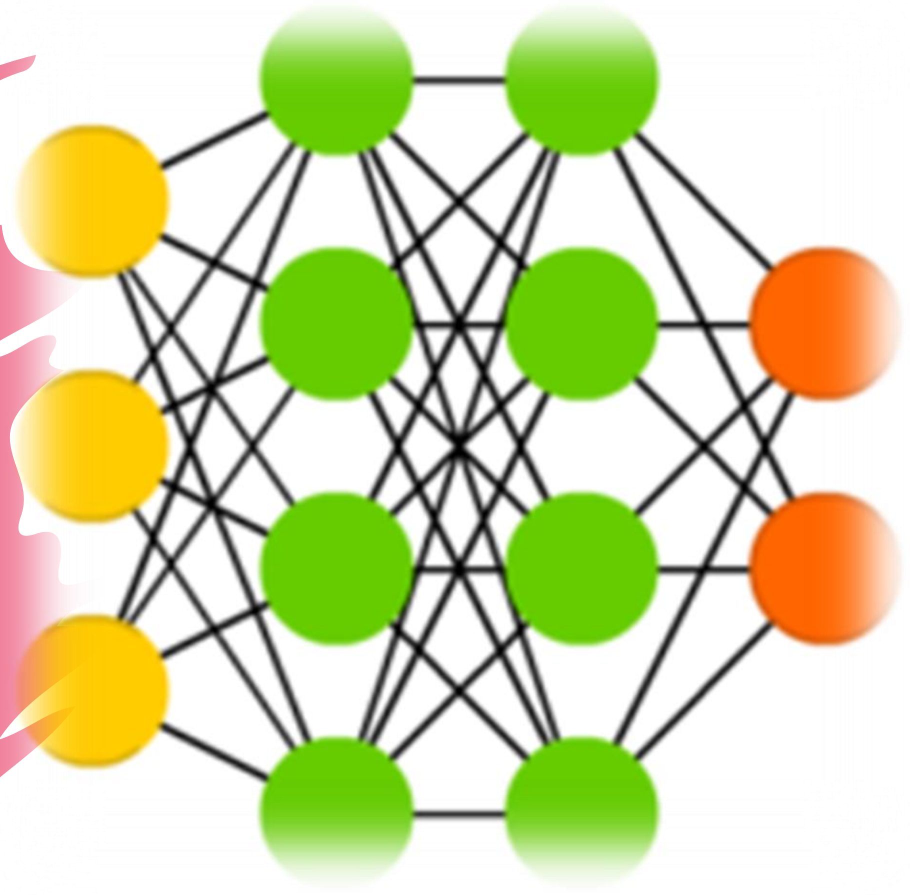
手計算
ニューラルネットワーク
入門　②逆伝播編



予定

第１回　理論：順伝播

第２回　理論：逆伝播

第３回　実装

第４回　実験

第５回

ニューラルネットワーク
を完全に理解したい

pythonのライブラリ使うだけ

なんとなく理論は知ってるけど
結局呪文を呪文になってしまってる

引数やパラメータの説明を聞いても
多分理解できない（自分が）

```
from tensorflow.python.keras.models import Sequential
from tensorflow.python.keras.layers import Dense

model = Sequential()
model.add(Dense(...))
```

手計算ニューラルネットワーク

完全に理解したい

⇔手計算できるレベルで理解する

⇔for文if文で実装ができる

最終目標（最低限）

ニューラルネットワークを完全に理解する

C++でニューラルネットワークを実装し
何かしらの分類問題を解く

C++でニューラルネットワークを実装し
何かしらの回帰問題を解く

最終目標（理想）

Kaggle Titanic in C++

k KAGGLE · GETTING STARTED PREDICTION COMPETITION · ONGOING

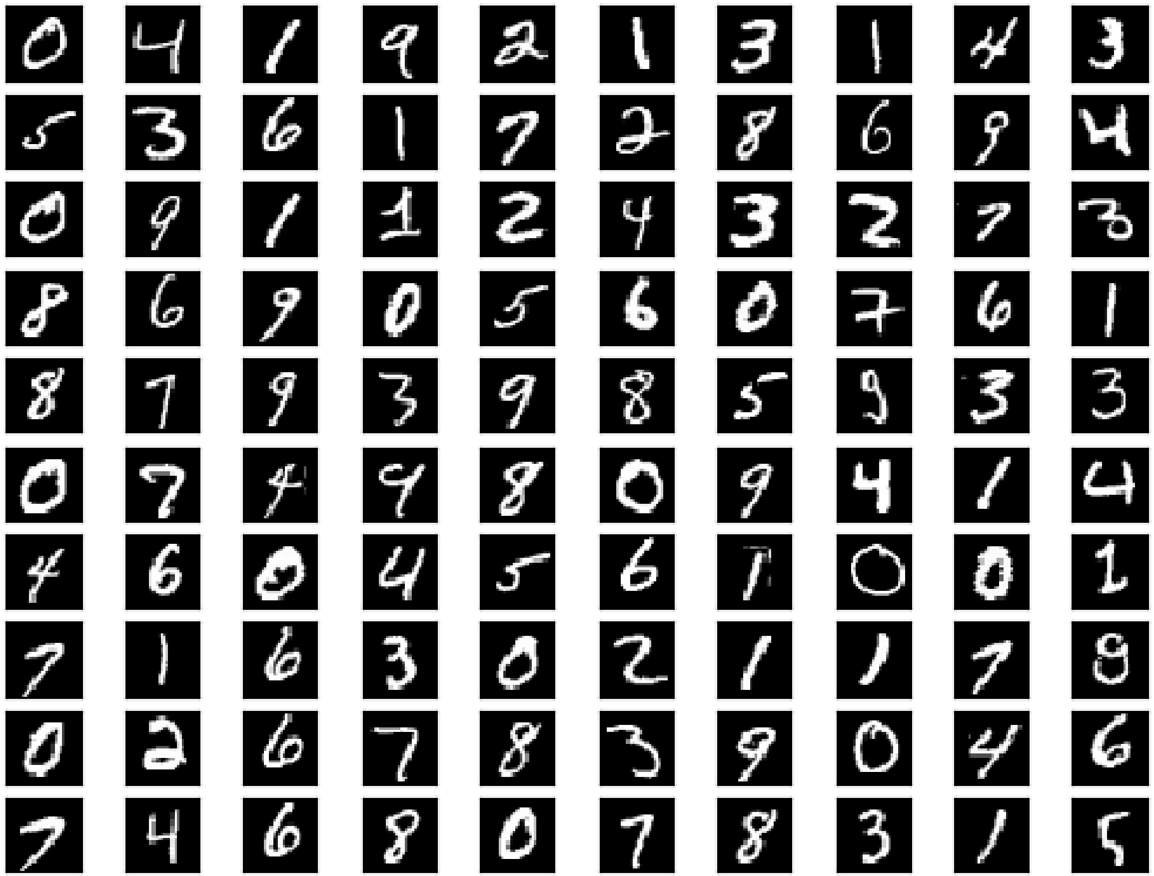
Submit Prediction

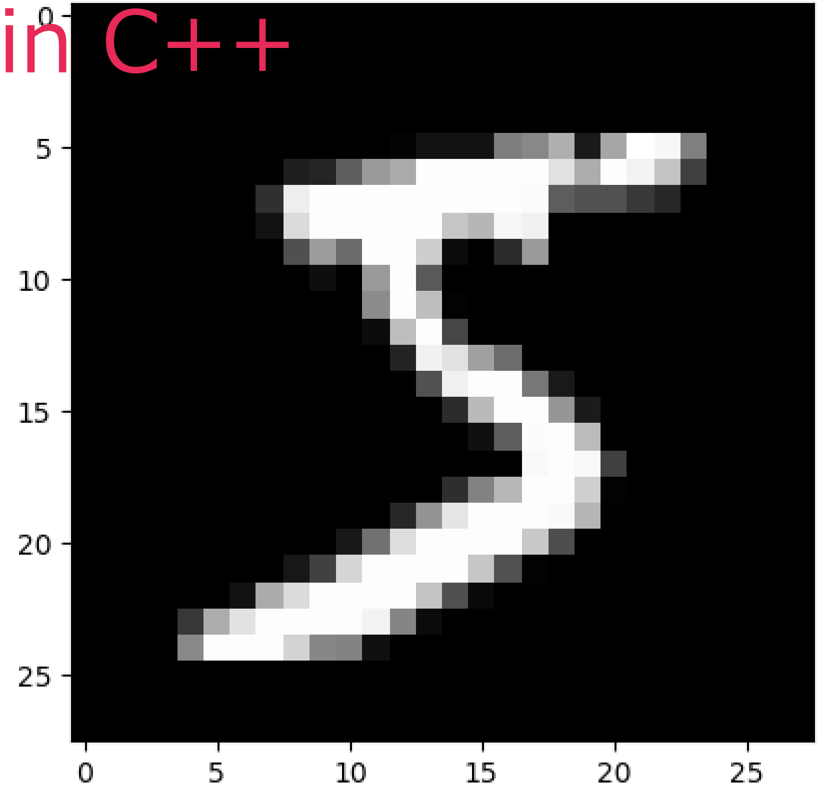
Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

最終目標（理想）

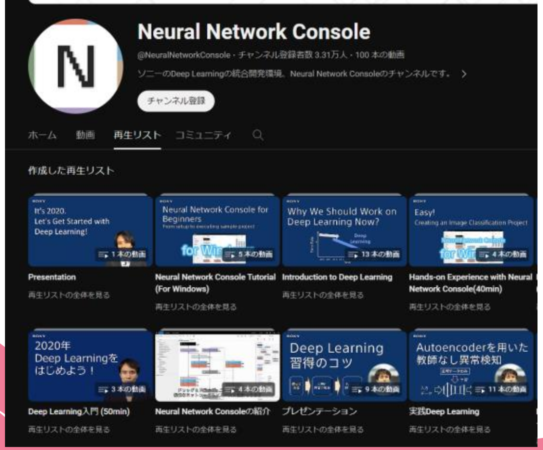
MNISTデータセットの手書き文字認識 in C++





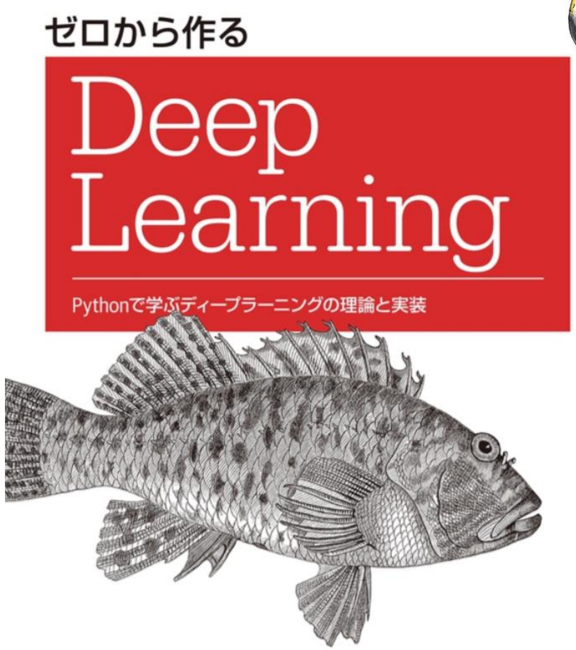
28×28×256

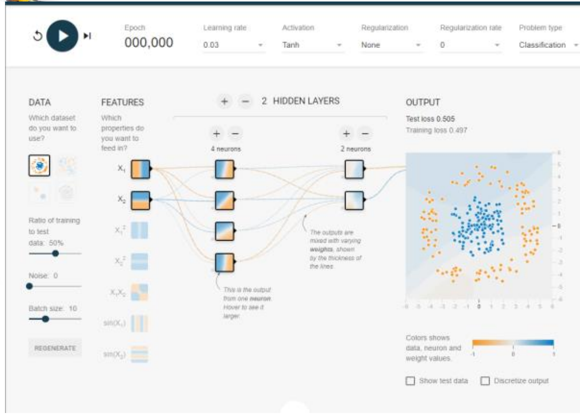
参考文献


Neural Network Console
チャンネル登録
ホーム
動画
再生リスト
コミュニティ
作成した再生リスト
Presentation
Neural Network Console Tutorial (For Windows)
Introduction to Deep Learning
Hands-on Experience with Neural Network Console(40min)
2020年 Deep Learningをはじめよう！
Deep Learning習得のコツ
Autoencoderを用いた教師なし異常検知
Deep Learning入門 (50min)
Neural Network Consoleの紹介
プレゼンテーション
実践Deep Learning

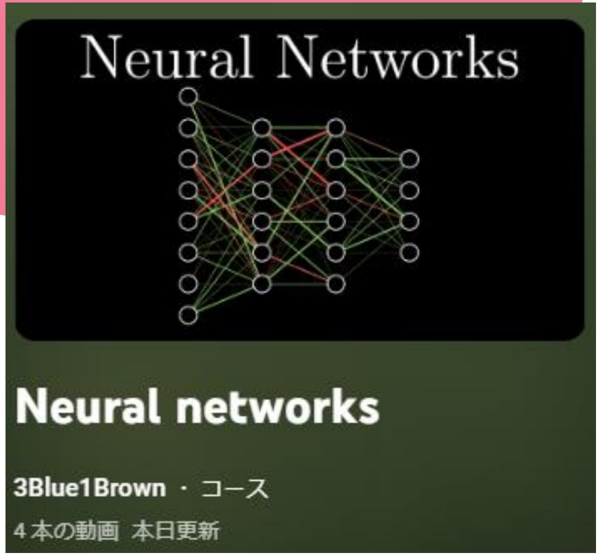

微分の連鎖律の復習からはじめる
誤差逆伝播法
何が「誤差」で
何が「逆」なのか
丁寧に説明します
50:14


O'REILLY
オライリー・ジャパン
第2版
scikit-learn、Keras、
TensorFlowによる
実践機械学習
powered by
jupyter
Aurélien Géron 著
下田 倫大 監訳
長尾 高弘 訳

