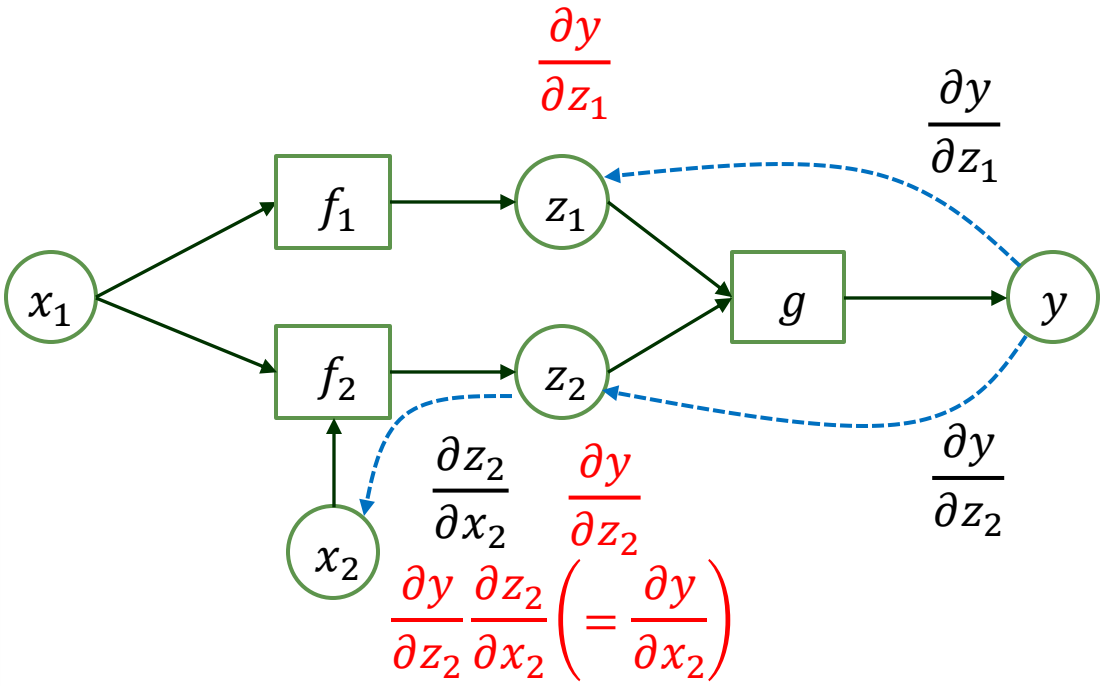
$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_1}, \qquad \frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



計算グラフによる解釈2

$$z_1 = f_1(x_1), z_2 = f_2(x_1, x_2), y = g(z_1, z_2)$$

$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_1}, \qquad \frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



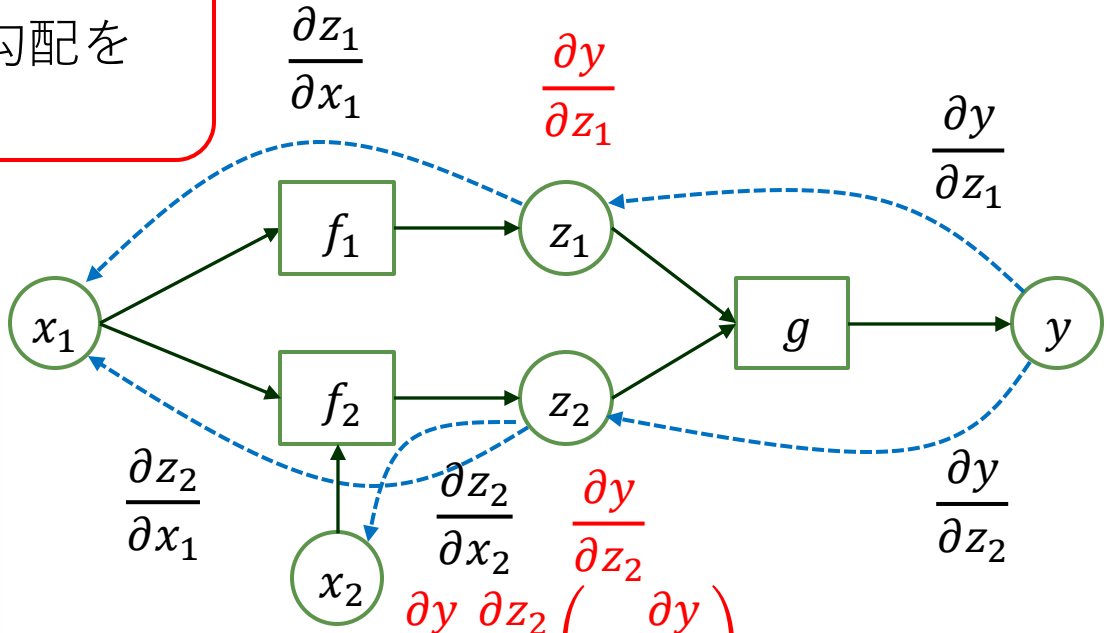
計算グラフによる解釈2

$$z_1 = f_1(x_1), z_2 = f_2(x_1, x_2), y = g(z_1, z_2)$$

$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_1}, \qquad \frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$

合流地点では各パスから伝播してきた勾配を足し合わせる

$$\frac{\partial y}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial x_1} \left(= \frac{\partial y}{\partial x_1}\right)$$



【参考】計算グラフ上で偏微分の計算ができる理由

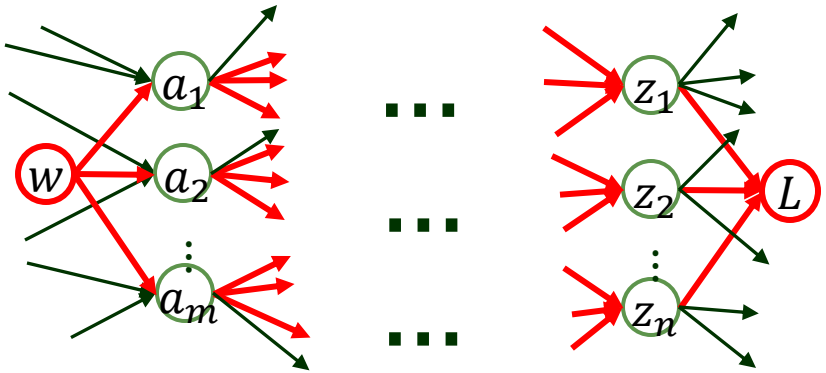
- **損失Lとあるパラメータwの偏微分$\frac{\partial L}{\partial w}$の計算**

  計算グラフ上の変数にだけ注目して、Lとwをつなぐ全てのパス集合をPとおくと、

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w} + \cdots + \frac{\partial L}{\partial z_n}\frac{\partial z_n}{\partial w} = \sum_{p \in P}\prod_{(u,v)\in P}\frac{\partial v}{\partial u}$$

  となり、全体の微分は各パスの微分の和となり、各パスの微分は局所的な微分の積となっている。従って、右から左に計算しても左から右に計算しても同じ。

$$\sum_{p \in P}\prod_{(u,v)\in P}\frac{\partial v}{\partial u} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial w} + \cdots + \frac{\partial L}{\partial a_m}\frac{\partial a_m}{\partial w}$$

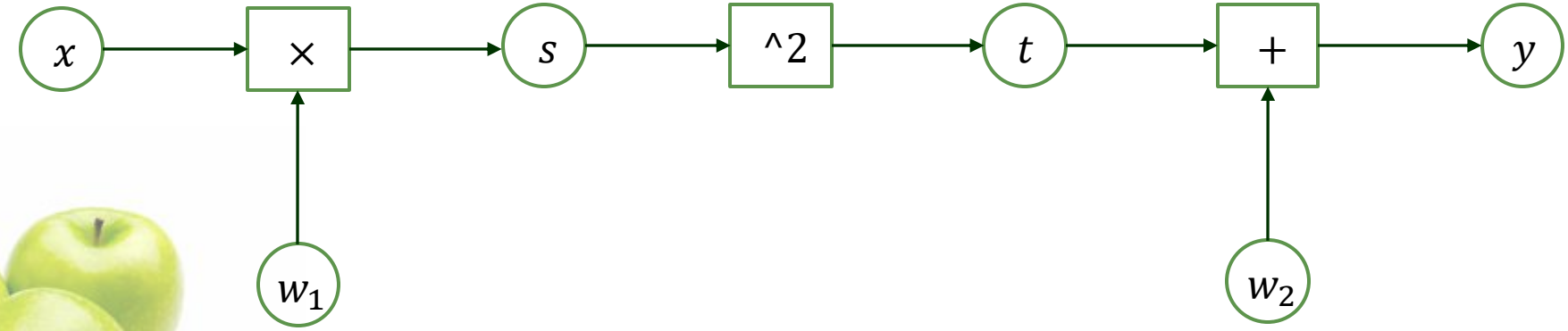


誤差逆伝播法の例

入力x、変数w_1, w_2から次のように出力値yを計算する過程を考える

$$s = x \times w_1$$
$$t = s^2$$
$$y = t + w_2$$

入力$x = 2$、変数$w_1 = 3\ w_2 = 4$における変数w_1, w_2の勾配$\nabla y = \left(\frac{\partial y}{\partial w_1}, \frac{\partial y}{\partial w_2}\right)$を求める



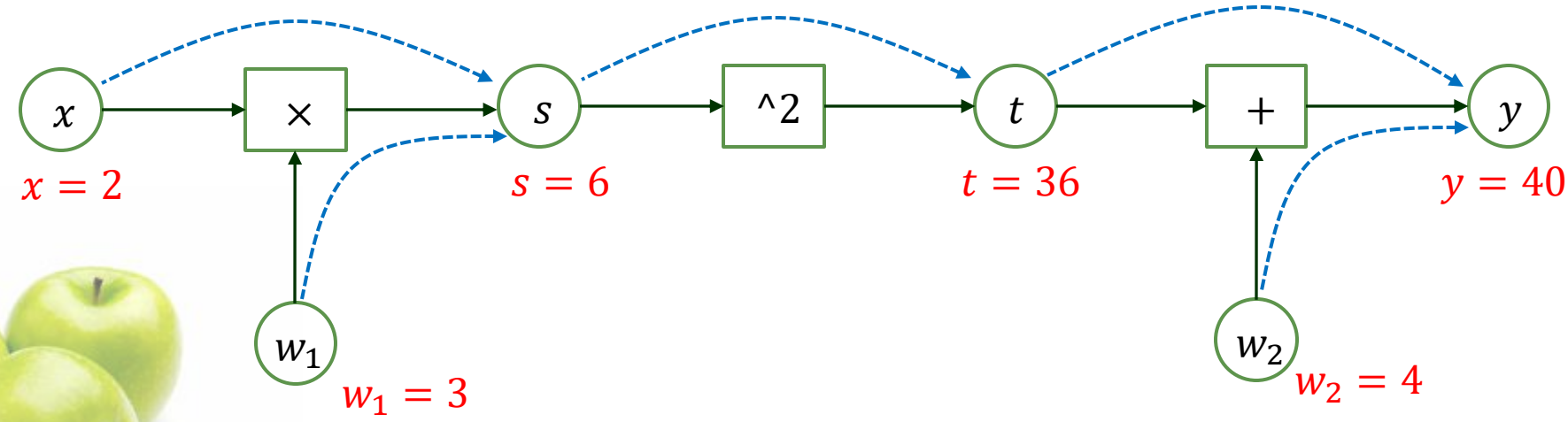
誤差逆伝播法の例

入力x、変数w_1, w_2から次のように出力値yを計算する過程を考える

$$s = x \times w_1$$
$$t = s^2$$
$$y = t + w_2$$

入力$x = 2$、重み変数$w_1 = 3, w_2 = 4$における重み変数w_1, w_2 の勾配$\nabla y = \left(\frac{\partial y}{\partial w_1}, \frac{\partial y}{\partial w_2}\right)$を求める

順伝播の計算



誤差逆伝播法の例

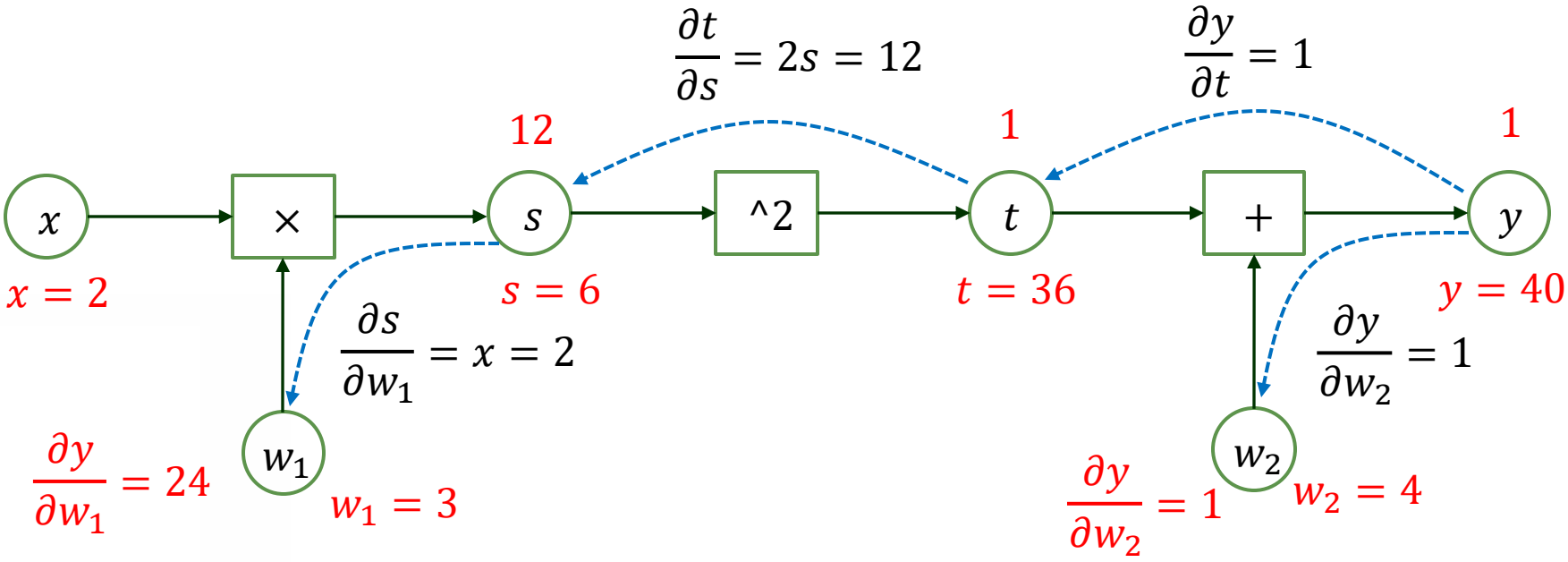
入力x、変数w_1, w_2から次のように出力値yを計算する過程を考える

$$s = x \times w_1, \qquad t = s^2, \qquad y = t + w_2$$

入力$x = 2$、重み変数$w_1 = 3, w_2 = 4$における重み変数w_1, w_2 の勾配

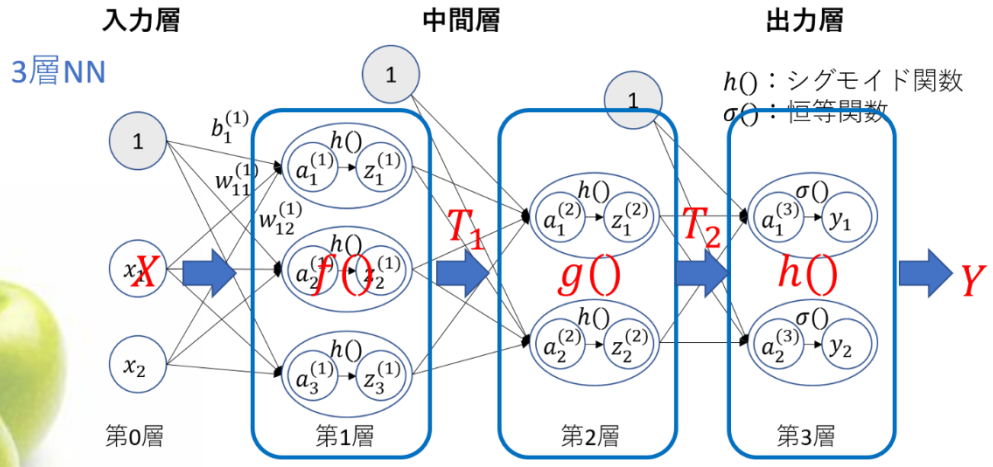
$\nabla y = \left(\frac{\partial y}{\partial w_1}, \frac{\partial y}{\partial w_2}\right)$を求める