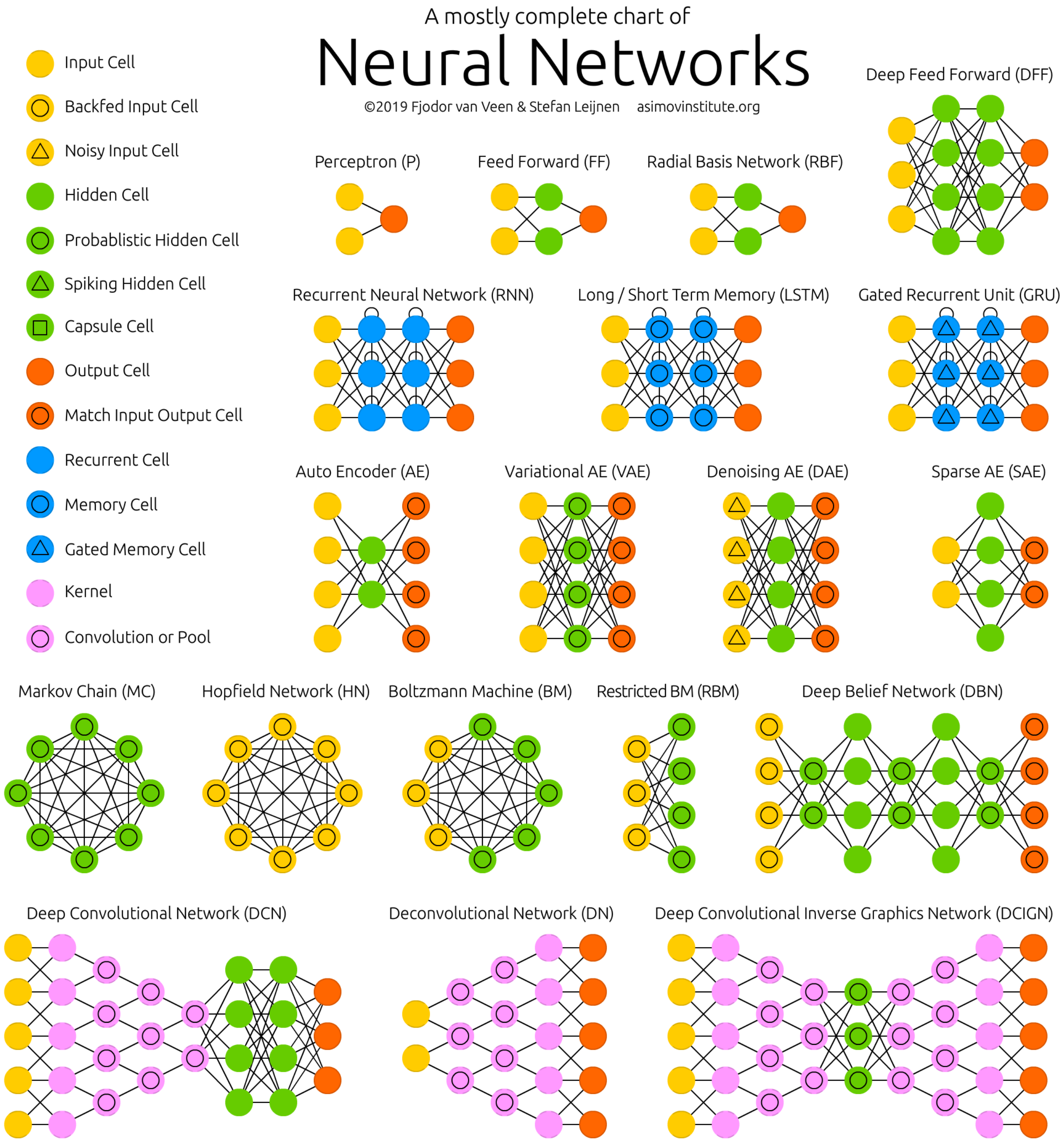

ゼロから作る
Deep Learning
Pythonで学ぶディープラーニングの理論と実装


Epoch
000,000
DATA
FEATURES
2 HIDDEN LAYERS
OUTPUT
REGENERATE

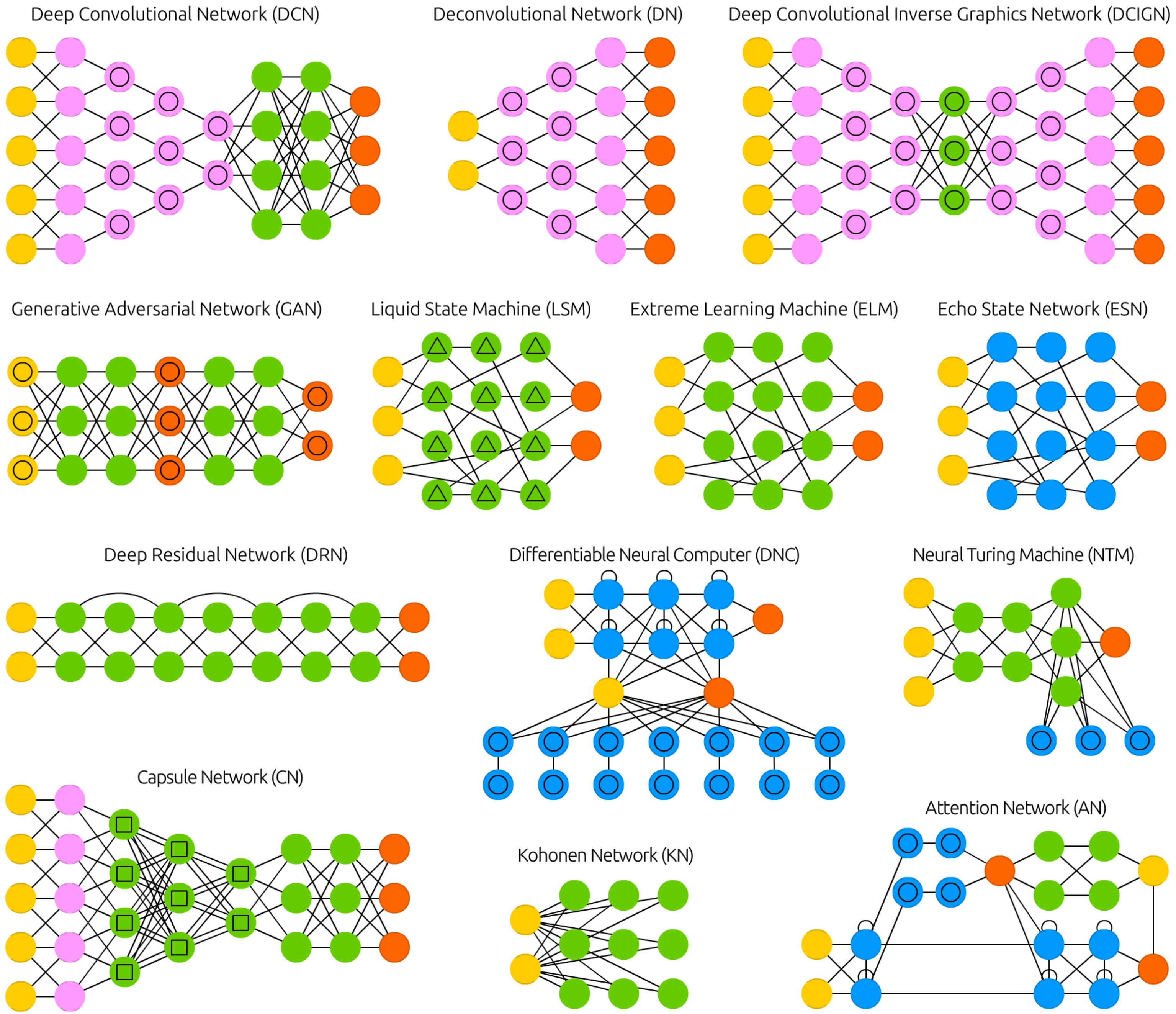

DEEP
LEARNING 入門
徹底解説
Part 0
初心者のため
の前準備
ディープラーニング入門：
仕組み理解から実装まで
某処生活_LiveSomewhere
5 本の動画 996 回視聴 最終更新日: 2021/02/28


Neural Networks
Neural networks
3Blue1Brown ・ コース
4 本の動画 本日更新

全体の流れ

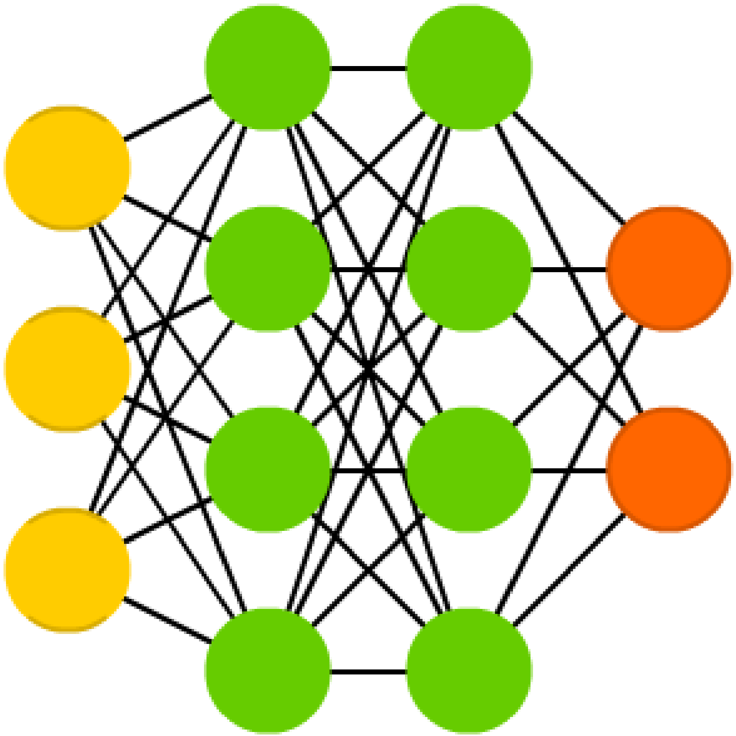

A mostly complete chart of
Neural Networks
©2019 Fjodor van Veen & Stefan Leijnen asimovinstitute.org
Input Cell
Backfed Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Capsule Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Gated Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)

全体の流れ


Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Differentiable Neural Computer (DNC)
Neural Turing Machine (NTM)
Capsule Network (CN)
Kohonen Network (KN)
Attention Network (AN)

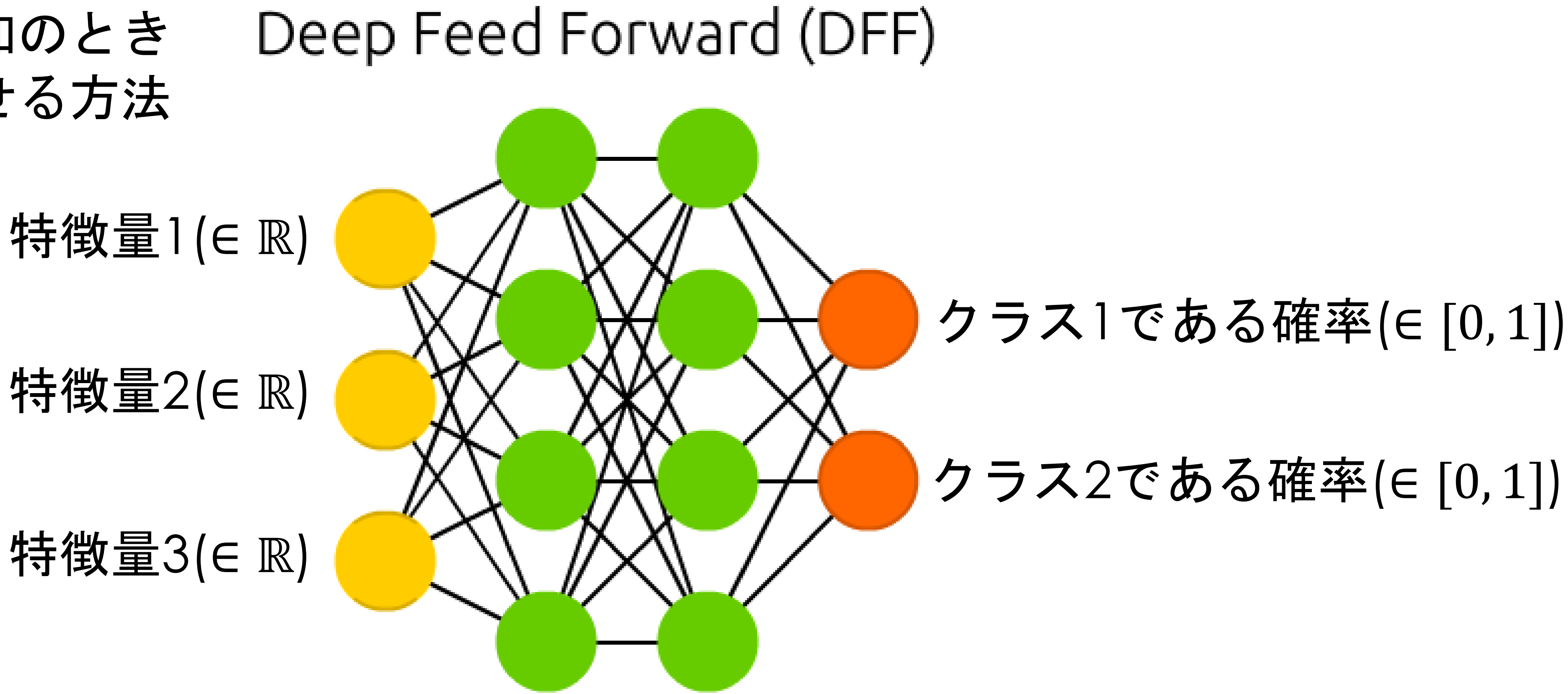
全体の流れ-全結合ニューラルネットワーク

Deep Feed Forward (DFF)



全体の流れ-前回

適切な重みが既知のとき
データを伝搬させる方法
を学んだ



全体の流れ-今回

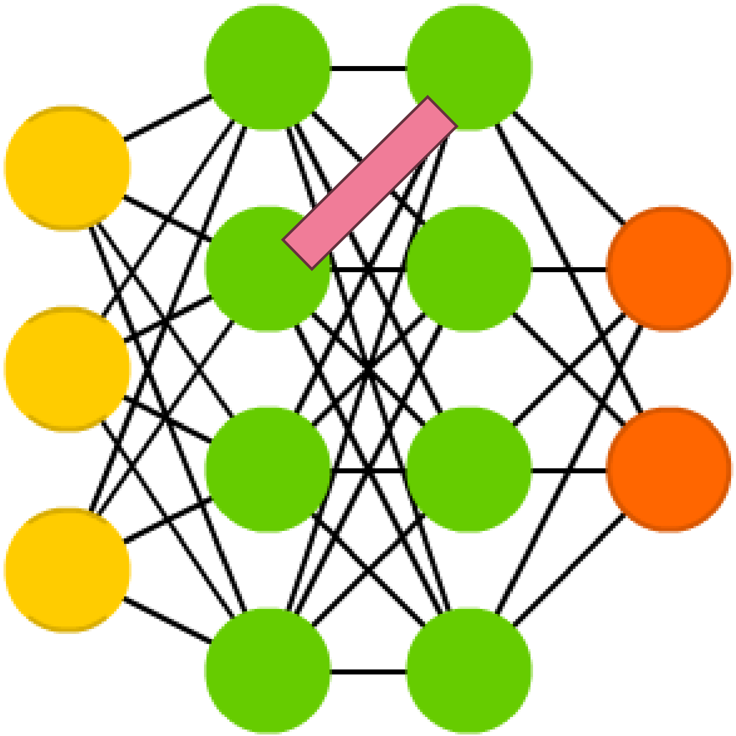
特徴量に対して
何が正解か
わかっているとき
重みを求める

Deep Feed Forward (DFF)

特徴量1($\in \mathbb{R}$)

特徴量2($\in \mathbb{R}$)

特徴量3($\in \mathbb{R}$)



正解ラベル1($\in [0,1]$)

正解ラベル0($\in [0,1]$)

偏微分復習

iPadのご準備を

偏微分-演習(ちょいムズ)

$$z = ax + by, \qquad \frac{\partial z}{\partial x} = ?$$

偏微分-演習(ちょいムズ)

$$z = ax + by, \qquad \frac{\partial z}{\partial x} = ?$$

冗談

偏微分-連鎖律 1変数

$$y = f(x), z = g(y)$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$$

偏微分-連鎖律 2変数

$$z = z(u, v),$$
$$u = u(x, y), v = v(x, y)$$

$$\frac{\partial z}{\partial x} = ?$$

偏微分-連鎖律 2変数

$$z = z(u, v),$$
$$u = u(x, y), v = v(x, y)$$

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

偏微分-連鎖律 多変数

$$z = z(u_1, u_2, \ldots, u_n),$$
$$u_1 = u_1(x_1, x_2, \ldots, x_m),$$
$$\ldots$$
$$u_n = u_n(x_1, x_2, \ldots, x_m)$$

$$\frac{\partial z}{\partial x_k} = ?$$

偏微分-連鎖律 多変数

$$z = z(u_1, u_2, \dots, u_n),$$
$$u_1 = u_1(x_1, x_2, \dots, x_m),$$
$$\dots$$
$$u_n = u_n(x_1, x_2, \dots, x_m)$$

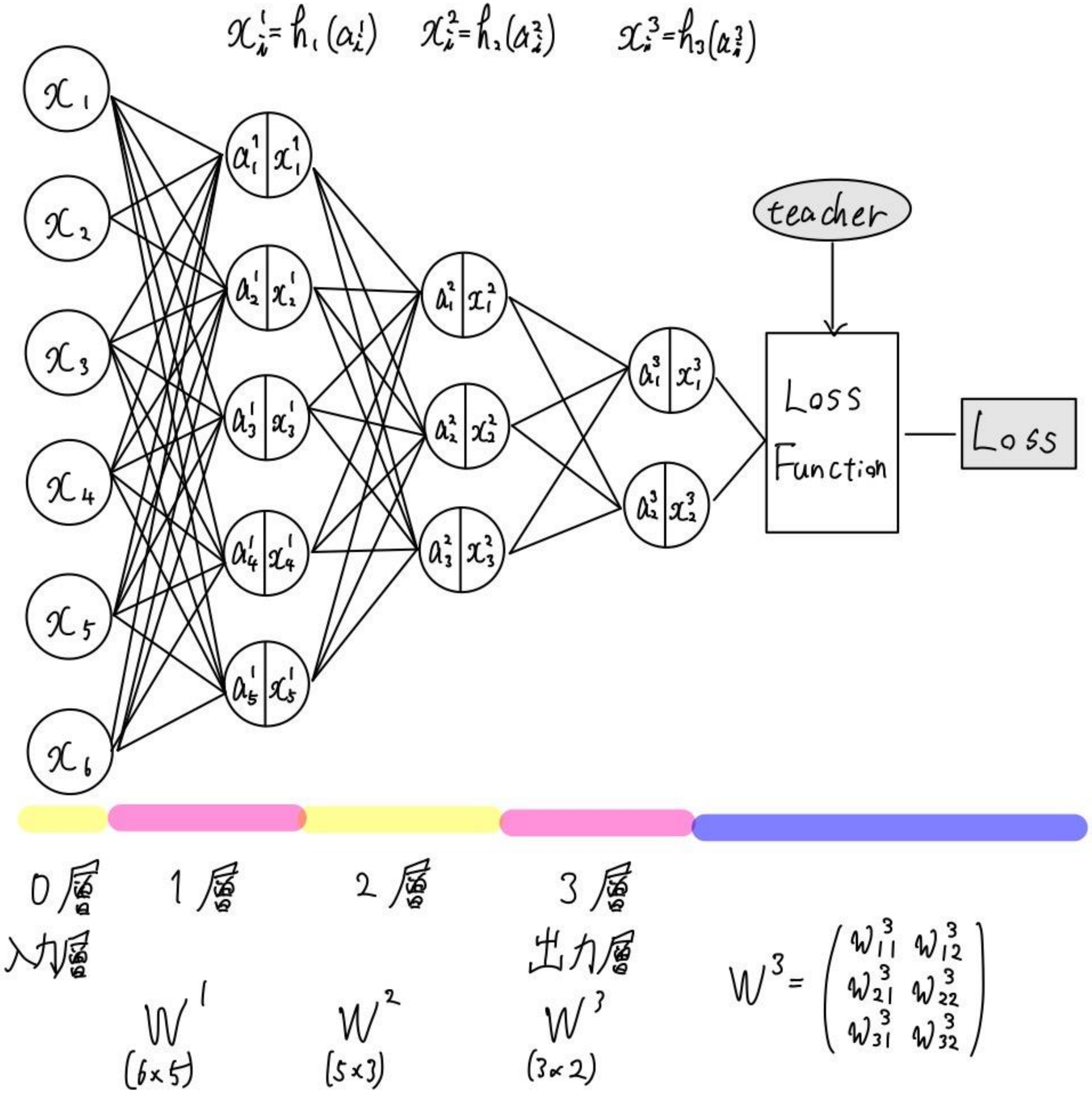
$$\frac{\partial z}{\partial x_k} = \frac{\partial z}{\partial u_1}\frac{\partial u_1}{\partial x_k} + \cdots + \frac{\partial z}{\partial u_n}\frac{\partial u_n}{\partial x_k} = \sum_{i=1}^{n} \frac{\partial z}{\partial u_i}\frac{\partial u_i}{\partial x_k}$$