逆伝播の計算



誤差逆伝播法による勾配の求め方のまとめ

1. **計算グラフを作成**
2. **順伝播の計算**
3. **逆伝播の計算**
   - 出力から入力へ、上流から伝播してきた勾配に局所的な微分を乗算して下流のノードに伝播
   - 合流地点では各パスから伝播してきた勾配を足し合わせる



各層の誤差逆伝播時の動作を定義すればそれらをつなぐことで誤差逆伝播を実行可能

誤差逆伝播法によるNN学習の全体像

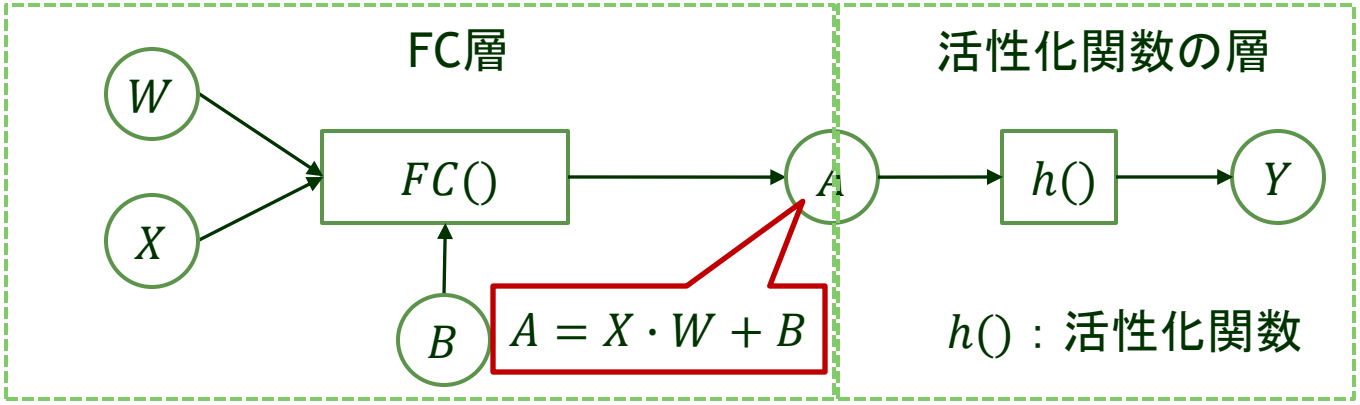
損失を最小にする重みパラメータを探す

1. 勾配の算出：損失に対する各重みパラメータの勾配を求める（誤差逆伝播法により算出）
2. パラメータの更新：重みパラメータを勾配方向とは逆方向に微小量更新する
3. 繰り返す：ステップ1に戻る

【参考】NN中間層の逆伝播

- 全結合層(Fully Connected Layer; FC)(Affine層ともいう)＋活性化関数の逆伝播

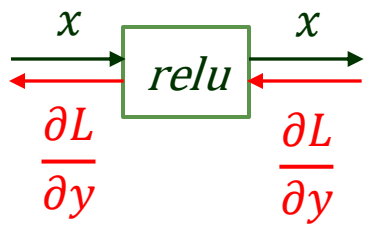
隠れ層の計算グラフ



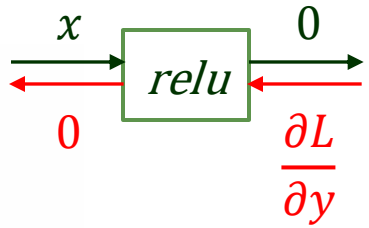
【参考】活性化関数の逆伝播：ReLU関数

$$y = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases}, \qquad \frac{\partial y}{\partial x} = \begin{cases} 1 & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

$x > 0$の時



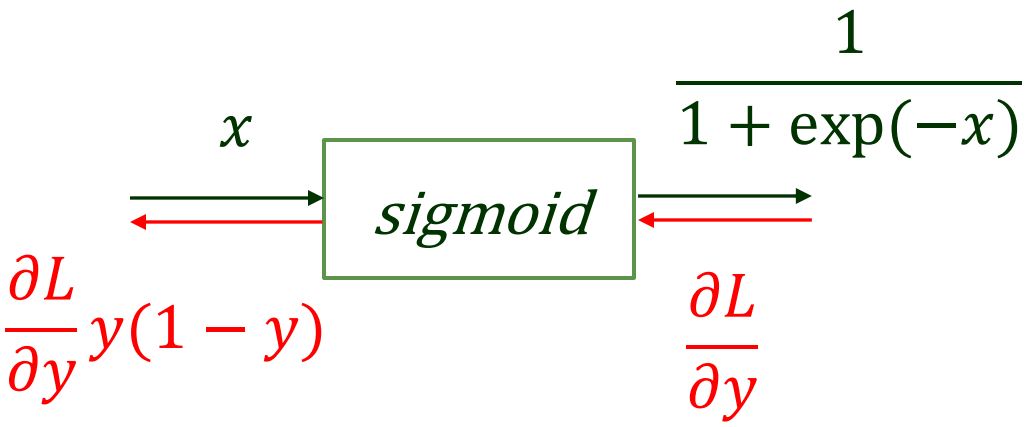
$x \leq 0$の時



順伝播時の入力xが0より大きい場合、
上流の値をそのまま下流に流す
0以下の場合、何も流さない

【参考】活性化関数の逆伝播：シグモイド関数

$$y = \frac{1}{1 + \exp(-x)} \quad , \quad \frac{\partial y}{\partial x} = y(1 - y)$$



※順伝播の出力yを用いて逆伝播の出力値を算出

【参考】シグモイド関数の微分

$y = \dfrac{1}{1+e^{-x}}$ の微分

$u = 1 + e^{-x}$ とおくと

$$y = \frac{1}{u}$$

$$\frac{dy}{du} = -\frac{1}{u^2}$$

$$\frac{du}{dx} = -e^{-x}$$

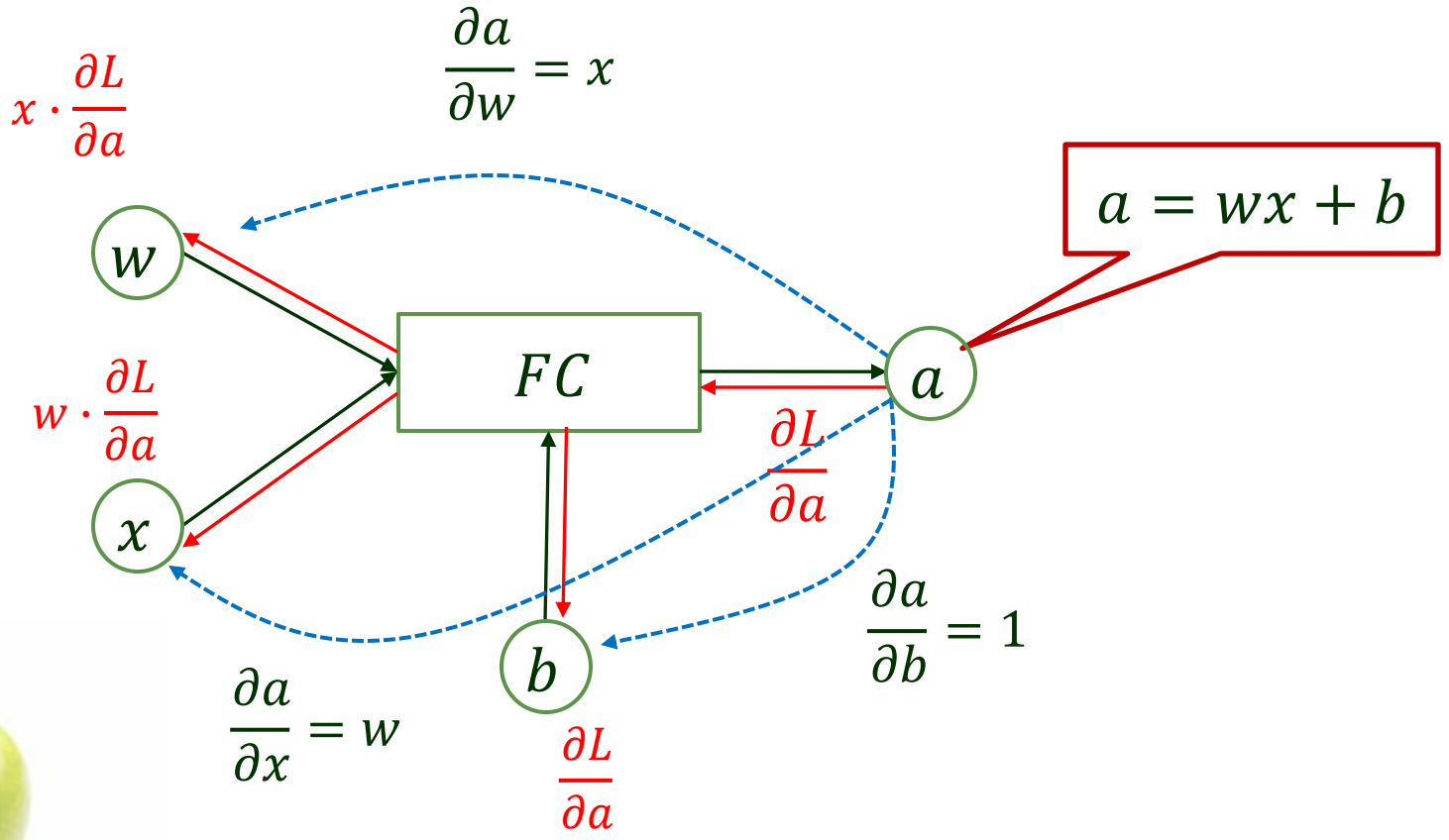
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$$= \frac{e^{-x}}{u^2} = \frac{e^{-x}}{(1+e^{-x})^2}$$

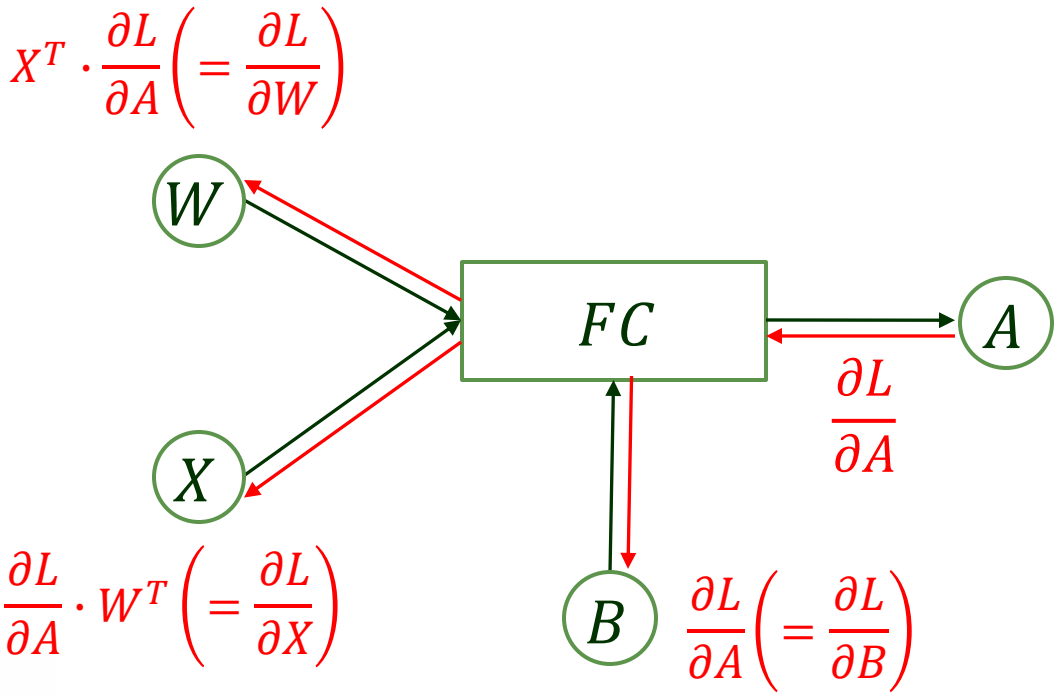
$$= \left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right)\frac{1}{1+e^{-x}}$$

$$= (1-y)y$$

【参考】全結合(FC)層の逆伝播 (1/2)



【参考】全結合(FC)層の逆伝播 (2/2)