偏微分復習完

誤差逆伝播法
オンライン

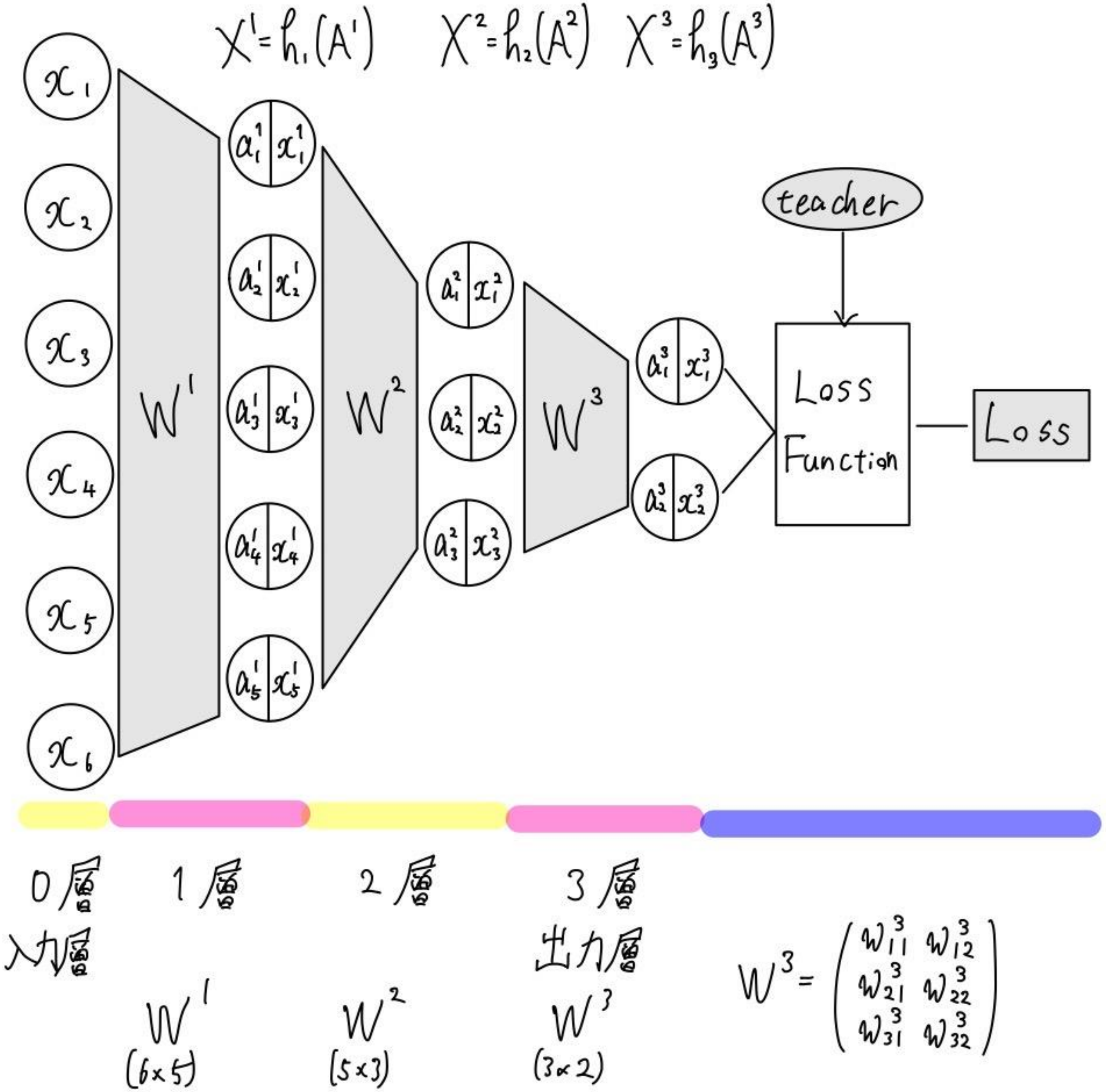
誤差逆伝播-記号導入

k層NN

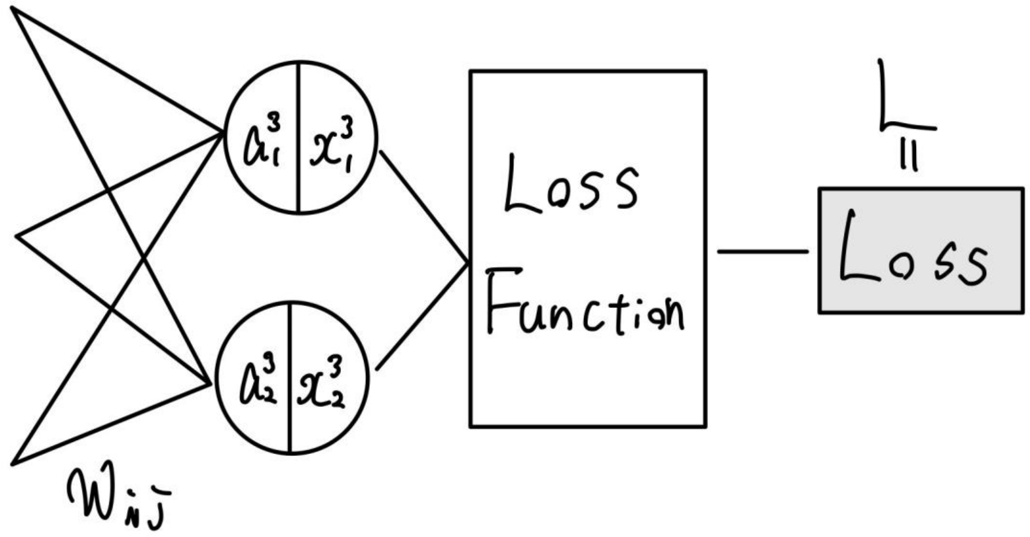


誤差逆伝播-記号導入

k層NN



誤差逆伝播-出力層の重みの更新

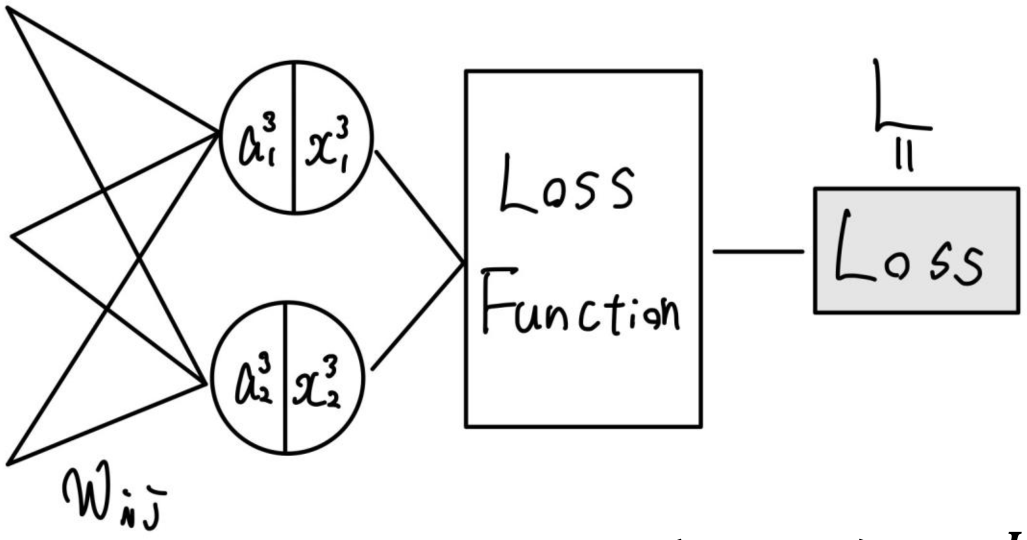


$$\frac{\partial L}{\partial {w_{ij}}^K} = \frac{\partial L}{\partial {x_j}^K}\frac{\partial {x_j}^K}{\partial {a_j}^K}\frac{\partial {a_j}^K}{\partial {w_{ij}}^K}$$

さっきのシグマは？

$$\frac{\partial z}{\partial x_k} = \sum_{i=1}^{n}\frac{\partial z}{\partial u_i}\frac{\partial u_i}{\partial x_k}$$

誤差逆伝播-出力層の重みの更新



$$\frac{\partial L}{\partial {w_{ij}}^K} = \frac{\partial L}{\partial {x_j}^K}\frac{\partial {x_j}^K}{\partial {a_j}^K}\frac{\partial {a_j}^K}{\partial {w_{ij}}^K}$$

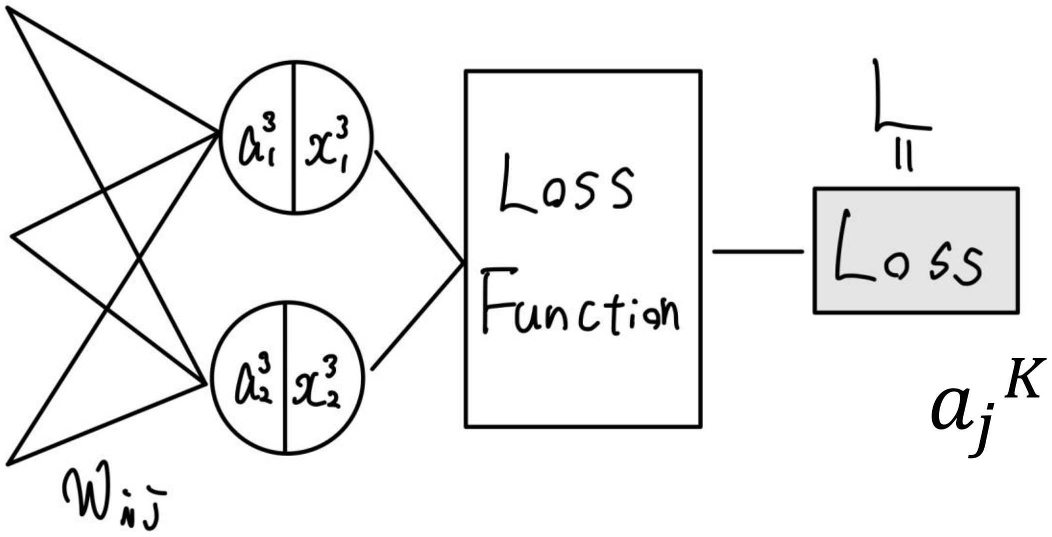
$$L = L(\dots, {a_{j-1}}^K, {a_j}^K, {a_{j+1}}^K, \dots)$$
$${a_j}^K = {a_j}^K({w_{1j}}^K \dots, {w_{i-1,j}}^K, {w_{ij}}^K, \dots)$$

$$(l \neq j), {a_l}^K = {a_l}^K({w_{1l}}^K \dots, {w_{i-1,l}}^K, {w_{il}}^K, \dots)$$

というように${w_{ij}}^K$が引数にない

$\therefore {w_{ij}}^K$での偏微分が0

誤差逆伝播-出力層の重みの更新

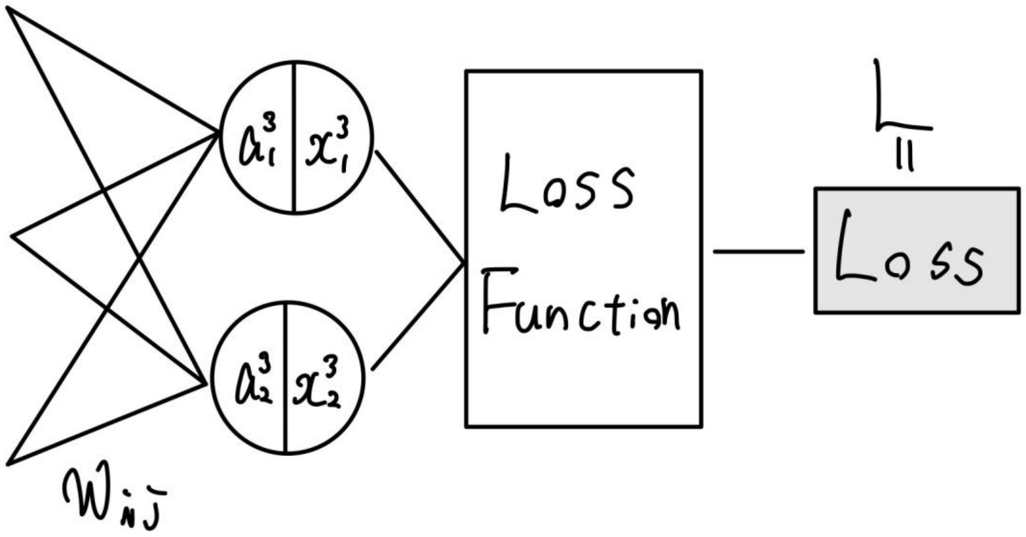


$$\frac{\partial L}{\partial w_{ij}{}^K} = \frac{\partial L}{\partial x_j{}^K}\frac{\partial x_j{}^K}{\partial a_j{}^K}\frac{\partial a_j{}^K}{\partial w_{ij}{}^K}$$

$$a_j{}^K = w_{1j}{}^K x_1{}^{K-1} + \cdots + w_{ij}{}^K x_i{}^{K-1} + \cdots$$

$$\frac{\partial a_j{}^K}{\partial w_{ij}{}^K} = x_i{}^{K-1}$$←データが入れば既知

誤差逆伝播-出力層の重みの更新



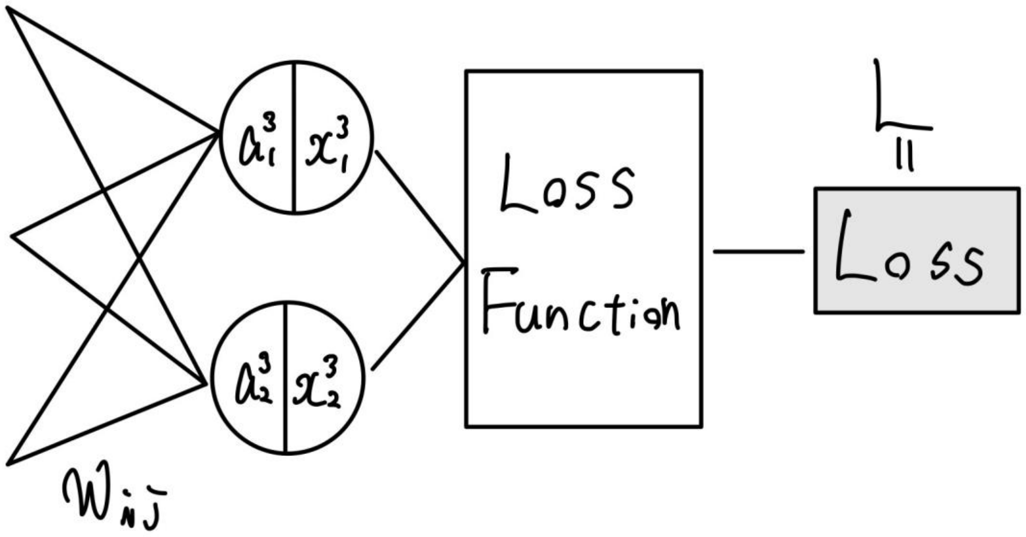
$$\frac{\partial L}{\partial w_{ij}{}^K} = \frac{\partial L}{\partial x_j{}^K} \frac{\partial x_j{}^K}{\partial a_j{}^K} \frac{\partial a_j{}^K}{\partial w_{ij}{}^K}$$

$x_j{}^K = h_K(a_j{}^K)$：hは活性化関数

出力層の活性化関数はsoftmax関数

$$x_j{}^K = \frac{e^{a_j{}^K}}{\sum_{i=1}^{n} e^{a_i{}^K}}$$

誤差逆伝播-出力層の重みの更新



$$\frac{\partial x_j{}^K}{\partial a_j{}^K} \qquad x_j{}^K = \frac{e^{a_j{}^K}}{\sum_{i=1}^{n} e^{a_i{}^K}}$$

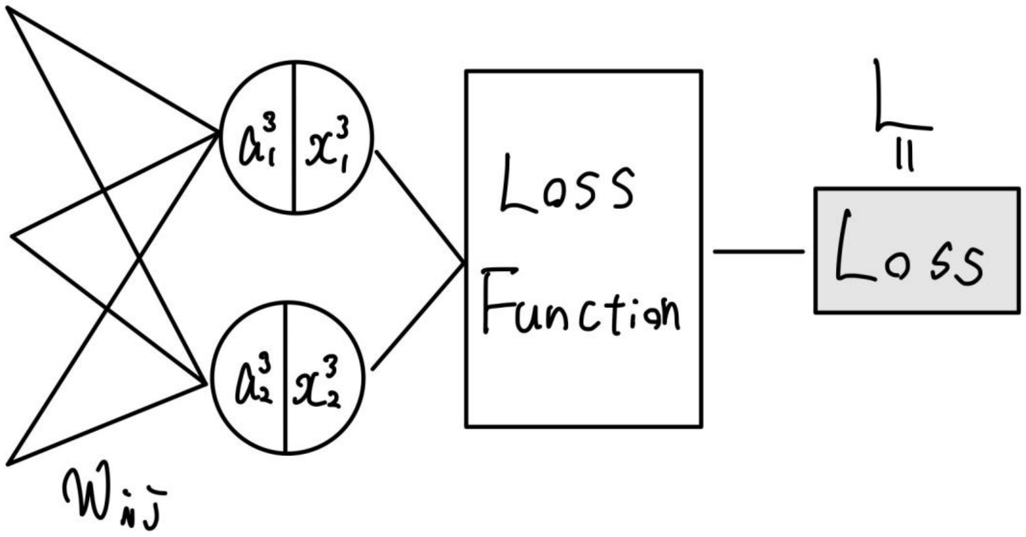
$$\frac{\partial x_j{}^K}{\partial a_j{}^K} = \frac{e^{a_j{}^K}\sum_{i=1}^{n} e^{a_i{}^K} - e^{a_j{}^K}e^{a_j{}^K}}{\left(\sum_{i=1}^{n} e^{a_i{}^K}\right)^2}$$

$$= e^{a_j{}^K}\frac{\sum_{i=1}^{n} e^{a_i{}^K} - e^{a_j{}^K}}{\left(\sum_{i=1}^{n} e^{a_i{}^K}\right)^2}$$

$$= \frac{e^{a_j{}^K}}{\sum_{i=1}^{n} e^{a_i{}^K}}\frac{\sum_{i=1}^{n} e^{a_i{}^K} - e^{a_j{}^K}}{\sum_{i=1}^{n} e^{a_i{}^K}}$$

$= x_j{}^K\left(1 - x_j{}^K\right)$ ←データが入れば既知

誤差逆伝播-出力層の重みの更新



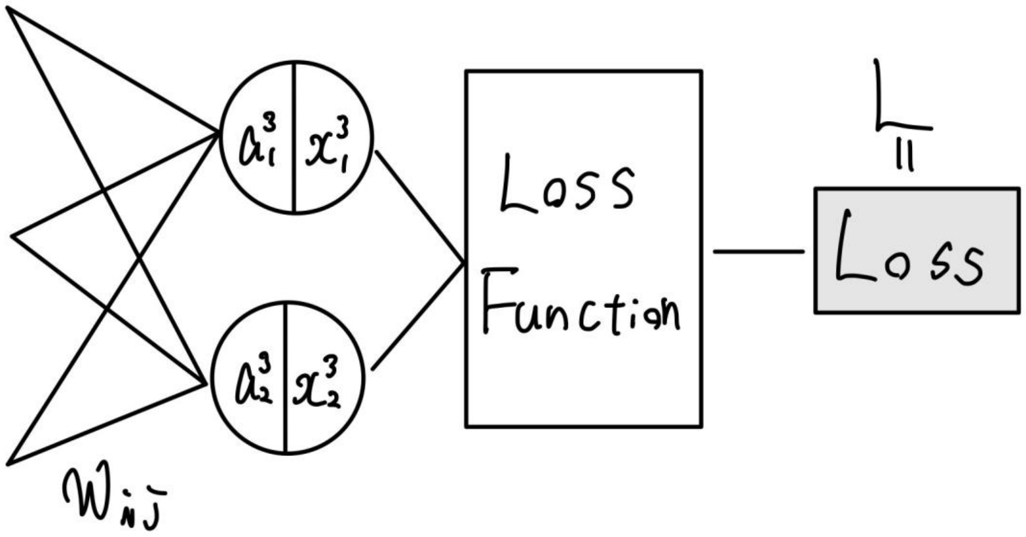
$$\frac{\partial L}{\partial {w_{ij}}^K} = \frac{\partial L}{\partial {x_j}^K} \frac{\partial {x_j}^K}{\partial {a_j}^K} \frac{\partial {a_j}^K}{\partial {w_{ij}}^K}$$

Loss Functionは今はCross Entropy
(平均二乗誤差MSEやRMSEなど様々)

$$L = -\sum_{i=1}^{n} t_i \ln {x_i}^K$$

誤差逆伝播-出力層の重みの更新

$$\frac{\partial L}{\partial {x_j}^K} \qquad L = -\sum_{i=1}^{n} t_i \ln {x_i}^K$$



$$\frac{\partial L}{\partial {x_j}^K} = -\sum_{i=1}^{n} t_i \frac{\partial}{\partial {x_j}^K}(\ln {x_i}^K)$$

$$= -\sum_{i=1}^{n} t_i \frac{\partial}{\partial {x_i}^K}(\ln {x_i}^K)\frac{\partial {x_i}^K}{\partial {x_j}^K}$$

$$= -\sum_{i=1}^{n} t_i \frac{1}{{x_i}^K}\frac{\partial {x_i}^K}{\partial {x_j}^K}$$

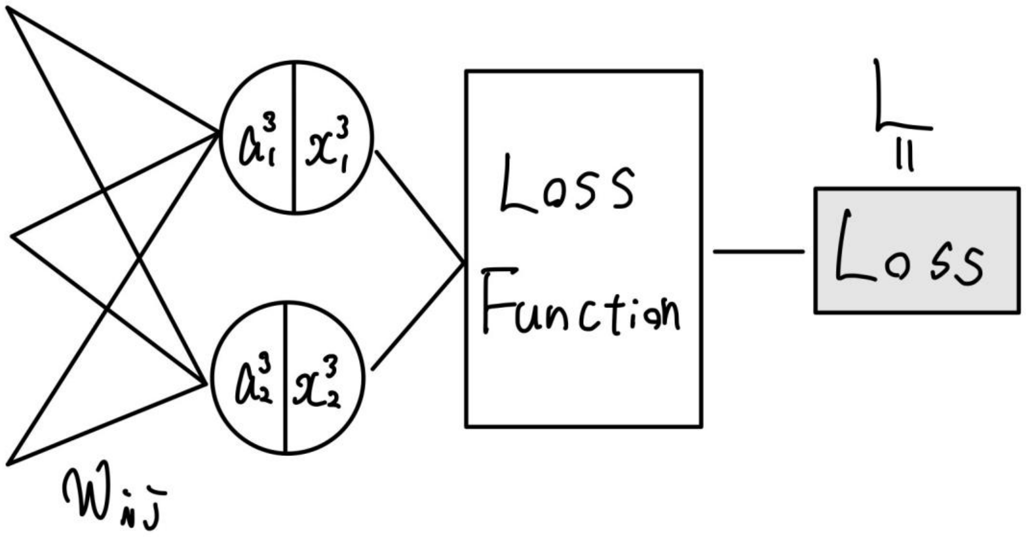
$$({x_1}^K + {x_2}^K + \cdots + {x_n}^K = 1$$
$$\Leftrightarrow {x_j}^K = 1 - {x_i}^K - \cdots)$$

$$\frac{\partial {x_i}^K}{\partial {x_j}^K} = \begin{cases} 1(i = j) \\ -1(i \neq j) \end{cases}$$

$$= -\frac{t_j}{{x_j}^K} + \sum_{i=1 \land i \neq j}^{n} \frac{t_i}{{x_i}^K}$$

←データが入れば既知

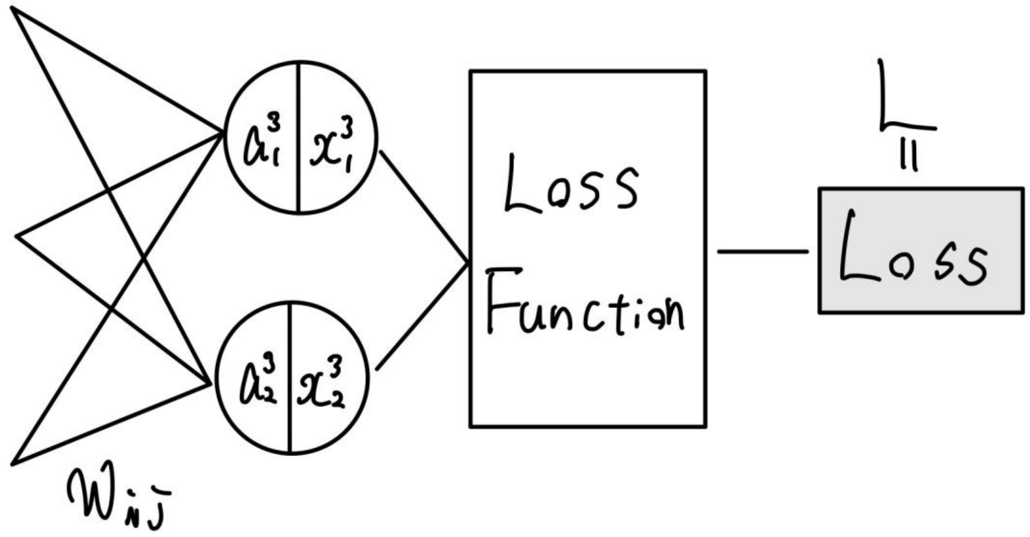
誤差逆伝播-出力層の重みの更新



$$\frac{\partial L}{\partial {w_{ij}}^K} = \frac{\partial L}{\partial {x_j}^K}\frac{\partial {x_j}^K}{\partial {a_j}^K}\frac{\partial {a_j}^K}{\partial {w_{ij}}^K}$$

$$\frac{\partial L}{\partial {w_{ij}}^K} = \left(-\frac{t_j}{{x_j}^K} + \sum_{i=1 \wedge i \neq j}^{n} \frac{t_i}{{x_i}^K}\right) {x_j}^K\left(1 - {x_j}^K\right) {x_i}^{K-1}$$

誤差逆伝播-演習①出力層のバイアスの更新

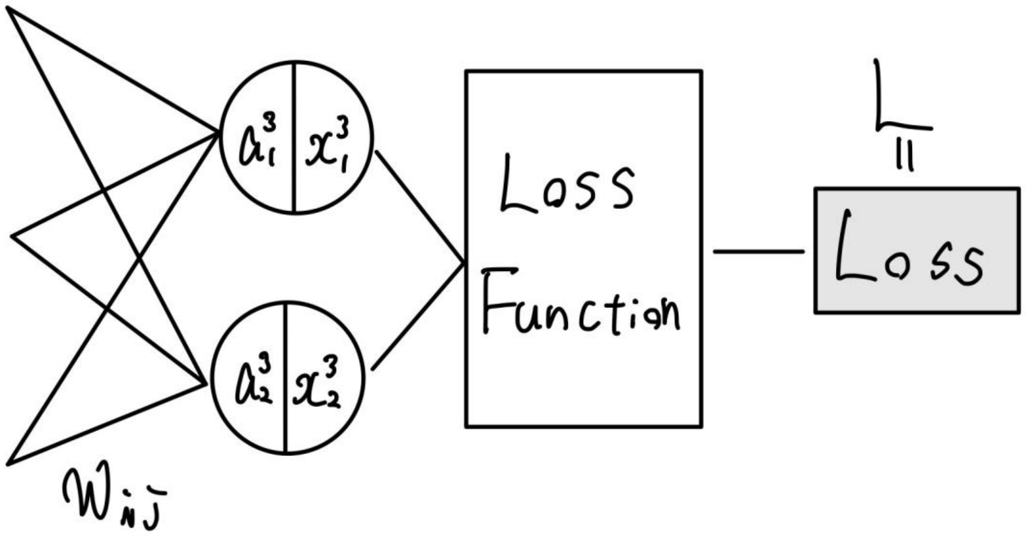


$\dfrac{\partial L}{\partial b_j{}^K}$を求めよ

既知なx, tのみを用いて表すこと

$$\frac{\partial L}{\partial w_{ij}{}^K} = \left(-\frac{t_j}{x_j{}^K} + \sum_{i=1 \wedge i \neq j}^{n} \frac{t_i}{x_i{}^K}\right) x_j{}^K\left(1 - x_j{}^K\right) x_i{}^{K-1}$$

誤差逆伝播-演習①出力層のバイアスの更新

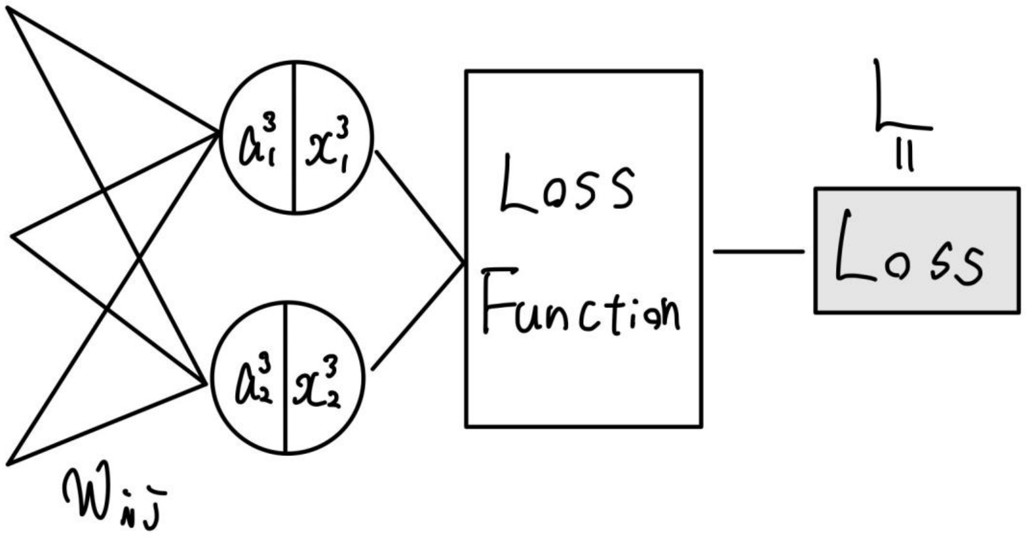


$$\frac{\partial L}{\partial {b_j}^K}$$を求めよ

$$\frac{\partial L}{\partial {b_j}^K} = \frac{\partial L}{\partial {x_j}^K} \frac{\partial {x_j}^K}{\partial {a_j}^K} \frac{\partial {a_j}^K}{\partial {b_j}^K}$$

$$\frac{\partial L}{\partial {b_j}^K} = \left(-\frac{t_j}{{x_j}^K} + \sum_{i=1 \wedge i \neq j}^{n} \frac{t_i}{{x_i}^K} \right) {x_j}^K \left(1 - {x_j}^K\right) 1$$