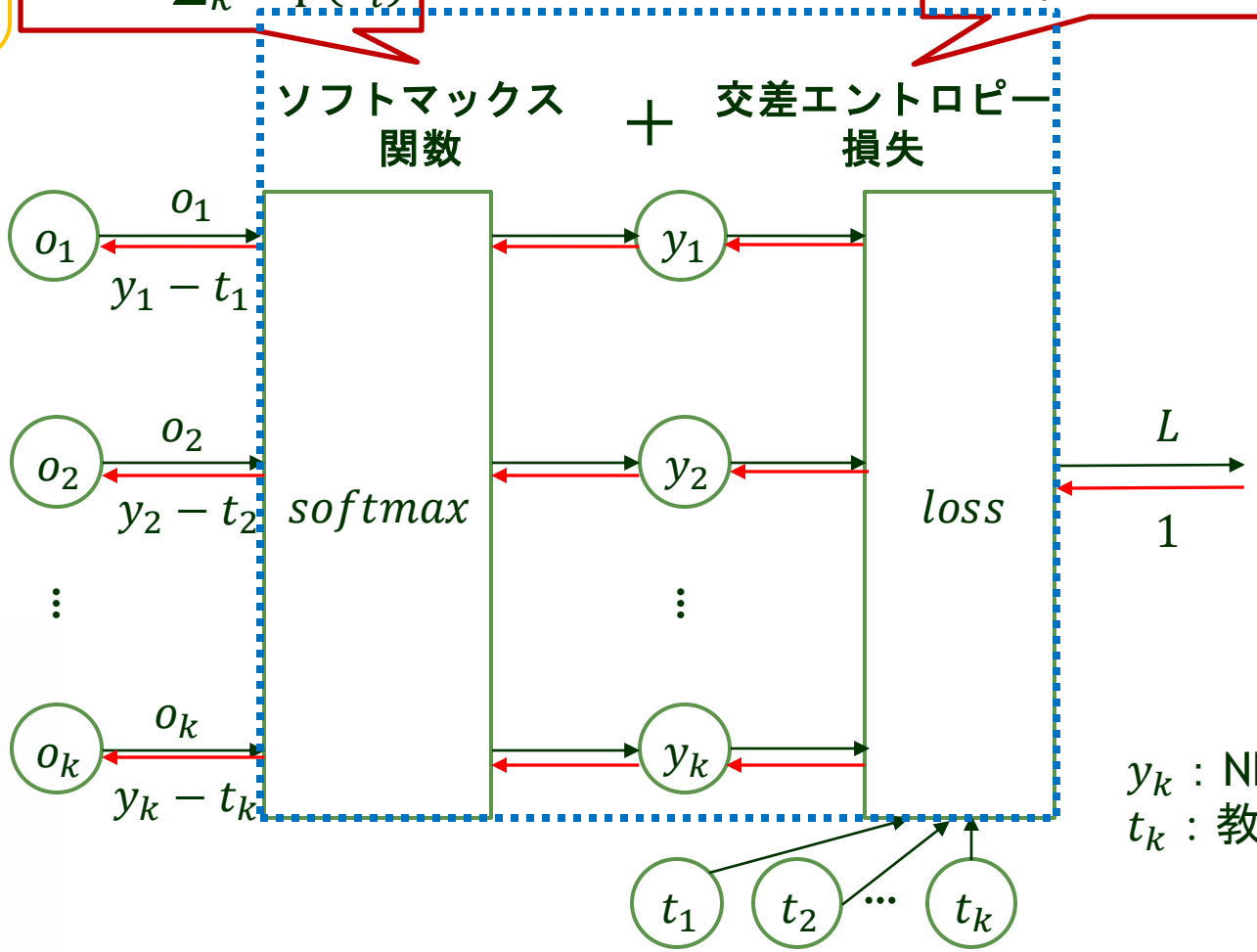


【参考】出力層の逆伝播：ソフトマックス層＋交差エントロピー損失

分類問題で使用

$$y_i = \frac{\exp(o_i)}{\sum_k \exp(o_i)}$$

$$-\sum_k t_k log_e y_k$$



y_k：NNの出力、
t_k：教師データ

【参考】ソフトマックス層＋交差エントロピー損失の逆伝播の導出

$$y_i = \frac{\exp(o_i)}{\sum_k \exp(o_i)}、\quad L = -\sum_k t_k log_e y_k \quad において\frac{\partial L}{\partial o_i}を求める$$

$$\frac{\partial L}{\partial o_i} = \sum_c \frac{\partial L}{\partial y_c} \cdot \frac{\partial y_c}{\partial o_i}$$

$$= \frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial o_i} + \sum_{c \neq i} \frac{\partial L}{\partial y_c} \cdot \frac{\partial y_c}{\partial o_i}$$

着目クラス　それ以外

$$= -\frac{t_i}{y_i} \cdot y_i(1 - y_i) - \sum_{c \neq i} \frac{t_c}{y_c} \cdot (-y_i y_c)$$

$$= y_i \cdot t_i - t_i + \sum_{c \neq i} t_c \cdot y_i$$

$$= \sum_c t_c \cdot y_i - t_i = y_i - t_i$$

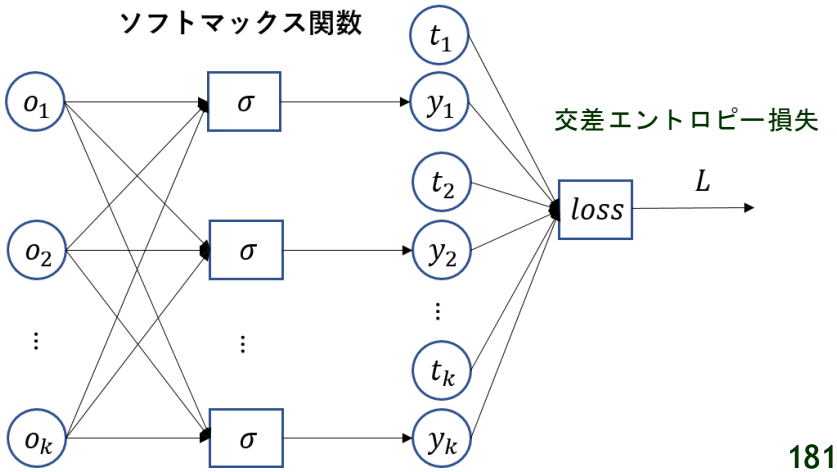
$$\frac{\partial L}{\partial y_c} = -\frac{t_c}{y_c} \quad (\mathrm{c} = 1 \ldots k)$$

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$$

$$\frac{\partial y_i}{\partial o_i} = \frac{\exp(o_i)\sum_k \exp(o_k) - \exp(o_i)\exp(o_i)}{(\sum_k \exp(o_k))^2}$$

$$= y_i - y_i^2 = y_i(1 - y_i)$$

$$\frac{\partial y_c}{\partial o_i} = \frac{-\exp(o_c)\exp(o_i)}{(\sum_k \exp(o_k))^2} = -y_i y_c \quad (c \neq i)$$

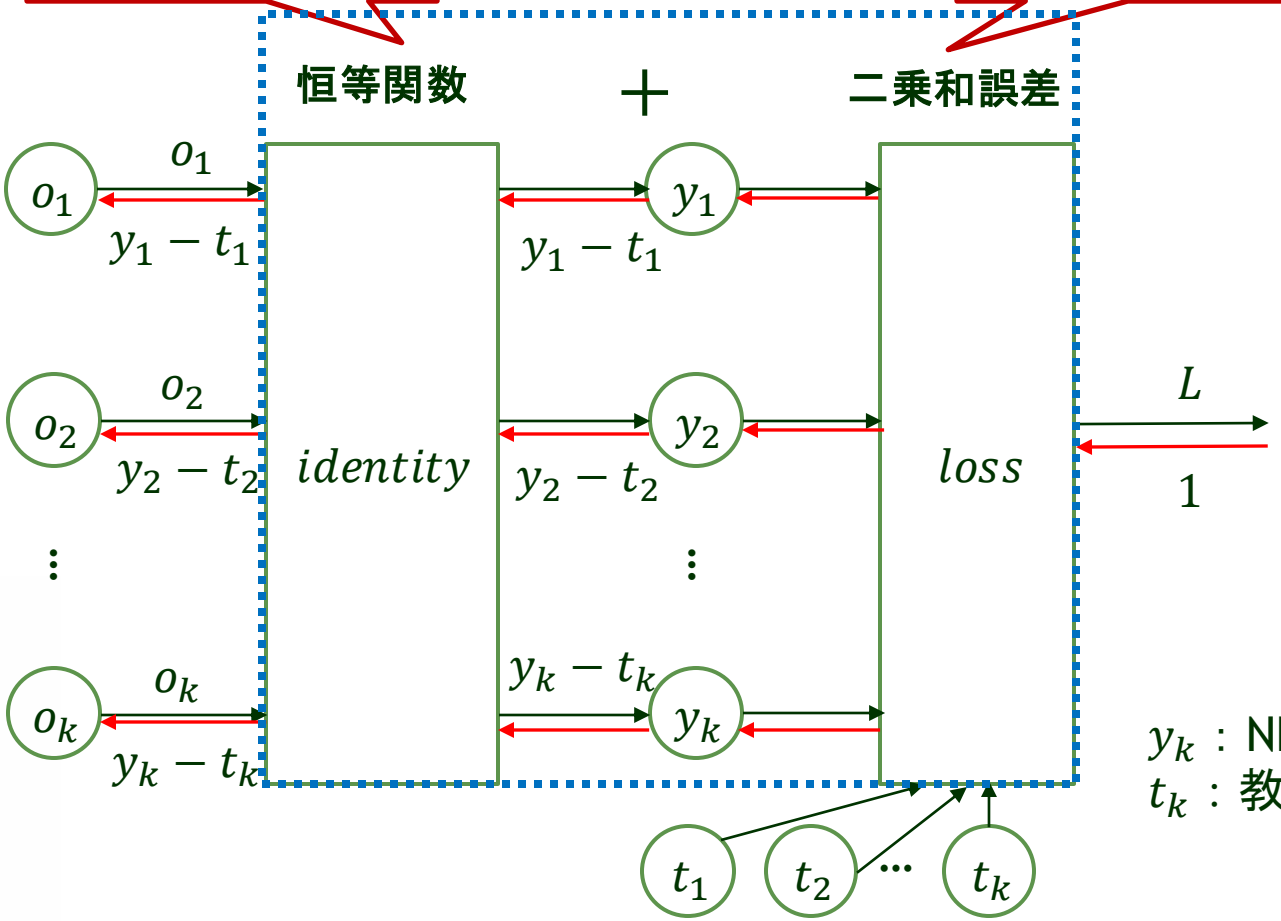


【参考】出力層の逆伝播：恒等関数＋2乗和誤差

回帰問題で使用

入力をそのまま流す

$\frac{1}{2}\sum_k (y_k - t_k)^2$



y_k：NNの出力、
t_k：教師データ

まとめ

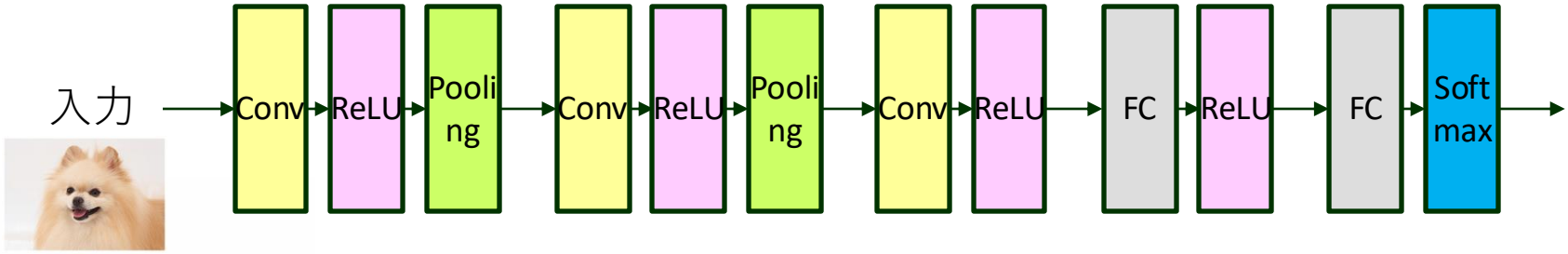
- **NNの推論**
  - NNは入力層、中間層、出力層の多層構造
  - 信号は入力層、中間層、出力層へと階層的に伝わる
  - 中間層、出力層には活性化関数を用いる
- **NNの学習**
  - 損失が最小になる重みパラメータを探す
  - 重みパラメータを勾配に基づき更新する作業を繰り返すことで探す（勾配降下法）
  - 誤差逆伝播により勾配を求める
- **計算グラフにより計算過程を視覚的に把握できる**
  - ノードは局所的な計算で構成
  - 順伝播：通常の計算、逆伝播：各ノードの微分（出力からのパスの積）を求められる

畳み込みニューラルネットワーク (CONVOLUTIONAL NEURAL NETWORK; CNN)

畳み込みニューラルネットワーク(Convolutional Neural Network; CNN)

- **入力の形状／構造を捉えることに適したネットワーク**
  - 特に画像認識には有効なNN
- **畳み込み層(Convolution Layer)とプーリング層(Pooling Layer)を持つ**

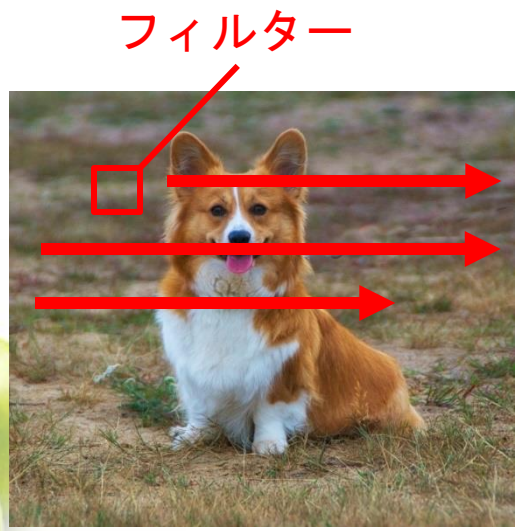
一般的なCNN



- Conv→ReLU(→Pooling)が典型的な要素
- 出力層に近くなるとFC→ReLU
- 出力層はFC→Softmax

畳み込み層

- **重み付きフィルターによる画像処理**
  - 全ての画素に対しフィルターをかけて画像全体になんらかの効果を与える
  - フィルターには重みがついていて、周りの画素情報から中心の画素の計算を行う
  - 通常のフィルターは人手で与えたものを用いるが、CNNではフィルターを自動的に学習する



畳込み層

- **データの次元数を維持する**
  - 画像の場合、3次元（縦、横、チャネル）のデータを入力として受け取り3次元のデータを出力して次の層に渡す
  - フィルターによる畳み込み演算を行う（3x3のフィルターの場合）

$$I'_{i,j} = \sum_{a=-1}^{1} \sum_{b=-1}^{1} w_{a,b} I_{i+a,j+b}$$

$I_{i,j}$

1	2	3	0	1
0	1	2	3	0
3	0	1	2	3
2	3	0	1	2
1	2	3	0	1

入力画像(5,5)