誤差逆伝播-出力層の重みの更新

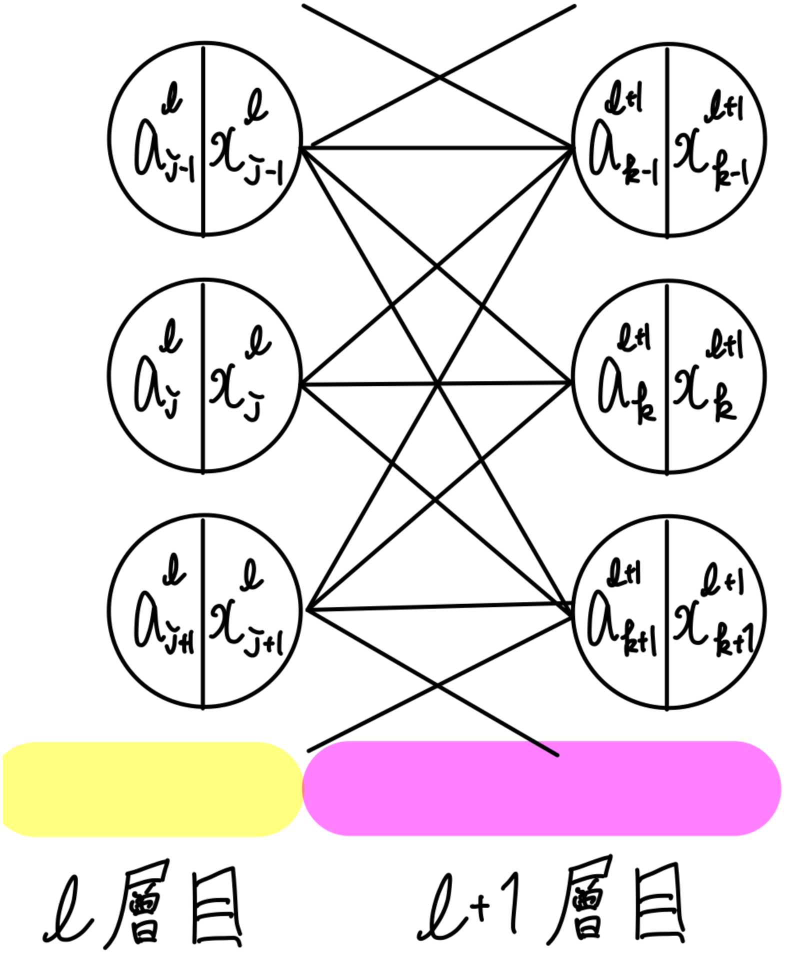


$$\frac{\partial L}{\partial w_{ij}{}^K} = \frac{\partial L}{\partial x_j{}^K}\frac{\partial x_j{}^K}{\partial a_j{}^K}\frac{\partial a_j{}^K}{\partial w_{ij}{}^K}$$

$$\delta_j{}^K := \frac{\partial L}{\partial a_j{}^K}$$

誤差

誤差逆伝播-中間層の重みの更新

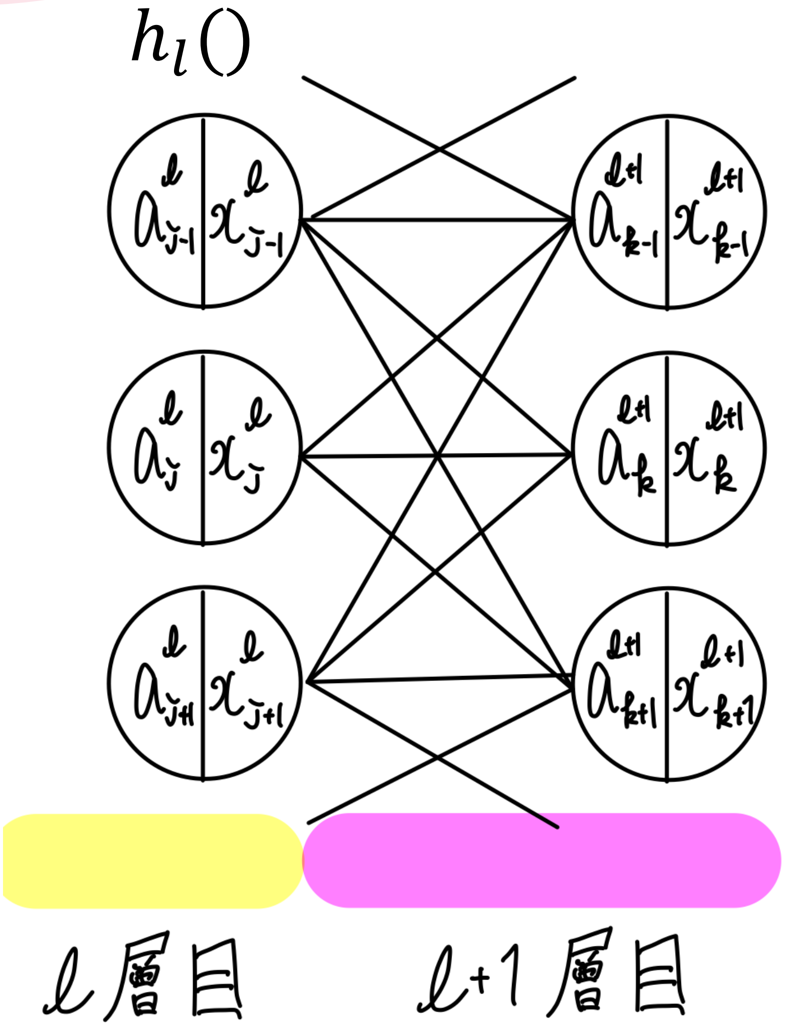


$$\frac{\partial L}{\partial {w_{ij}}^{l}} = {\delta_j}^{l} \frac{\partial {a_j}^{l}}{\partial {w_{ij}}^{l}}$$

$${a_j}^{l} = {w_{1j}}^{l} {x_1}^{l-1} + \cdots + {w_{ij}}^{l} {x_i}^{l-1} + \cdots$$

$$\frac{\partial {a_j}^{l}}{\partial {w_{ij}}^{l}} = {x_i}^{l-1}$$

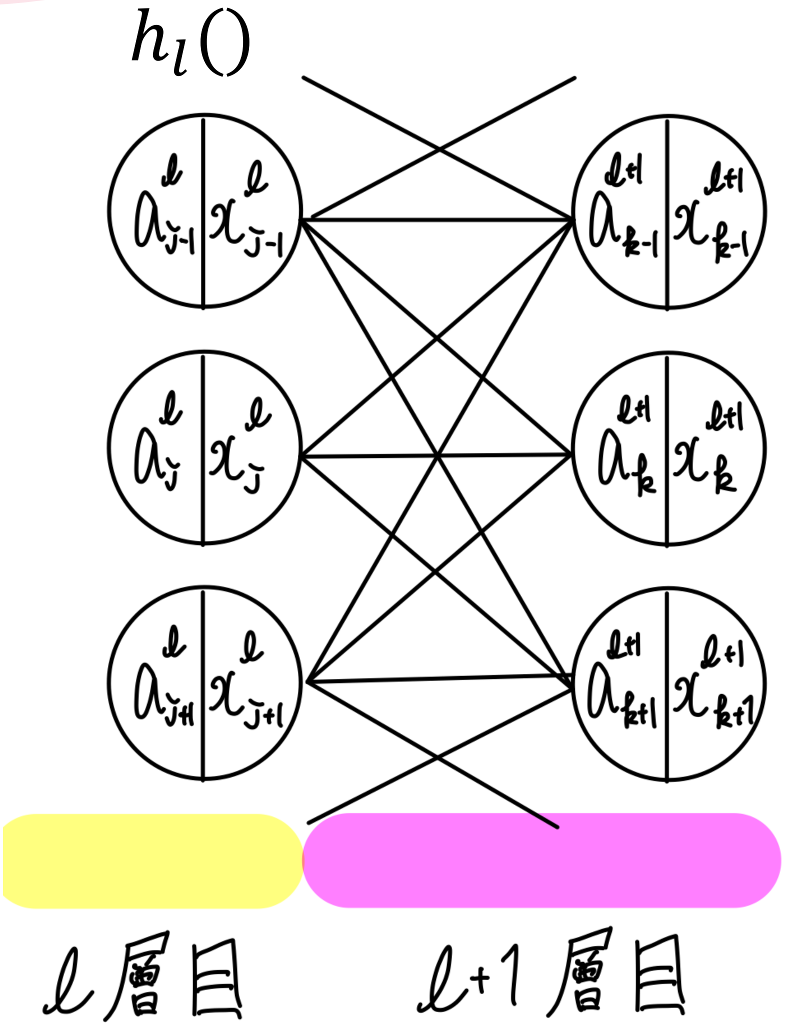
誤差逆伝播-中間層の重みの更新



$$\frac{\partial L}{\partial {w_{ij}}^{l}} = {\delta_j}^{l} \frac{\partial {a_j}^{l}}{\partial {w_{ij}}^{l}}$$

$$\begin{aligned}
{\delta_j}^{l} = \frac{\partial L}{\partial {a_j}^{l}} &= \sum_{k=1}^{n} \frac{\partial L}{\partial {a_k}^{l+1}} \frac{\partial {a_k}^{l+1}}{\partial {a_j}^{l}} \\
&= \sum_{k=1}^{n} \frac{\partial L}{\partial {a_k}^{l+1}} \frac{\partial {a_k}^{l+1}}{\partial {x_j}^{l}} \frac{\partial {x_j}^{l}}{\partial {a_j}^{l}} \\
&= \sum_{k=1}^{n} {\delta_k}^{l+1} \frac{\partial {a_k}^{l+1}}{\partial {x_j}^{l}} \frac{\partial {x_j}^{l}}{\partial {a_j}^{l}}
\end{aligned}$$

誤差逆伝播-中間層の重みの更新

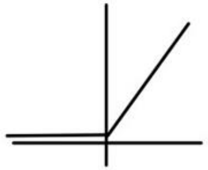


$$\frac{\partial L}{\partial {w_{ij}}^l} = {\delta_j}^l \frac{\partial {a_j}^l}{\partial {w_{ij}}^l}$$

$$ {\delta_j}^l = \frac{\partial L}{\partial {a_j}^l} = \sum_{k=1}^{n} \frac{\partial L}{\partial {a_k}^{l+1}} \frac{\partial {a_k}^{l+1}}{\partial {x_j}^l} \frac{\partial {x_j}^l}{\partial {a_j}^l}$$

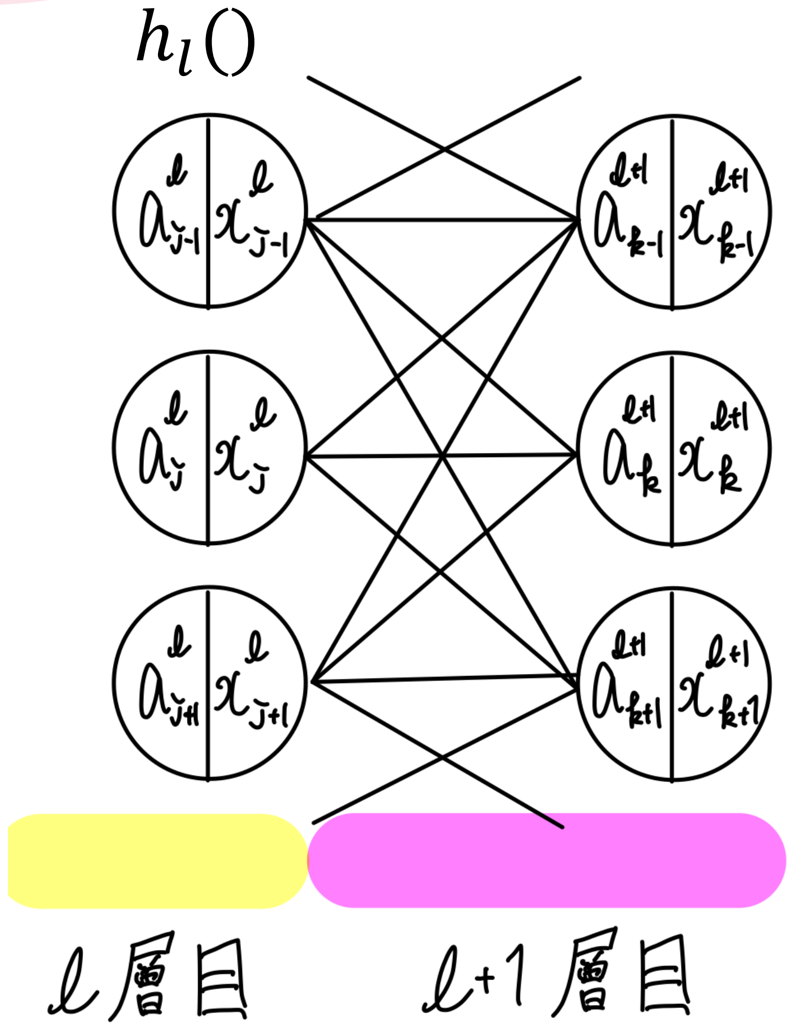
$$= \sum_{k=1}^{n} {\delta_k}^{l+1} \frac{\partial {a_k}^{l+1}}{\partial {x_j}^l} {h_l}'({a_j}^l)$$