$\circledast$

$w_{a,b}$

2	0	1
0	1	2
1	0	2

フィルター(3,3)
（畳み込み層の重みパラメータ）

$I'_{i,j}$

15		

出力画像(3,3)

畳込み層

- **データの次元数を維持する**
  - 画像の場合、3次元（縦、横、チャネル）のデータを入力として受け取り3次元のデータを出力して次の層に渡す
  - フィルターによる畳み込み演算を行う（3x3のフィルターの場合）

$$I'_{i,j} = \sum_{a=-1}^{1} \sum_{b=-1}^{1} w_{a,b} I_{i+a,j+b}$$

$I_{i,j}$

1	2	3	0	1
0	1	2	3	0
3	0	1	2	3
2	3	0	1	2
1	2	3	0	1

入力画像(5,5)

$\circledast$

$w_{a,b}$

2	0	1
0	1	2
1	0	2

フィルター(3,3)
（畳み込み層の重みパラメータ）

$I'_{i,j}$

15	16	

出力画像(3,3)

畳込み層

- **データの次元数を維持する**
  - 画像の場合、3次元（縦、横、チャネル）のデータを入力として受け取り3次元のデータを出力して次の層に渡す
  - フィルターによる畳み込み演算を行う（3x3のフィルターの場合）

$$I'_{i,j} = \sum_{a=-1}^{1} \sum_{b=-1}^{1} w_{a,b} I_{i+a,j+b}$$

$I_{i,j}$

1	2	3	0	1
0	1	2	3	0
3	0	1	2	3
2	3	0	1	2
1	2	3	0	1

入力画像(5,5)

$\circledast$

$w_{a,b}$

2	0	1
0	1	2
1	0	2

フィルター(3,3)
（畳み込み層の重みパラメータ）

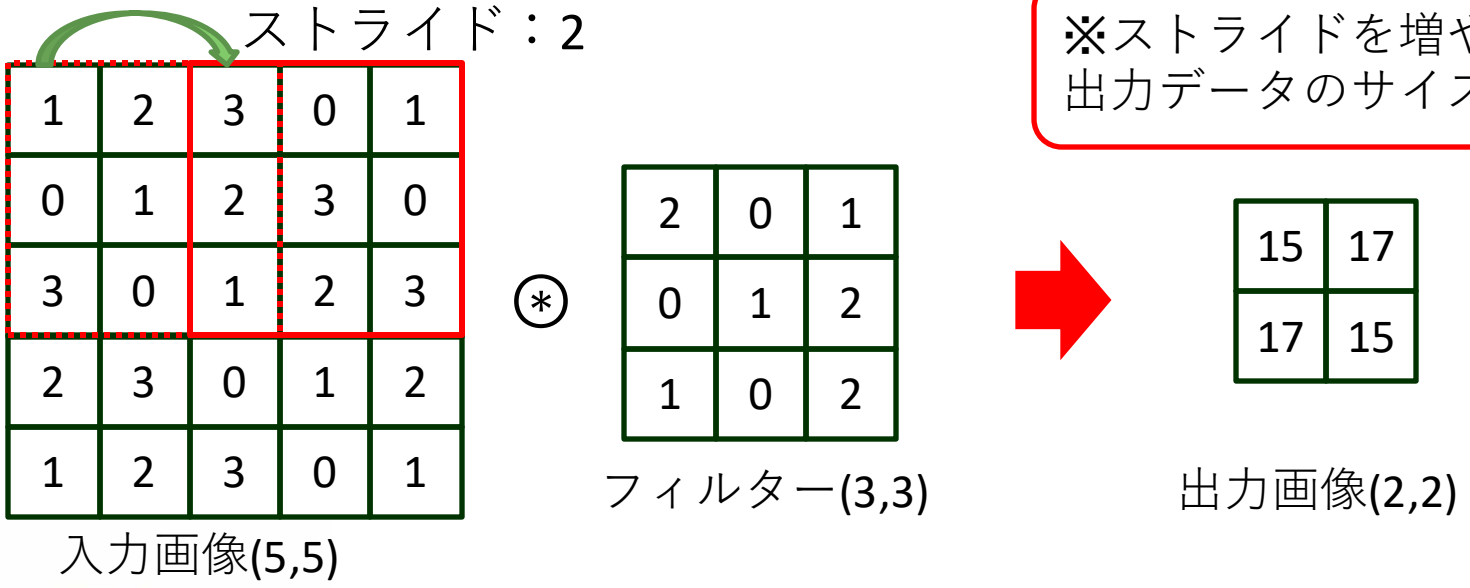
$I'_{i,j}$

15	16	17
6	15	16
17	6	15

出力画像(3,3)

ストライド

- **ストライド: フィルターの移動幅**



※ストライドを増やすと
出力データのサイズが減る

畳み込み演算にバイアスを導入

$I_{i,j}$

1	2	3	0	1
0	1	2	3	0
3	0	1	2	3
2	3	0	1	2
1	2	3	0	1

入力画像(5,5)

$\circledast$

$w_{a,b}$

2	0	1
0	1	2
1	0	2

フィルター(3,3)

＋ 4

バイアス

※フィルタに対して
バイアスは１つ

ストライド：1

15	16	17
6	15	16
17	6	15

＋ 4

バイアス

19	20	21
10	19	20
21	10	19

全ての要素に
バイアスを加算

パディング

- 畳み込み層の処理をする度に画像の周辺が1画素ずつ削れていく(小さい画像や深いNNであるときに問題)
- パディング: 入力データの周囲を固定データ(例えば0)で埋めること

幅1のパディング

0	0	0	0	0	0
0	1	2	3	0	0
0	0	1	2	3	0
0	3	0	1	2	0
0	2	3	0	1	0
0	0	0	0	0	0

入力画像(4+2,4+2)

⊛

2	0	1
0	1	2
1	0	2

フィルター(3,3)

ストライド：1

7	12	10	2
4	15	16	10
10	6	15	6
8	10	4	3

出力画像(4,4)

パディングによって入力画像と
出力画像のサイズが同じになる

【参考】入力サイズと出力サイズの関係

入力データのサイズ（H, W）、フィルターサイズ（FH, FW）
ストライドS、パディングの幅 Pの時

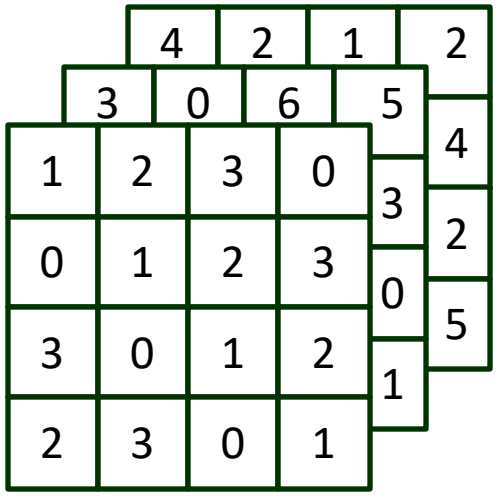
出力データのサイズ（OH, OW）は

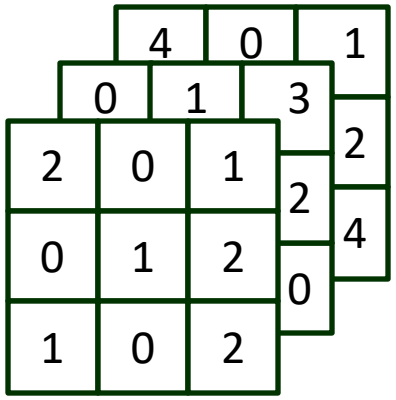
$$OH = \frac{H + 2P - FH}{S} + 1、\quad OW = \frac{W + 2P - FW}{S} + 1$$

3次元データの畳み込み演算(1/2)

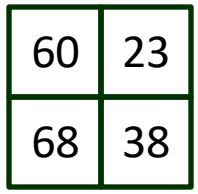
- **画像データは3次元データ(縦、横、チャネル)**

3次元データの畳み込み演算（ストライド：1）

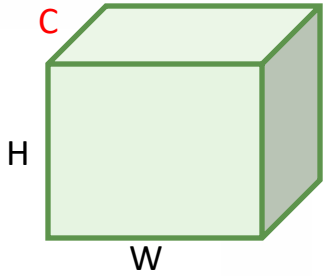


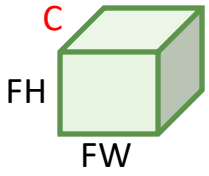


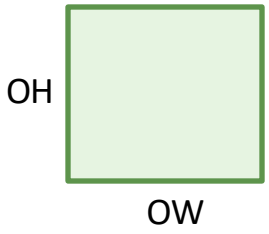
積和演算



※入力データのチャネル数とフィルターのチャネル数は一致させる

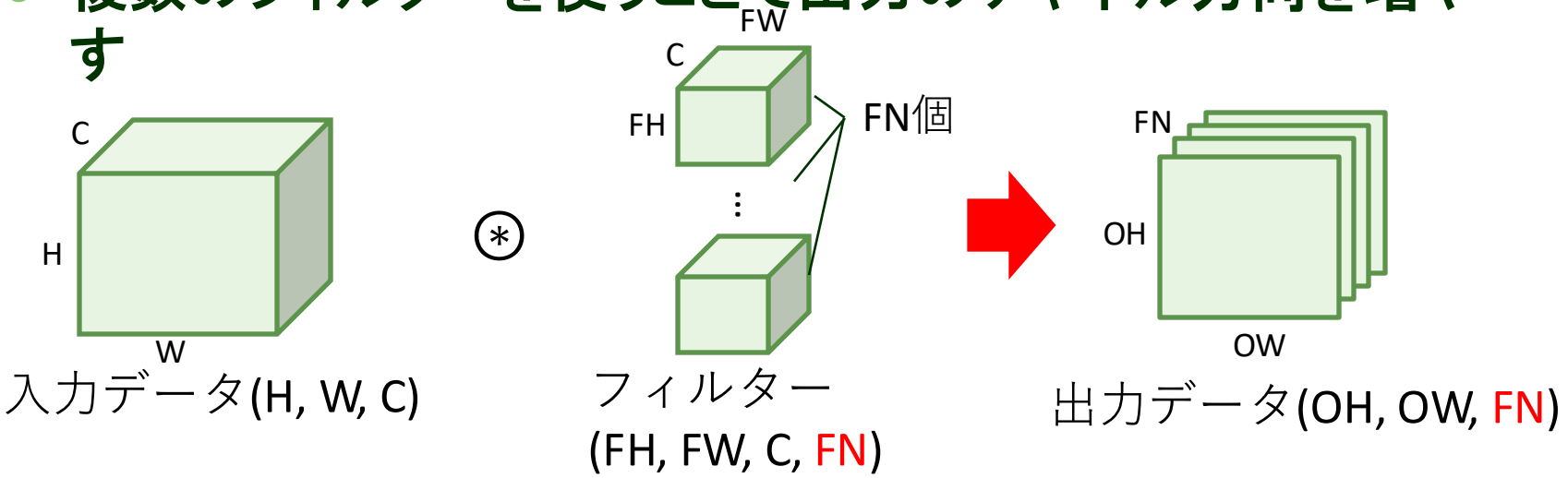




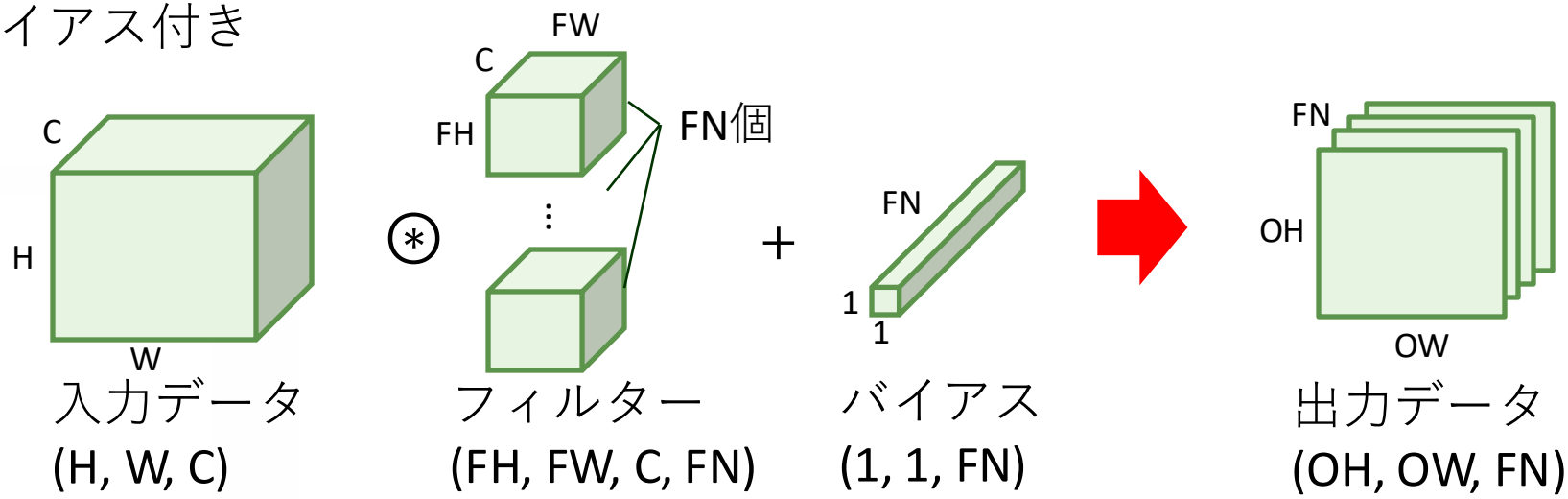


3次元データの畳み込み演算(2/2)

- **複数のフィルターを使うことで出力のチャネル方向を増やす**

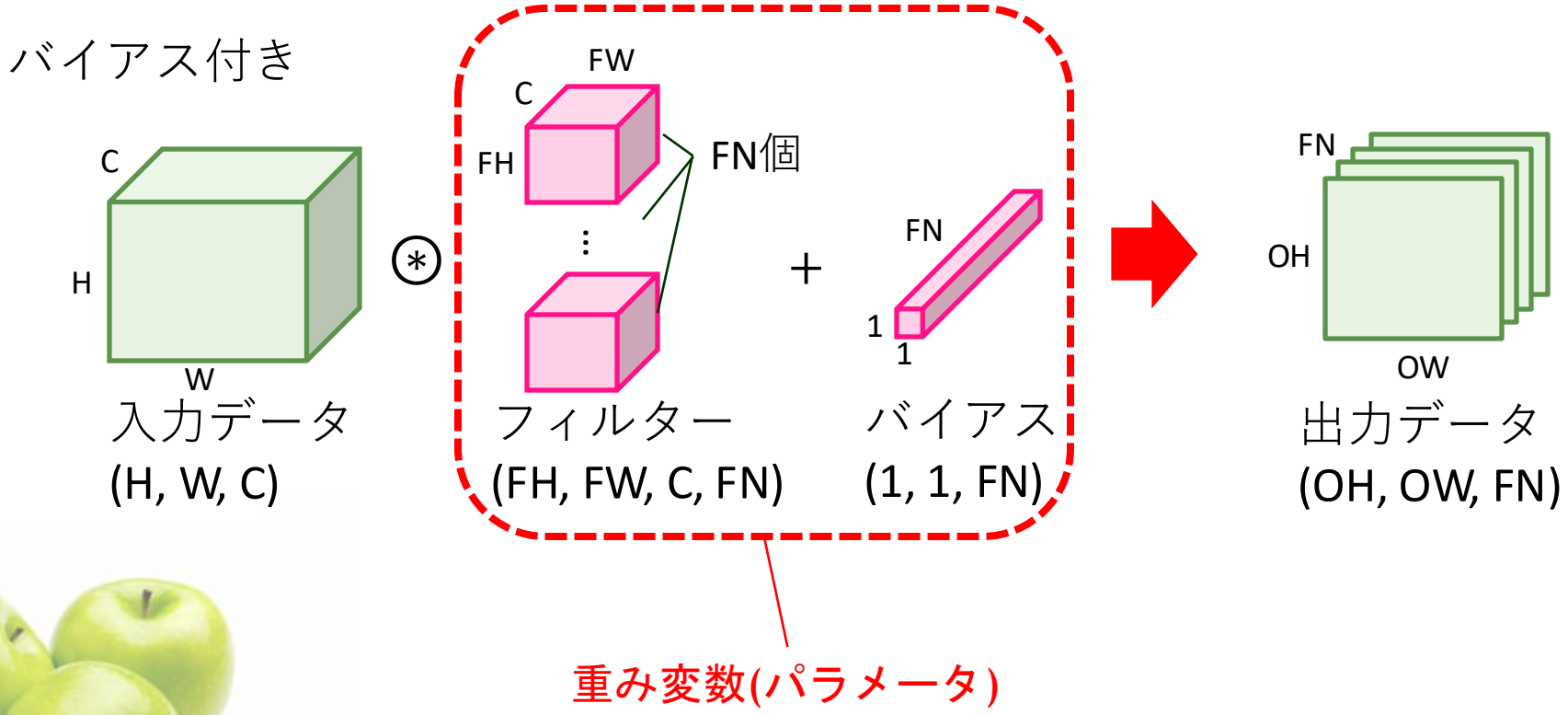


バイアス付き



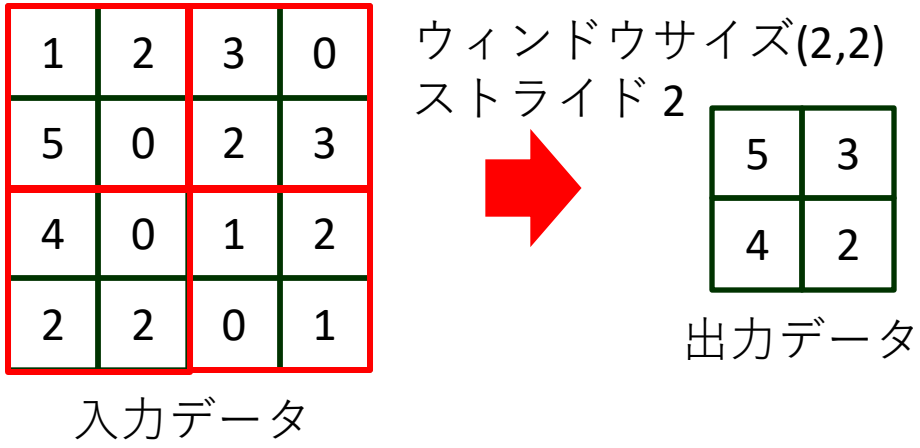
畳み込み層の重み変数(パラメータ)

- フィルターとバイアスが重み変数(パラメータ)



プーリング層

- **縦・横方向の空間を小さくする演算(ダウンサンプリング)**
  - 学習パラメータはない
  - チャネル数は変化しない
- **最大(Max)プーリング**
  - 領域内の最大の要素を取り出す



入力データの微小なズレを吸収する働きがある

Maxプーリング以外のプーリング演算

- **平均(Average)プーリング**
  - 領域内の平均を各要素とする

入力データ

1	2	3	0
5	0	2	3
4	0	1	2
2	2	0	1

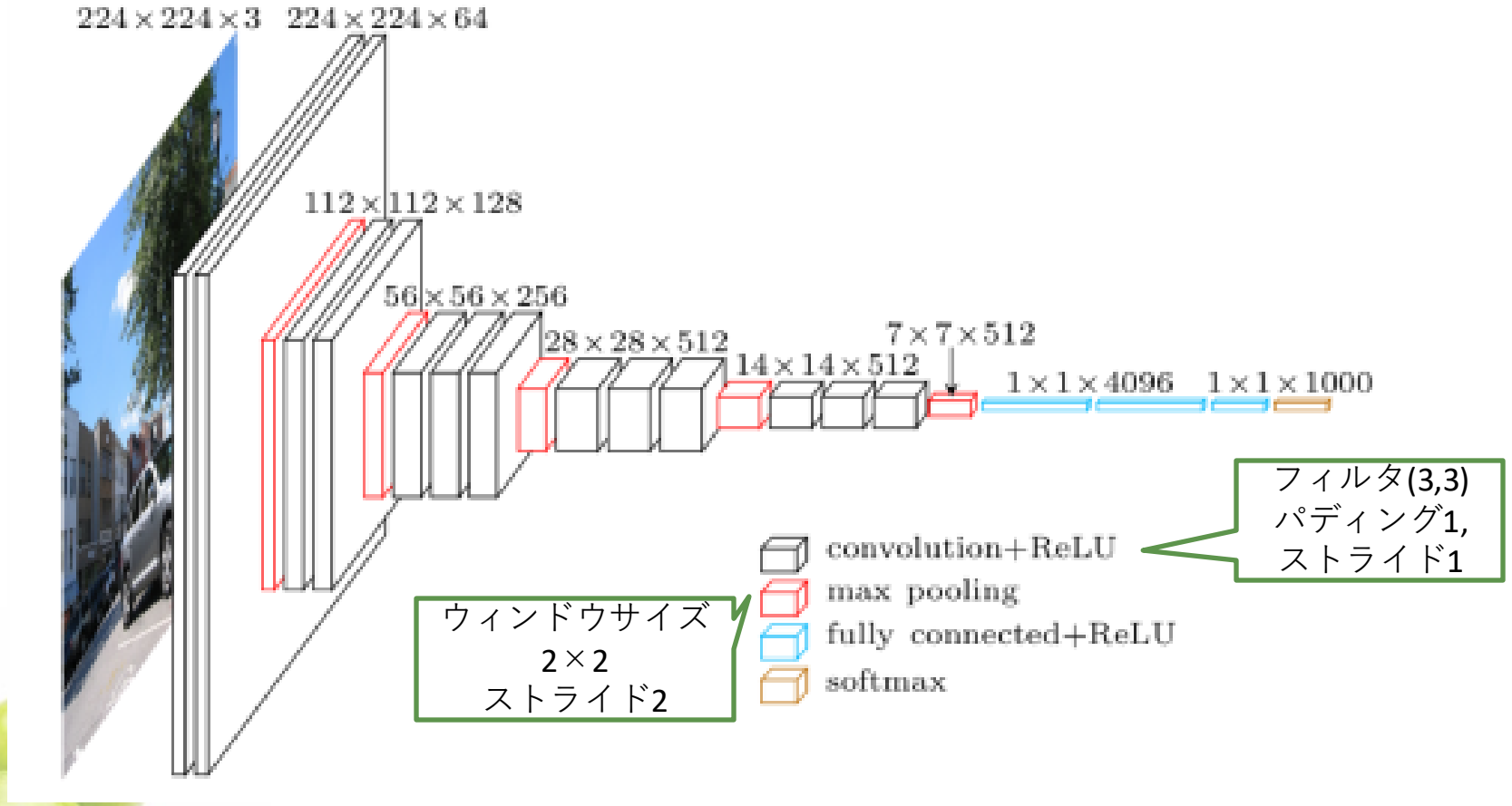
ウィンドウサイズ (2,2)
ストライド 2

2	2
2	1

出力データ

代表的なCNN：VGG16

- 畳み込み層 13層＋FC層 3層



Karen Simonyan and Andrew Zisserman, Very Deep Convolutoinal Networks for Large-Scale Image Recognition, 2015.

CNNのまとめ

- CNNは入力の形状／構造を捉えることに適したNN
- 畳み込み層(Convolution Layer)とプーリング層(Pooling Layer)を持つ
  - 畳み込み演算においてはパディングにより出力データのサイズを調整する
- 畳み込み層、プーリング層のforward、backward演算を実装することで、組み合わせるだけでCNNを実現できる

実習編: PYTORCH

テンソルを作る

- torch.zeros: 0で埋まったテンソル(多次元行列)を作る
- torch.ones: 1で埋まったテンソルを作る
- torch.tensor: リストで表された行列からテンソルを作る
  - requires_grad=Trueをつけることで勾配を計算してくれる

```
[1]  import torch

     a = torch.zeros((10, 15), requires_grad=True)
     print(a)
     b = torch.ones((10, 15), requires_grad=True)
     print(b)
     c = torch.tensor([1.0, 2.0, 3.0], requires_grad=True)
     print(c)
```

実行結果

```
tensor([[0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.],
        [0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]],
       requires_grad=True)
tensor([[1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.],
        [1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.]],
       requires_grad=True)
tensor([1., 2., 3.], requires_grad=True)
```

うまくいけば10x15の0の行列と10x15の1の行列と[1,2,3]の行列が表示される

ネットワークの作成と計算

- 次の関数fの$f(1,2,3)$と$\nabla f(1,2,3)$の計算をしてみよう

$$f(x_1, x_2, x_3) = (x_1 - 2x_2 - 1)^2 + (x_2 x_3 - 1)^2 + 1$$

- 解析的な答え

$f(1,2,3) = 42$

$$\frac{\partial f}{\partial x_1} = 2(x_1 - 2x_2 - 1)$$

$$\frac{\partial f}{\partial x_2} = -4(x_1 - 2x_2 - 1) + 2(x_2 x_3 - 1)x_3$$

$$\frac{\partial f}{\partial x_3} = 2(x_2 x_3 - 1)x_2$$

$$\nabla f(1,2,3) = \left(\frac{\partial f}{\partial x_1}(1,2,3), \frac{\partial f}{\partial x_2}(1,2,3), \frac{\partial f}{\partial x_3}(1,2,3)\right) = (-8, 46, 20)$$

関数と勾配の計算

- 次の関数fの$f(1,2,3)$と$\nabla f(1,2,3)$の計算をしてみよう

$$f(x_1, x_2, x_3) = (x_1 - 2x_2 - 1)^2 + (x_2 x_3 - 1)^2 + 1$$

```
x1 = torch.tensor(1.0, requires_grad=True)
x2 = torch.tensor(2.0, requires_grad=True)
x3 = torch.tensor(3.0, requires_grad=True)

z = (x1 - 2*x2 - 1)**2 + (x2*x3 - 1)**2 + 1

print("z: ", z)
print("z: ", z.item())
z.backward()

print("x1: ", x1.grad)
print("x2: ", x2.grad)
print("x3: ", x3.grad)
```

$(x_1, x_2, x_3) = (1, 2, 3)$

$f(1,2,3)$の計算
計算結果をzに代入

zでテンソル型の結果が表示される。
z.item()とすることでzの結果が実数型で得られる

z.backward()でzに対する勾配の計算

x.gradで、zをxで偏微分した結果が得られる

関数と勾配の計算

- **計算結果**

```
z:  tensor(42., grad_fn=<AddBackward0>)
z:  42.0
x1:  tensor(-8.)
x2:  tensor(46.)
x3:  tensor(20.)
```

解析的な計算結果と一致している

- **解析的な計算結果(2ページ前のスライド)**

$$f(1,2,3) = 42$$

$$\nabla f(1,2,3) = \left(\frac{\partial f}{\partial x_1}(1,2,3), \frac{\partial f}{\partial x_2}(1,2,3), \frac{\partial f}{\partial x_3}(1,2,3)\right) = (-8, 46, 20)$$

PyTorchにはこのように関数と勾配の計算を自動的に行う機能がある

線形回帰モデルの学習

- **教師データ$D = \{(1,2),(2,3),(3,5)\}$から最小二乗法により学習される直線の式$f(x) = ax + b$を求めよ**

$$L(a,b) = \sum_{(x,y)\in D} \big(y - f(x)\big)^2$$

$$L(a,b) = (2 - a - b)^2 + (3 - 2a - b)^2 + (5 - 3a - b)^2$$

$$\frac{\partial L(a,b)}{\partial a} = -2(2 - a - b) - 4(3 - 2a - b) - 6(5 - 3a - b) = -46 + 28a + 12b = 0$$

$$\frac{\partial L(a,b)}{\partial b} = -2(2 - a - b) - 2(3 - 2a - b) - 2(5 - 3a - b) = -20 + 12a + 6b = 0$$

解くと、$a = \frac{3}{2}, b = \frac{1}{3}$。従って、$f(x) = \frac{3}{2}x + \frac{1}{3}$。

PyTorchを使った勾配降下法による最小二乗法

```
import torch

lr = 0.01
X = [1,2,3]
Y = [2,3,5]
a = 0.0
b = 0.0

for e in range(500):
  __a = torch.tensor(a, requires_grad=True)
  __b = torch.tensor(b, requires_grad=True)
  loss = torch.tensor(0.0)
  for i in range(len(X)):
    loss = loss + (Y[i] - __a*X[i]-__b)**2
  loss.backward()
  a = a - lr * __a.grad.item()
  b = b - lr * __b.grad.item()
  print("epoch: ", e, "a: ", a, "b: ", b, "loss: ", loss.item())
```