・ReLU関数 $f(x) = \begin{cases} x & , x \geq 0 \\ 0 & , x < 0 \end{cases}$



主にこれを使っていく

誤差逆伝播-中間層の重みの更新



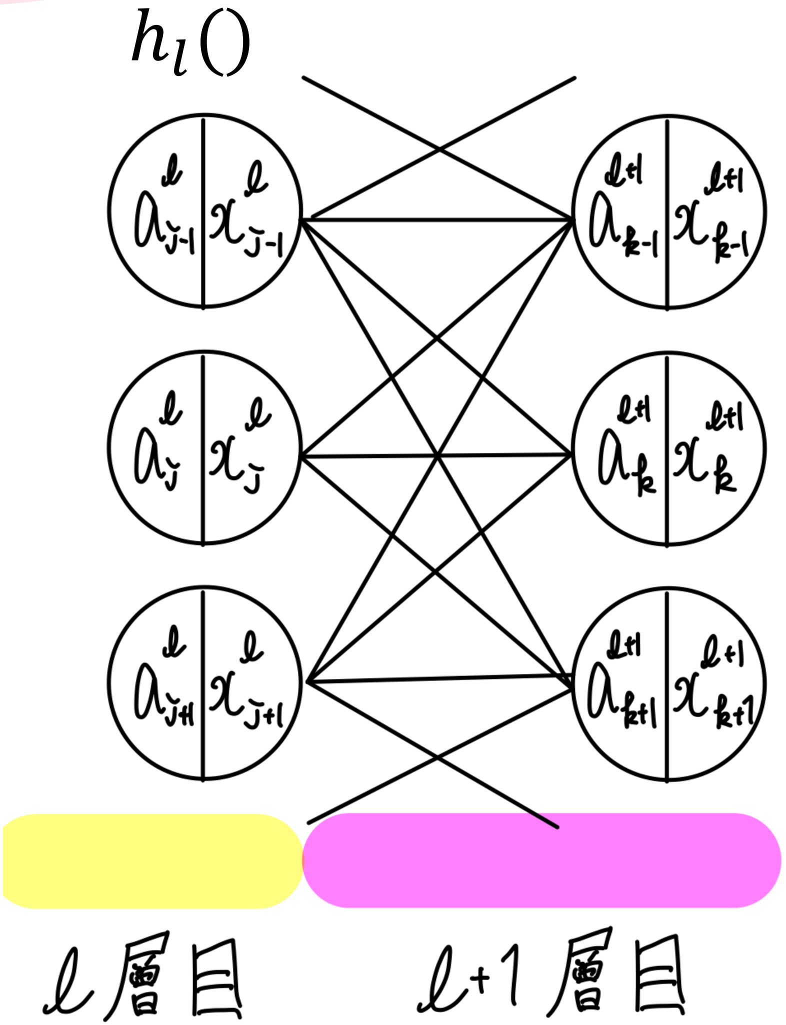
$$\frac{\partial L}{\partial {w_{ij}}^{l}} = {\delta_j}^{l} \frac{\partial {a_j}^{l}}{\partial {w_{ij}}^{l}}$$

$${\delta_j}^{l} = \frac{\partial L}{\partial {a_j}^{l}} = \sum_{k=1}^{n} \frac{\partial L}{\partial {a_k}^{l+1}} \frac{\partial {a_k}^{l+1}}{\partial {x_j}^{l}} \frac{\partial {x_j}^{l}}{\partial {a_j}^{l}}$$

$$= \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{jk}}^{l+1} h_l'({a_j}^{l})$$

更新前の重み

誤差逆伝播-中間層の重みの更新



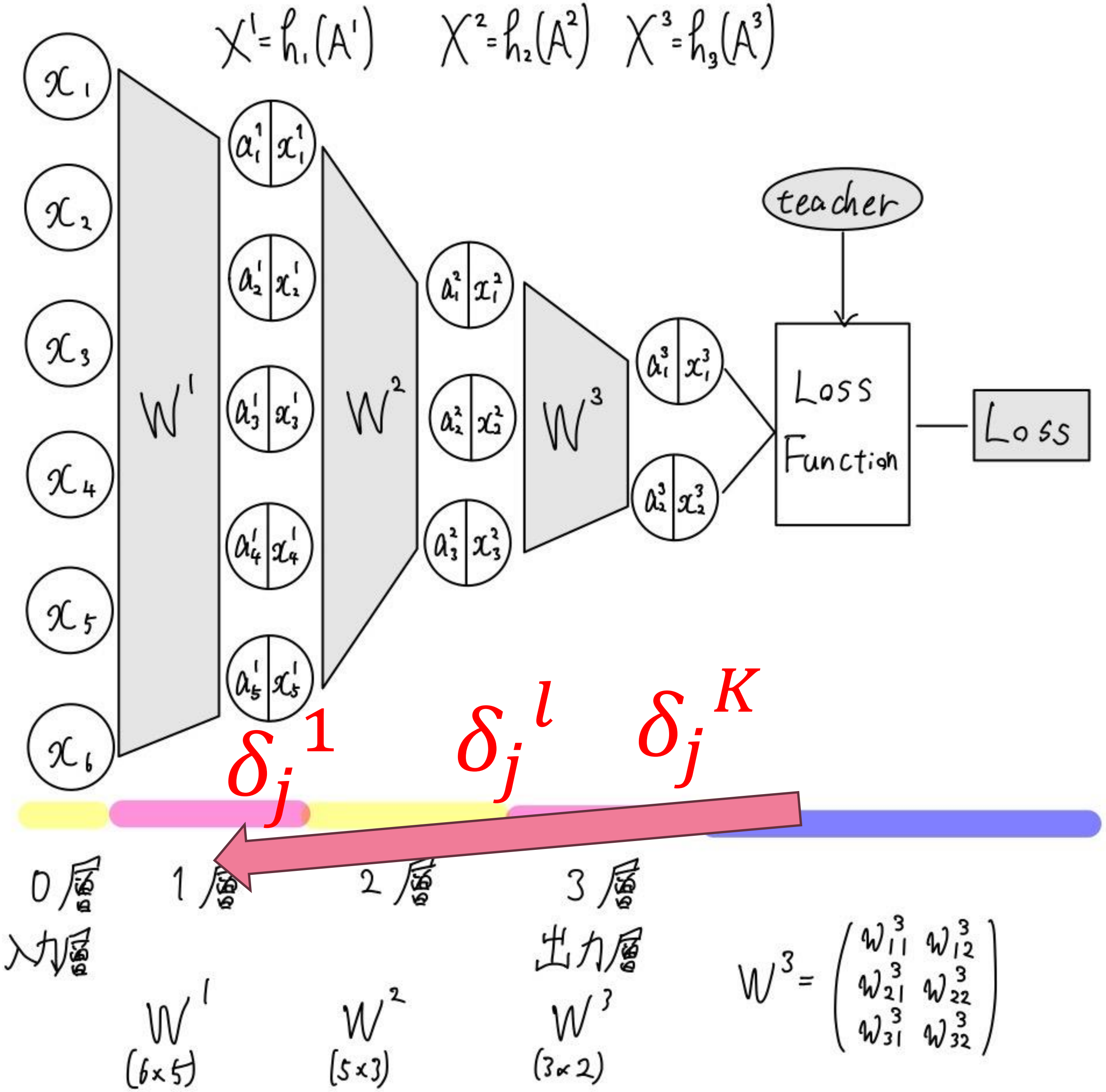
$$\frac{\partial L}{\partial w_{ij}{}^{l}} = \delta_j{}^{l} \frac{\partial a_j{}^{l}}{\partial w_{ij}{}^{l}}$$

$$\delta_j{}^{l} = \sum_{k=1}^{n} \delta_k{}^{l+1} w_{jk}{}^{l+1} h_l'(a_j{}^{l})$$

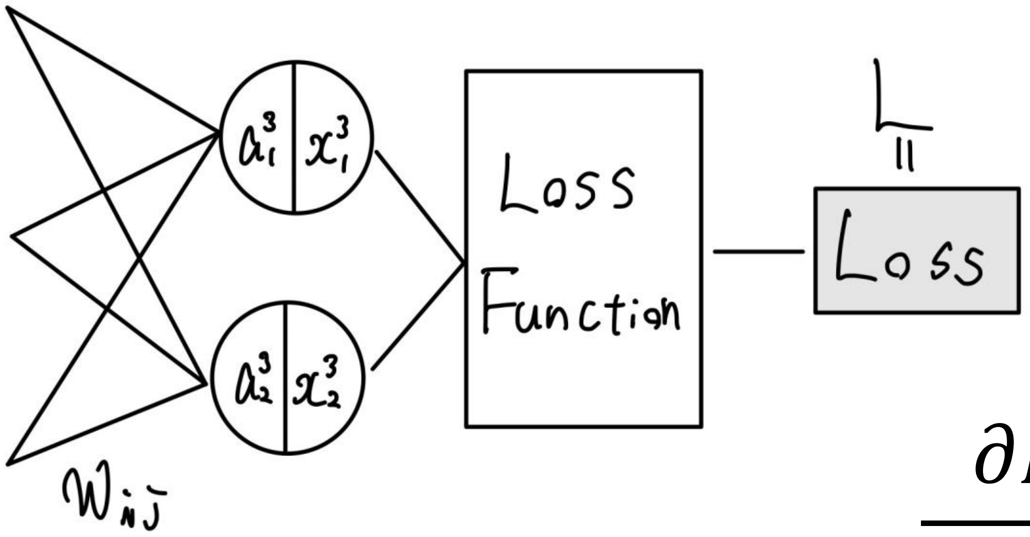
これって誤差の漸化式
後ろの層の誤差がわかればよい

誤差逆伝播

誤差$\delta_j{}^l$が
逆方向に
伝播してる



誤差逆伝播-演習②中間層のバイアスの更新



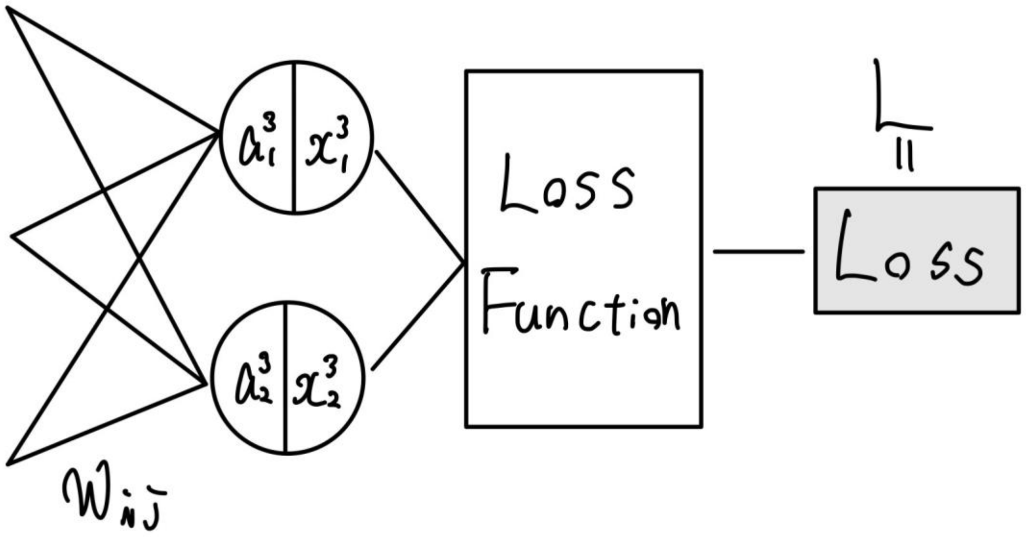
$\dfrac{\partial L}{\partial {b_j}^l}$を求めよ

既知なa, x, w, δ^{l+1}などを用いて表すこと

$$\frac{\partial L}{\partial {w_{ij}}^l} = {\delta_j}^l {x_i}^{l-1}$$

$${\delta_j}^l = \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{jk}}^{l+1} h_l'({a_j}^l)$$

誤差逆伝播-演習②中間層のバイアスの更新

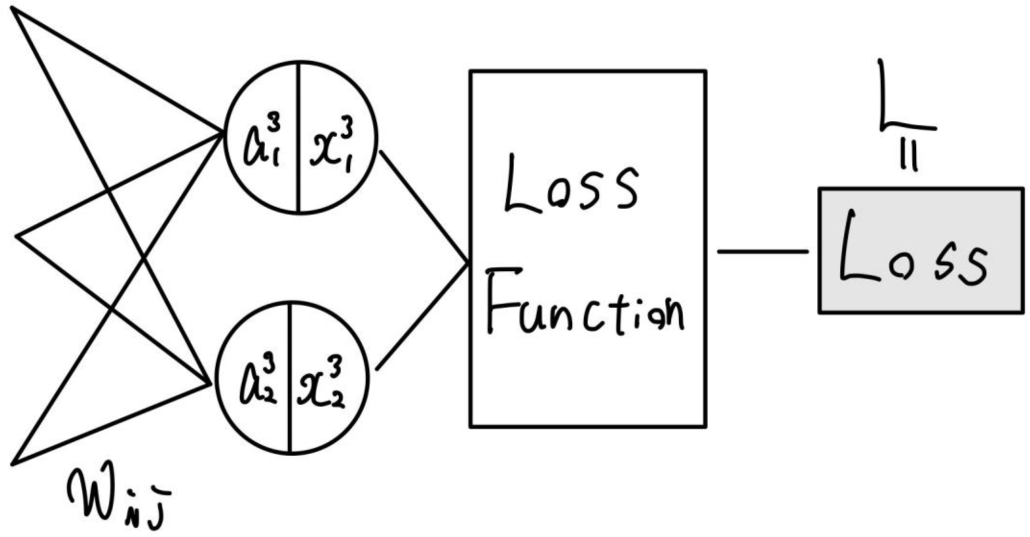


$\frac{\partial L}{\partial {b_j}^l}$を求めよ

既知なa, x, w, δ^{l+1}のみを用いて表すこと

$$\frac{\partial L}{\partial {b_j}^l} = {\delta_j}^l \frac{\partial {a_j}^l}{\partial {b_j}^l}$$