パラメータ__a, __bの設定
損失lossの初期化

各データの損失の累積を計算

損失に対する勾配を自動計算

勾配を使ってパラメータ更新

回帰モデルをPyTorchで計算してみよう

$D = \{(1,2),(2,3),(3,5)\}$

$f(x) = ax + b$

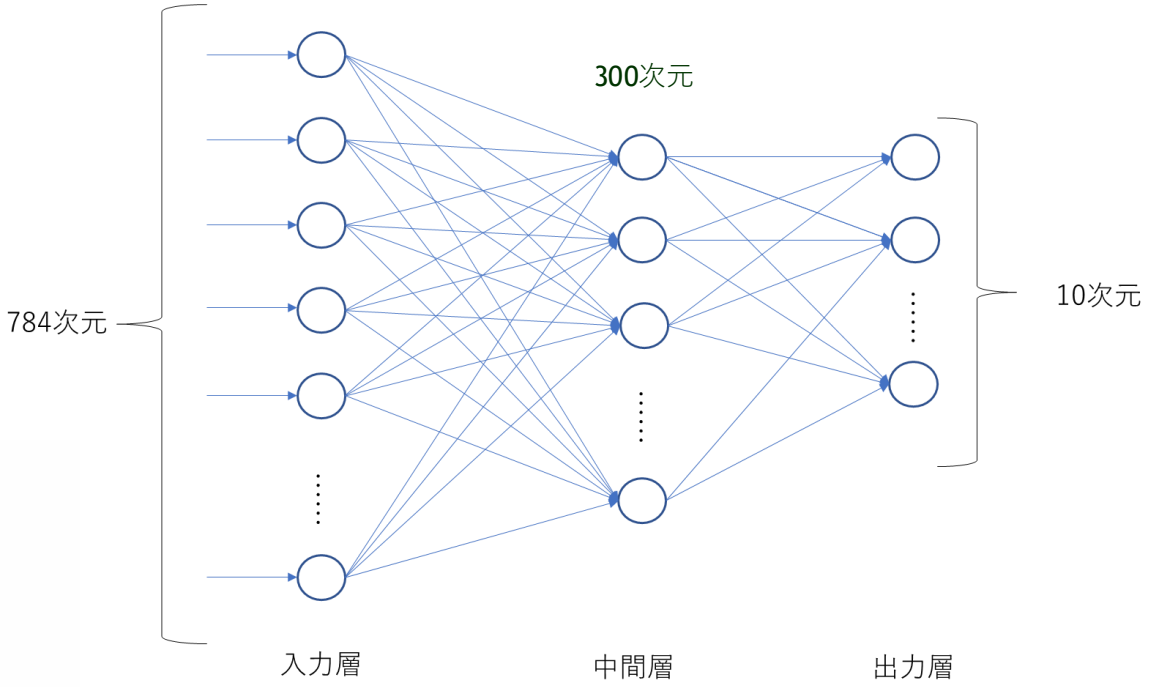
実行結果 (解析解 $a = \frac{3}{2}, b = \frac{1}{3}$)

```
epoch:  496 a:  1.4967026257514946 b:  0.34082910180091835 loss:  0.16669124364852905
epoch:  497 a:  1.4967263859510413 b:  0.34077503502368905 loss:  0.16669084131717682
epoch:  498 a:  1.4967499971389762 b:  0.34072136759757976 loss:  0.16669055819511414
epoch:  499 a:  1.4967734265327444 b:  0.34066808342933635 loss:  0.16669011116027832
```

PYTORCHによる深層学習 (1)

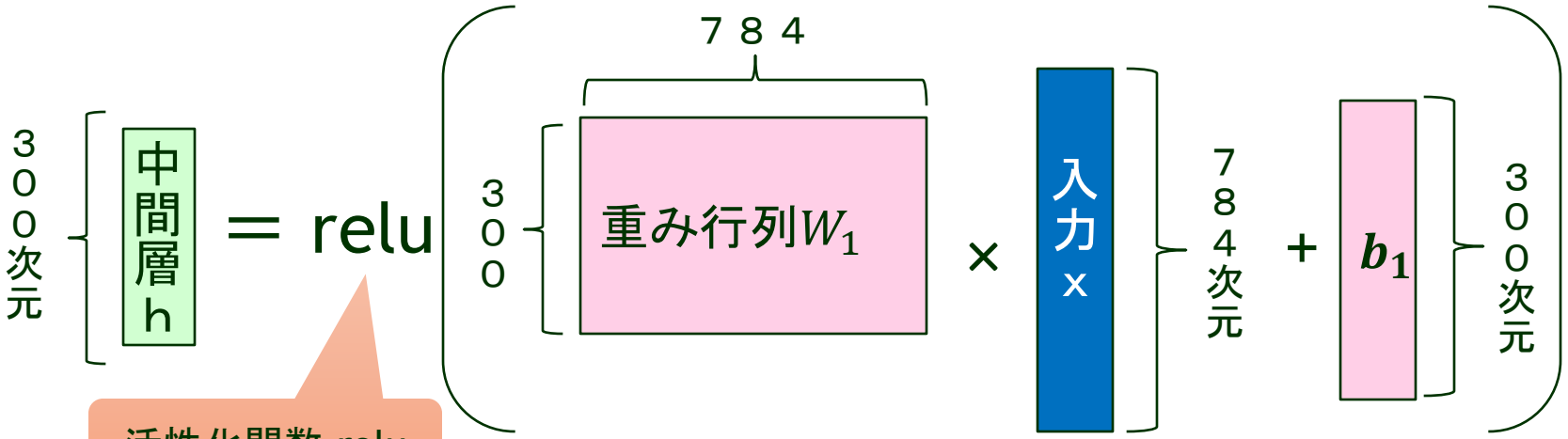
多層パーセプトロン

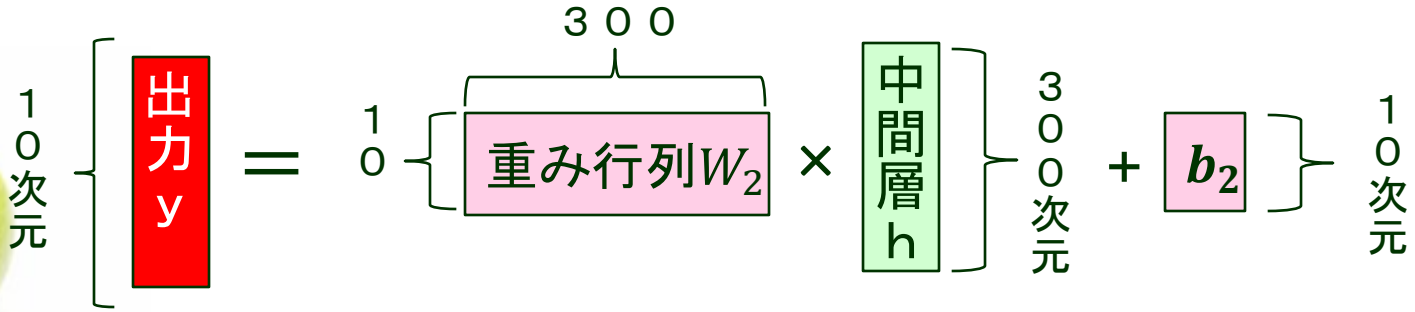
- 入力層(784次元)、中間層(300次元)、出力層(10次元)の3層のニューラルネットワークを作って白黒画像(28x28)の画像認識を行います



多層パーセプトロン

- 入力層(784次元)、中間層(300次元)、出力層(10次元)の3層のニューラルネットワークの計算





線形変換(全結合層) torch.nn.Linear

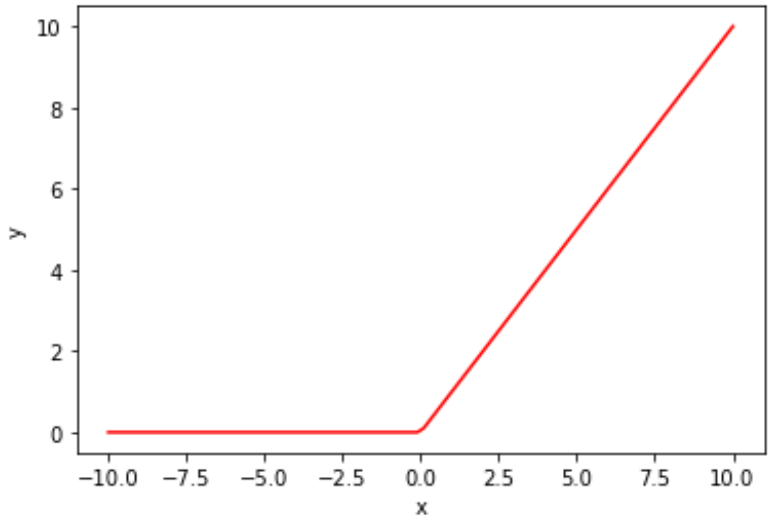
- torch.nn.Linear
  - 線形変換(全結合層)のモジュール
    - $\boldsymbol{y} = \mathrm{W}\boldsymbol{x} + \boldsymbol{b}$を計算する
    - W(行列)と$\boldsymbol{b}$(ベクトル)がパラメータ
  - `torch.nn.Linear(in_dim, out_dim)`
    - `in_dim`次元のベクトルから`out_dim`次元のベクトルに射影する線形変換を計算する層を作る
  - 例: `torch.nn.Linear(3, 4)`
    - 3次元ベクトルから4次元ベクトルに射影する線形変換

活性化関数 ReLU (Rectified Linear Unit) 関数

- **ReLU関数**
  - 出力の活性化(オン・オフ)を表す非線形関数

$$h(x) = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

  - `torch.nn.functional.relu(x)`
    - xに対して、 $h(x)$の計算をする



データ MNIST

- MNIST
  - 手書き文字の認識を行うタスク
  - 画像に対して、どの数字が書かれているかを予測する。

  - 訓練データは60,000データ
  - テストデータは10,000データ
  - 各データは画像とラベルのペア
  - 各画像は28×28ピクセル
  - ラベルは0～9までのいずれかの数字。画像に対する正解の数字。

データの読み込み

- データ読み込みのためのプログラム

```
import numpy as np
import torch
import torch.nn.functional as F
import torchvision as tv

train_dataset = tv.datasets.MNIST(root="./", train=True,
                                  transform=tv.transforms.ToTensor(),
                                  download=True)
test_dataset = tv.datasets.MNIST(root="./", train=False,
                                 transform=tv.transforms.ToTensor(),
                                 download=True)
train_loader = torch.utils.data.DataLoader(dataset=train_dataset,
                                           batch_size=100,
                                           shuffle=True)
test_loader = torch.utils.data.DataLoader(dataset=test_dataset,
                                          batch_size=100,
                                          shuffle=False)
```

ライブラリの読み込み

訓練データとテストデータの読み込み(初めて実行するときはデータをネットからダウンロードする)

訓練データとテストデータのミニバッチ処理
・ミニバッチサイズ=100
・データの順番をシャッフル

DataLoaderがしていること

- **ミニバッチ形式への変換**
  - もともとのデータ（1×28×28のリストと数値のペアのリスト）

```
[([[[0,0,…画像データ…,0,0]]], 5),
 ([[[0,0,…画像データ…,0,0]]], 2),
 …
 ([[[0,0,…画像データ…,0,0]]], 8)]
```

60000個(=データサイズ)

  - 変換後のデータ(1×28×28のリストと数値のリストのペア

```
([[[[0,0,…画像データ…,0,0]]],
  [[[0,0,…画像データ…,0,0]]],
  …
  [[[0,0,…画像データ…,0,0]]]],
 [5, 2, …, 8])
```

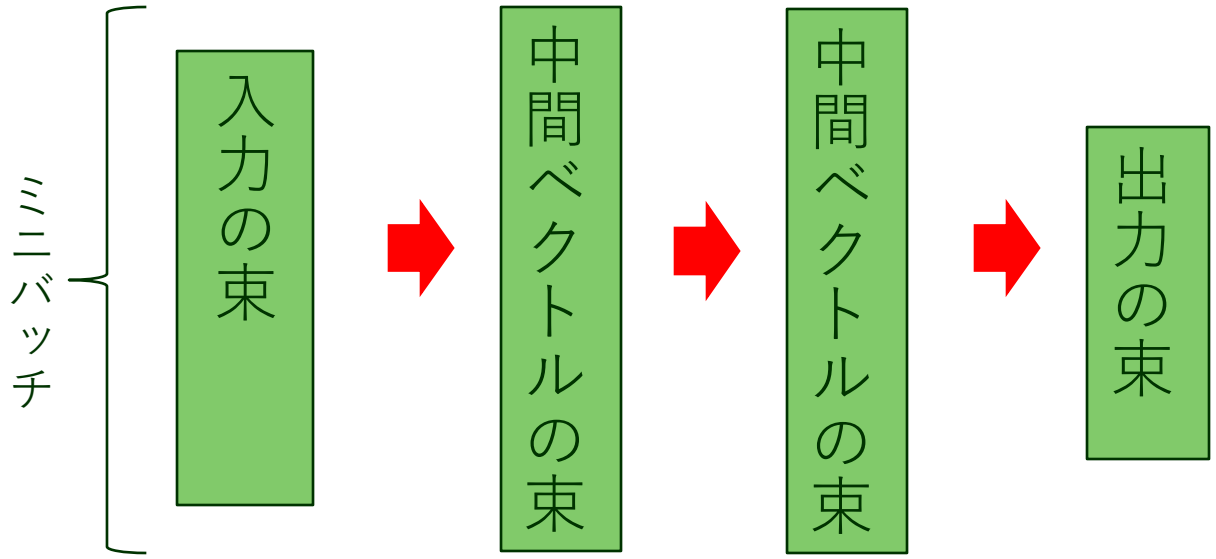
入力となる画像だけを集めたもの
(100個=ミニバッチサイズ)

出力となる数値だけを集めたもの
(100個=ミニバッチサイズ)

100個のミニバッチデータが600個続く

ミニバッチ

- ミニバッチ内のデータをまとめて左から右に流す(計算する)



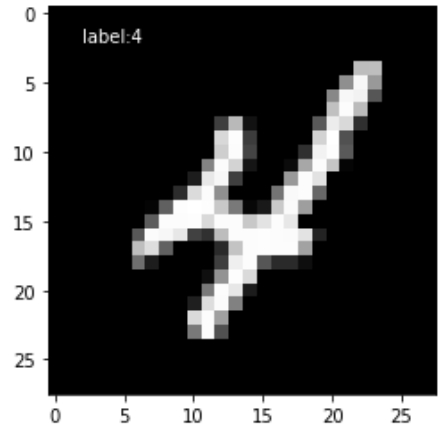
- まとめて計算できる行列演算が増える
  - 行列演算が得意なGPUにとって非常に効率の良い計算方法

(参考)どんな画像が入っているのか みてみよう

- 先程のプログラムの下に次のプログラムを追加して実行してみよう(ただし、コマンドプロンプトでpip3 install matplotlibの実行が必要)

```
import matplotlib.pyplot as plt
for i in range(10):
  print(train_dataset[i])
  plt.imshow(train_dataset[i][0][0], cmap="gray")
  txt = "label:"+str(train_dataset[i][1])
  plt.text(2, 2, txt, color="white")
  plt.show()
```

train_dataset[i][0][0]でi番目の画像(28x28の白黒)
train_dataset[i][1]でi番目の正解ラベル



・具体的なデータの中身が表示される
・このような画像が10枚表示される(左上の「label:4」はこの画像の正解ラベル)

(参考) カラー画像だったら、
train_dataset[i][0][0] 赤(R)の画像
train_dataset[i][0][1] 緑(G)の画像
train_dataset[i][0][2] 青(B)の画像

今回は白黒なので、train_dataset[i][0][0]だけ使う

コメント化

- **せっかく書いたプログラムを消すのはもったいないので、コメント化しておこう**
  - 「#」をいれるとそこから行末までは無視される(コメント化)

```
for i in range(10):
  print(train_dataset[i])
  plt.imshow(train_dataset[i][0][0], cmap="gray")
  txt = "label:"+str(train_dataset[i][1])
  plt.text(2, 2, txt, color="white")
  plt.show()
```

```
# for i in range(10):
#   print(train_dataset[i])
#   plt.imshow(train_dataset[i][0][0], cmap="gray")
#   txt = "label:"+str(train_dataset[i][1])
#   plt.text(2, 2, txt, color="white")
#   plt.show()
```

コメント化された部分は実行されない

プログラムをみやすくするために積極的にコメントをいれよう
(例) #初期設定
　　#モデルの定義
　　#ここから訓練

ネットワークの作成

- ネットワークの設定とパラメータ最適化の設定

```
l1 = torch.nn.Linear(784, 300)
l2 = torch.nn.Linear(300, 10)
params = list(l1.parameters())+ list(l2.parameters())
optimizer = torch.optim.Adam(params)

def mynet(x):
  h = F.relu(l1(x))
  y = l2(h)
  return y
```

l1とl2は線形変換の関数

l1/l2.parameters()でl1とl2に含まれるパラメータを取り出す

Adamというパラメータ最適化手法を使う。パラメータ更新に関わるパラメータ群paramsを渡す

ネットワーク全体の計算を行う関数 mynet
・xが入力、yが出力
・hはニューラルネットワーク前半(784次元→300次元への線形変換l1と活性化関数ReLUの適用)
・yはニューラルネットワーク後半(hに対して300次元→10次元への線形変換l2を適用)

ネットワークの学習

- **訓練データからの学習**

```
#学習
for e in range(10):
  loss = 0
  for images, labels in train_loader:
    images = images.view(-1, 28*28)
    optimizer.zero_grad()
    y = mynet(images)
    batchloss = F.cross_entropy(y, labels)
    batchloss.backward()
    optimizer.step()
    loss = loss + batchloss.item()
  print("epoch:", e, "loss:", loss)
```

データ全体を10回学習する

train_loaderから100個ずつデータを受け取る
Imagesは画像データ100枚
labelsは正解ラベル100個

(100×1x28×28)から(100×784)に変形

勾配を初期化

ネットワークの計算(ラベルの予測)

正解ラベルに対する出力の損失

誤差に対する勾配を計算

パラメータ更新

ミニバッチでの損失をlossに足す

ネットワークによる推論

- テストデータに対して学習したネットワークを評価する

```
#テスト
correct = 0
total = len(test_loader.dataset)
for images, labels in test_loader:
  images = images.view(-1, 28*28)
  y = mynet(images)
  pred_labels = y.max(dim=1)[1]
  correct = correct + (pred_labels == labels).sum()

print("correct:", correct.item())
print("total:", total)
print("accuracy:", correct.item()/total)
```

test_loaderから100個ずつデータを受け取る
Imagesは画像データ100枚
labelsは正解ラベル100個

(100×1x28×28)から(100×784)に変形

ネットワークの計算

ラベルの予測(最大値となるラベル)

100個のうち正解数がいくつか数える

correctは正解数
totalはテストデータのデータ数
accuracyは精度

実行結果

```
epoch: 0 loss: 195.63006672263145
epoch: 1 loss: 82.39563230797648
epoch: 2 loss: 54.87096862308681
epoch: 3 loss: 40.70517487358302
epoch: 4 loss: 31.048207819461823
epoch: 5 loss: 23.900350784882903
epoch: 6 loss: 19.02593113365583
epoch: 7 loss: 14.900030710035935
epoch: 8 loss: 12.13549662521109
epoch: 9 loss: 9.818495093728416
```

損失がだんだん減っていることがわかる

```
correct: 9798
total: 10000
accuracy: 0.9798
```

10000データ中、9798データの正解
(=97.98%の正解率)

- http://yann.lecun.com/exdb/mnist/ の結果と比較してみよう

PYTORCHのクラスライブラリ

テンソル型の定義

- 計算グラフのノード(変数)を表す
- Pytorchではtorch.float32(実数), torch.float64(実数)の型を用いる

```
import torch

x = torch.zeros((10, 20), dtype=torch.float32, requires_grad=True)
print(x)
x = torch.ones((10, 20), dtype=torch.float32, requires_grad=True)
print(x)
x = torch.tensor([1,2,3], dtype=torch.float32, requires_grad=True)
print(x)
```

torch.nn.functional

- **パラメータ(求めたい重み変数)を含まない関数はtorch.nn.functionalパッケージに用意されています**
  - 活性化関数
  - 損失関数など

torch.nn.functional

- 簡単な例

```
x1 = torch.tensor(1.5)
x2 = torch.tensor(0.75)
y = torch.nn.functional.relu(x1)
z = torch.nn.functional.mse_loss(x1, x2)
print("y: "+str(y))
print("z: "+str(z))
```

y: 1.5000
z: 0.5625

torch.nn

- パラメータ(求めたい重み変数)を含む関数はtorch.nnパッケージに用意されています。
  - 多くのNNモデルはtorch.nnの直下に用意されていて、活性化関数などの簡単な関数はtorch.nn.functionalの下に用意されている、ということになります。

torch.nn.Paremeter

- **PyTorchでの重み変数(パラメータ)を生成するクラス**
  - テンソル型を生成する際にrequires_grad=Trueとするだけでは、pytorchの中ではパラメータ(保存と更新の対象となる変数)として認識されない
  - nn.Parameter(x)とすることでxを中身とするパラメータを生成する

(ただし、通常使うことはあまり無い。カスタムレイヤーを作るときに必要)

  - 例: 2×3の重み行列

```
W = torch.nn.Parameter(torch.tensor([[1.,2.,3.],
                                     [4.,5.,6.]]))
```

↓

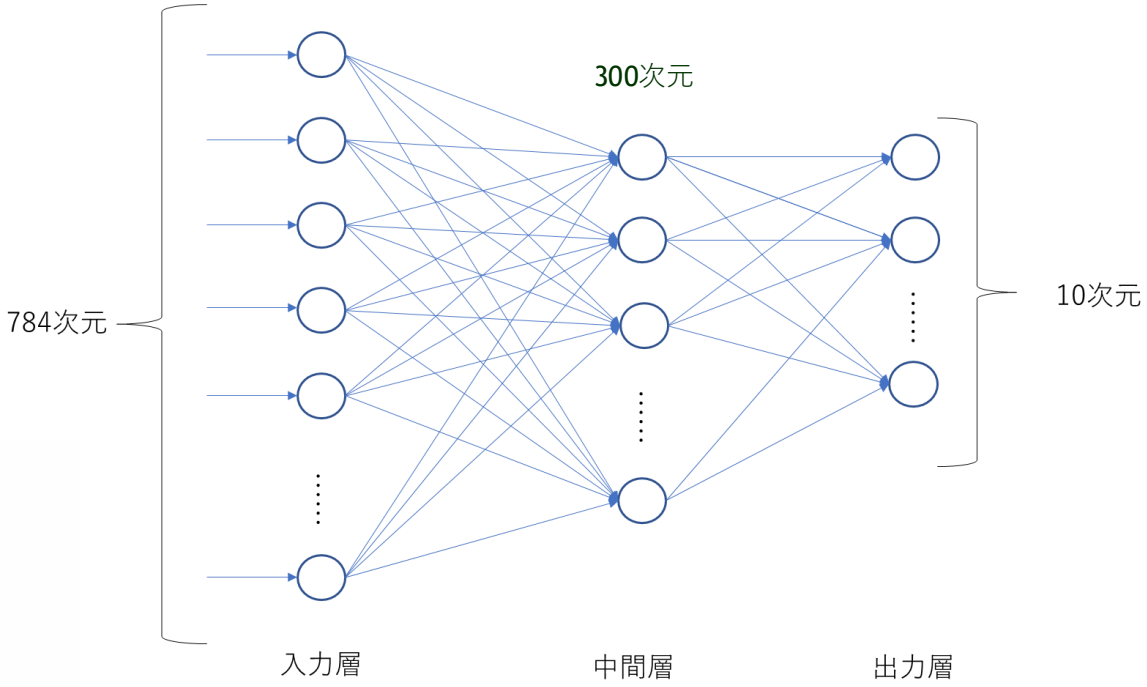
```
Parameter containing:
tensor([[1., 2., 3.],
        [4., 5., 6.]], requires_grad=True)
```

torch.nn.Module

- **Moduleクラス**
  - ネットワーク全体の構成(モデル)を与えるクラス
    - torch.nnやtorch.nn.functionalをパーツとして全体を構成し、そのネットワークが表現する合成関数を計算する
    - 合成関数に含まれるパラメータを学習する

Module

- 入力層(784次元)、中間層(300次元)、出力層(10次元)の3層のニューラルネットワークを作って白黒画像(28x28)の画像認識を行います



Module

- 例: 入力層(784次元)、中間層(300次元)、出力層(10次元)の3層のニューラルネットワーク (活性化関数はReLU)

```
class MyNet(torch.nn.Module):
    def __init__(self):
        super(MyNet,self).__init__()
        self.l1 = torch.nn.Linear(784,300)
        self.l2 = torch.nn.Linear(300,10)
    def forward(self,x):
        h = torch.nn.functional.relu(self.l1(x))
        y = torch.nn.functional.relu(self.l2(h))
        return y
```

__init__()関数でネットワークの要素を定義

forward計算のときにforward()が呼ばれる。同時にネットワークを構築

optim

- **optimクラス**
  - 最適化ツール
  - 様々な最適化アルゴリズムを選択して利用できる
  - 最初に次の定義をして最適化クラスを選択、初期化

```
model = MyNet()
optimizer = torch.optim.SGD(model.parameters())
```

optimizerを選択し
パラメータをセット

  - パラメータ更新時

```
optimizer.zero_grad()
loss = model(x)
loss.backward()
optimizer.step()
```

勾配の初期化

パラメータを更新

PYTORCHクラスを用いたMNIST

データの読み込み

- データ読み込みのためのプログラム(以前のMNIST用プログラムと同じ)

```
import numpy as np
import torch
import torch.nn.functional as F
import torchvision as tv

train_dataset = tv.datasets.MNIST(root="./", train=True,
                                  transform=tv.transforms.ToTensor(),
                                  download=True)
test_dataset = tv.datasets.MNIST(root="./", train=False,
                                 transform=tv.transforms.ToTensor(),
                                 download=True)
train_loader = torch.utils.data.DataLoader(dataset=train_dataset,
                                           batch_size=100,
                                           shuffle=True)
test_loader = torch.utils.data.DataLoader(dataset=test_dataset,
                                          batch_size=100,
                                          shuffle=False)
```

ライブラリの読み込み

訓練データとテストデータの読み込み(初めて実行するときはデータをネットからダウンロードする)

訓練データとテストデータのミニバッチ処理
・ミニバッチサイズ=100
・データの順番をシャッフル

グローバル変数の定義

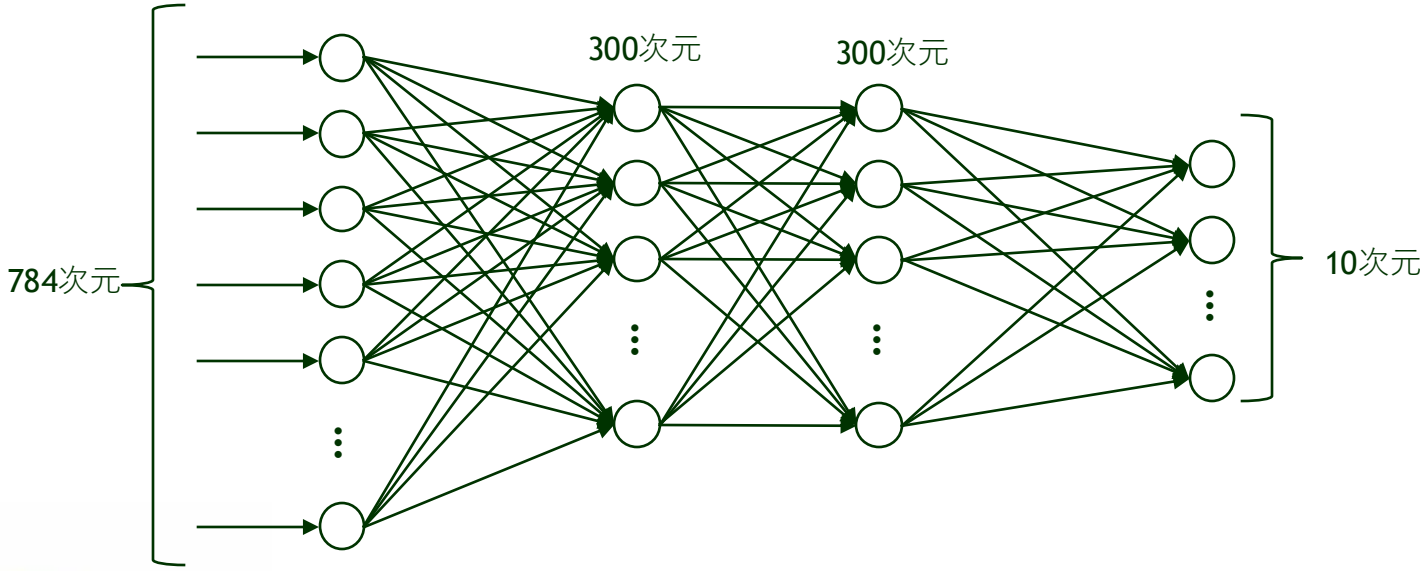
- グローバル変数を定義しておきます

```
MODELNAME = "mnist.model"
EPOCH = 10
DEVICE = "cuda" if torch.cuda.is_available() else "cpu"
print(DEVICE)
```

GPUが利用可能ならDEVICE=”cuda”
CPUを利用するのならDEVICE=”cpu”

モデルの定義

- **手書き文字の認識を行うモデル**
  - 入力層(784次元)、中間層1(300次元)、中間層2(300次元)、出力層(10次元)の中間層2層のネットワーク



  - 全結合層のみによるモデル
  - torch.nn.Conv2dを使用し、入力データの形式を少し変更すれば、畳込みニューラルネットワーク(CNN)も可能

モデルの定義

- **手書き文字の認識を行うモデルを構築**
  - 入力層(784次元)、中間層1(300次元)、中間層2(300次元)、出力層(10次元)の中間層2層のネットワーク

```
class MNIST(torch.nn.Module):
  def __init__(self):
    super(MNIST,self).__init__()
    self.l1 = torch.nn.Linear(784, 300)
    self.l2 = torch.nn.Linear(300, 300)
    self.l3 = torch.nn.Linear(300, 10)
  def forward(self,x):
    h = F.relu(self.l1(x))
    h = F.relu(self.l2(h))
    y = self.l3(h)
    return y
```

__init__()関数でネットワークの要素を定義

forward計算のときにforward()が呼ばれる。同時にネットワークを構築

モデルの定義　__init__メソッド

- **__init__メソッド**
  - 入力層の次元数: 768 (= ピクセル数)
  - 第1中間層の次元数: 300
  - 第2中間層の次元数: 300
  - 出力次元数: 10 (= 多クラス分類のクラス数)
  - ニューラルネットワークを定義する手順
    - 親クラスのコンストラクタを呼ぶ
    - __init__(): 以下に使用するネットワークを定義
  - torch.nn.Linear: 全結合層
    - torch.nn.Linearの第一引数: 全結合層の入力次元数
    - torch.nn.Linearの第二引数: 全結合層の出力次元数
    - 全結合層以外にも多くのネットワークが使用できます (https://pytorch.org/docs/stable/nn.html 参照)。

モデルの定義 forwardメソッド

- forwardメソッド
  - ネットワークの結合を定義
  - 784次元のベクトル(28×28ピクセルの画像)を受け取る
  - 10次元のベクトルを返す
  - l1が784次元(入力層)を中間次元(300次元)に圧縮
  - l2はl1の出力を受け取り、中間ベクトル(300次元)を生成
  - l3は中間ベクトルを受け取り、10次元のベクトル(出力層)に変換
    - MNISTの分類が10クラス(0～9の文字)であるため、10次元のベクトルを出力
  - L1, l2層に対しては、活性化関数のrelu関数

モデルの訓練

```
def train():
  model = MNIST().to(DEVICE) #モデルを定義してDEVICEにのせる
  optimizer = torch.optim.Adam(model.parameters()) #最適化にAdamを使う
  for epoch in range(EPOCH): #EPOCH回最適化を行う
    loss = 0
    for images, labels in train_loader:
      images = images.view(-1, 28*28).to(DEVICE)
      labels = labels.to(DEVICE)
      optimizer.zero_grad() #勾配の初期化
      y = model(images) #forward計算
      batchloss = F.cross_entropy(y, labels) #損失の計算
      batchloss.backward() #逆伝搬の計算
      optimizer.step() #重み変数の更新
      loss = loss + batchloss.item()

    print("epoch", epoch, ": loss", loss)
  torch.save(model.state_dict(), MODELNAME)
```

データをミニバッチサイズに切り出す

(100×1×28×28)から(100×784)に変形

ミニバッチ内の処理

モデルをファイルに保存

テスト

```
def test():
  total = len(test_loader.dataset)
  correct = 0
  model = MNIST().to(DEVICE)
  model.load_state_dict(torch.load(MODELNAME))
  model.eval()
  for images, labels in test_loader:
    images = images.view(-1, 28*28).to(DEVICE)
    labels = labels.to(DEVICE)
    y = model(images) #forward計算
    pred_labels = y.max(dim=1)[1]
    correct = correct + (pred_labels == labels).sum()
  print("correct:", correct.item())
  print("total:", total)
  print("accuracy:", (correct.item() / float(total)))
```

ファイルに保存したモデルをロード

テストデータに対してループ

最大値の次元IDを得る

正解率を計算

訓練・テスト

```
train()
#test()
```

訓練・テストのうちやりたくないこと
に対してコメントアウトする

Pythonでは#がコメントを表す記号

- train()を実行して訓練してみよう
- test()を実行して精度を測ってみよう
- 各データに対し、正解ラベルと予測ラベルをそれぞれ表示させてみよう
- http://yann.lecun.com/exdb/mnist/ の結果と比較してみよう

Jetson NanoでのMNISTの訓練・テスト

- **Jetson Nanoにファイルを転送する**
  - 全結合のニューラルネットワークのプログラム(mnist.py)

```
VirtualBoxのterminal
$ scp mnist.py (ユーザー名)@(Jetson NanoのIP):
```

- **まず、trainが動くか確認する**
  - EPOCH=10で8分ぐらいかかる
  - 学習が終わったらtestの動作確認をする

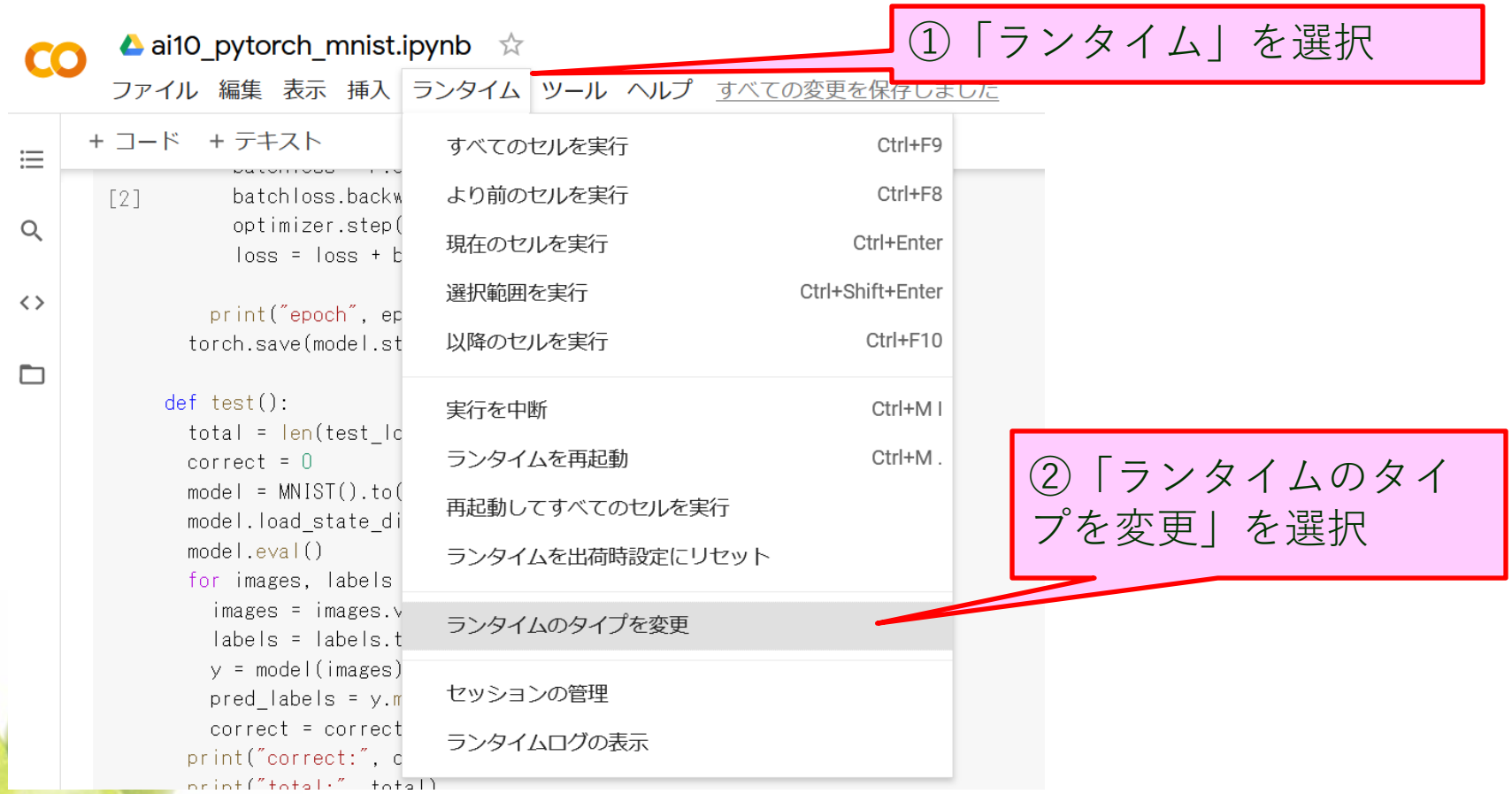
Jetson Nanoのterminal上でifconfigと入力するとIPがわかります

- **GPUの使用状況を確認するコマンド(jetson-stats)**

```
Jetson Nanoのterminal
$ sudo jtop
```

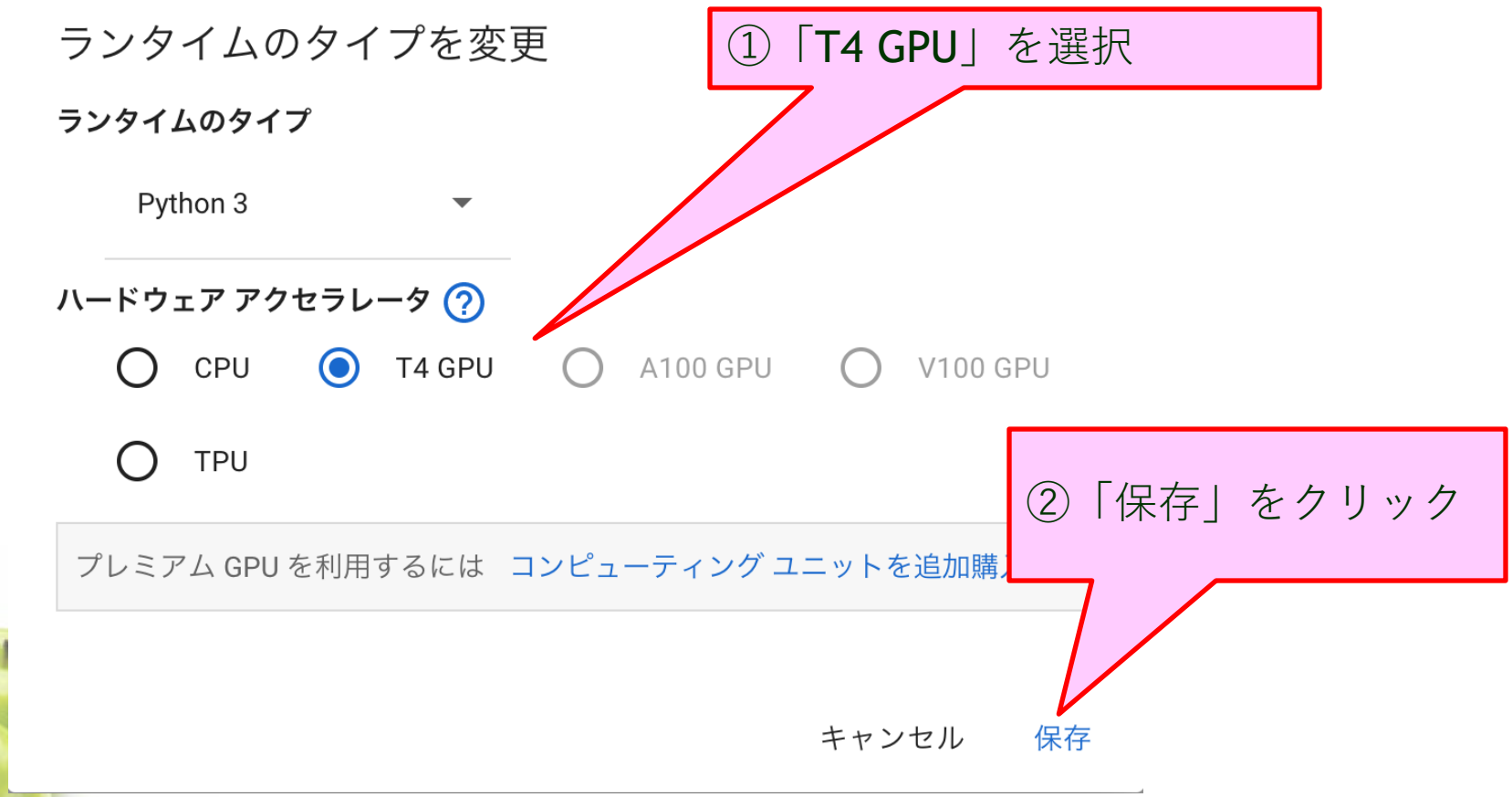
GPUを使ってみよう

- Google ColaboratoryではGPUを利用することができます



GPUを使ってみよう

- GPUの設定ができたらMNISTの学習を行ってみよう



まとめ

- Pythonのクラスについて学んだ
- テンソル型について勉強した
- クラスを用いたPyTorchの使い方を学んだ
- GPUを用いた深層学習を行った

オプション課題

畳み込みニューラルネットワークに変えてみよう

- 中間層までの全結合を畳み込みニューラルネットワークに変えてみよう
- `__init__`の中で、`nn.Conv2d`を使う
- データローダから受け取った入力画像のサイズ調整(view)
  - (ミニバッチサイズ, 1チャネル, 横幅, 縦幅)の4次元配列

畳み込みニューラルネットワークに変えてみよう

```
def __init__(self):
     super(MNIST,self).__init__()
     filtersize = 3
     in_channel = 1
     out_channel = 10
     self.convout = out_channel*(28-filtersize+1)*(28-filtersize+1)
     self.l1 = torch.nn.Conv2d(in_channel, out_channel, filtersize)
     self.l2 = torch.nn.Linear(self.convout, 300)
     self.l3 = torch.nn.Linear(300, 10)
def forward(self, x):
     h = F.relu(self.l1(x)).view(-1, self.convout)
     h = F.relu(self.l2(h))
     y = self.l3(h)
     return y
```

畳み込みニューラルネットワークに変えてみよう

- train, testの中で入力変数のサイズ調整(view不要)
  - (ミニバッチサイズ, 1チャネル, 横幅, 縦幅)の4次元配列
  - デフォルトの画像サイズのままで大丈夫

```
def train():
  model = MNIST().to(DEVICE)
  optimizer = torch.optim.Adam(model.parameters())
  for epoch in range(EPOCH):
        loss = 0
        for images, labels in train_loader:
              images = images.to(DEVICE)
              label = labels.to(DEVICE)
               . . .
```

```
def test():
…
      images = images.to(DEVICE)
…
```

CIFAR-10の画像分類

- CIFAR-10というデータセットに対して、学習と推論を行い、その精度評価の実験を行いましょう。
- CIFAR-10は10クラスの画像分類データであり、

```
train_dataset = tv.datasets.CIFAR10(root="./", train=True,
                                    transform=tv.transforms.ToTensor(),
                                    download=True)
test_dataset = tv.datasets.CIFAR10(root="./", train=False,
                                   transform=tv.transforms.ToTensor(),
                                   download=True)
```

を実行することによりMNIST同様、データが得られます。各画像は3色×32pixel×32pixelとなっています。

より高い精度の実現を目指して、モデル設計を行いましょう。