誤差逆伝播-演習②中間層のバイアスの更新



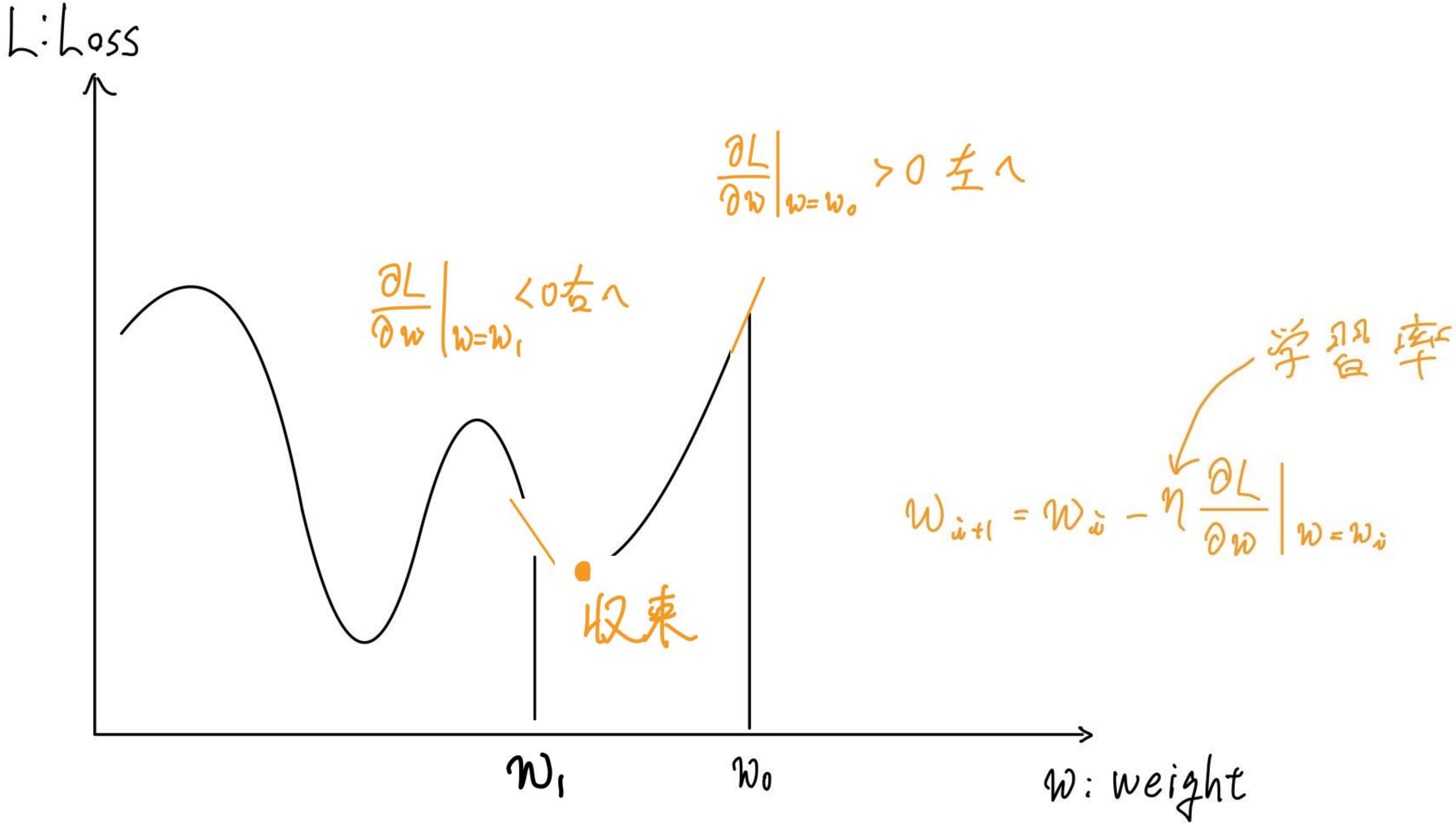
$\frac{\partial L}{\partial {b_j}^l}$を求めよ

既知なa, x, w, δ^{l+1}のみを用いて表すこと

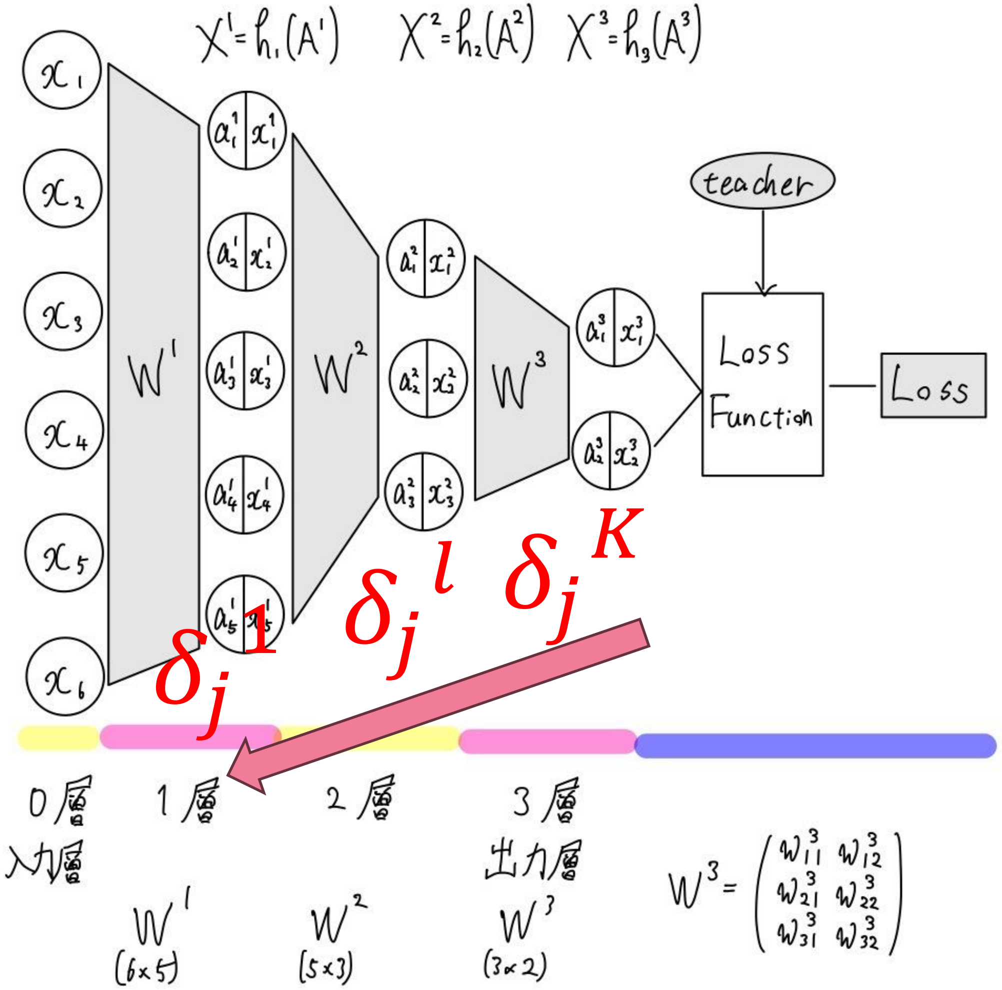
$$\frac{\partial L}{\partial {b_j}^l} = {\delta_j}^l 1$$

$${\delta_j}^l = \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{jk}}^{l+1} h_l'({a_j}^l)$$

ニューラルネットワーク-勾配法　再掲

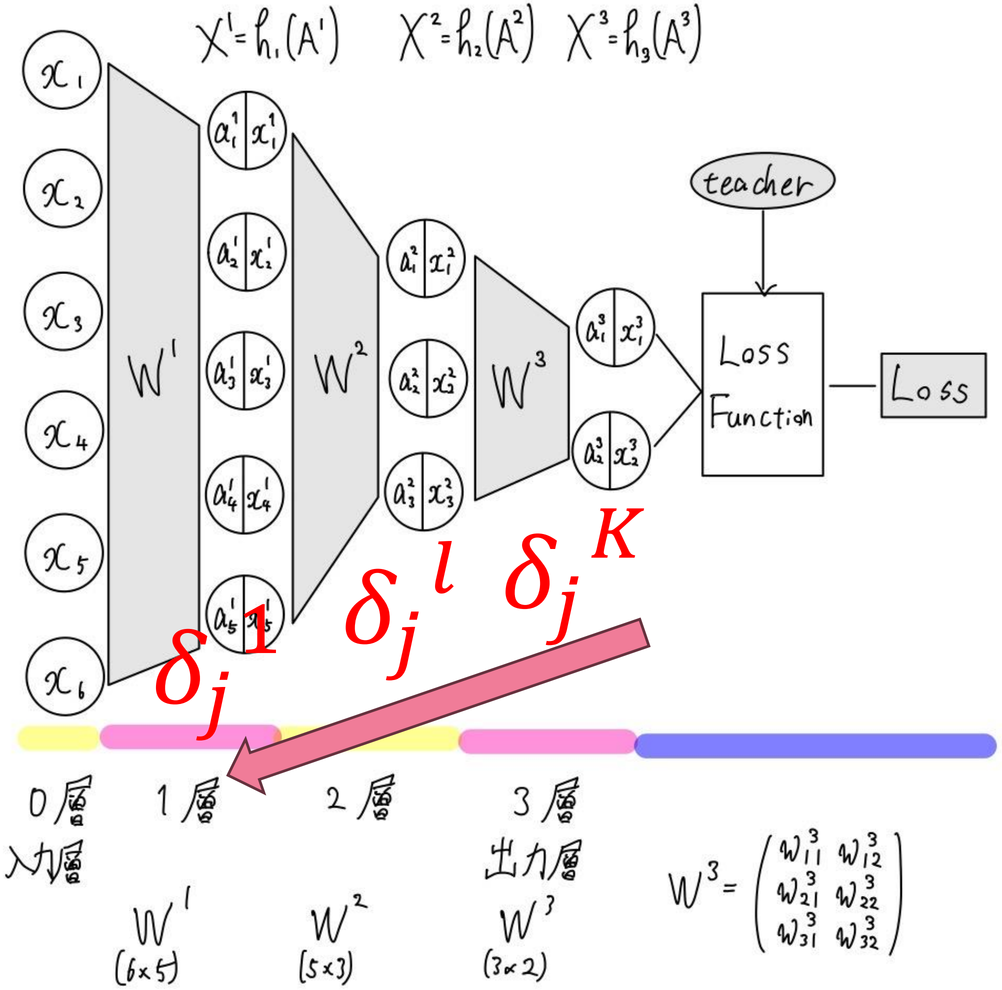


誤差逆伝播-中間層の重みの更新



$$w_{ij}{}^{l} = w_{ij}{}^{l} - \eta \frac{\partial L}{\partial w_{ij}{}^{l}}$$

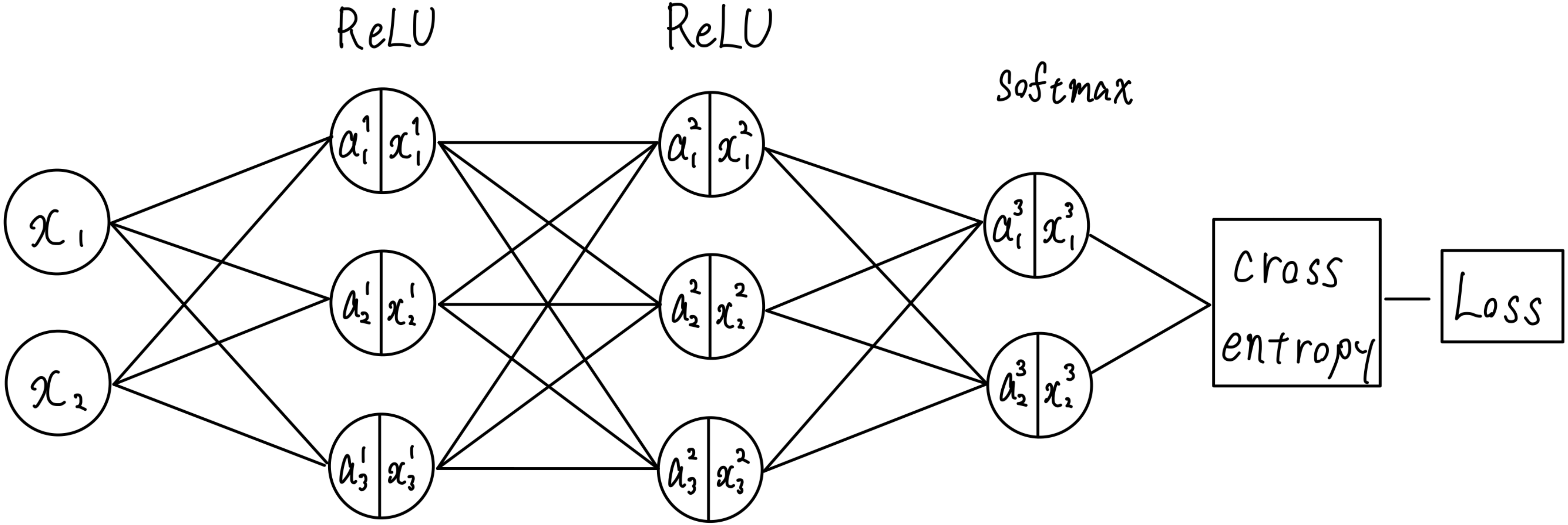
誤差逆伝播-中間層の重みの更新



$$w_{ij}{}^l = w_{ij}{}^l - \eta \frac{\partial L}{\partial w_{ij}{}^l}$$

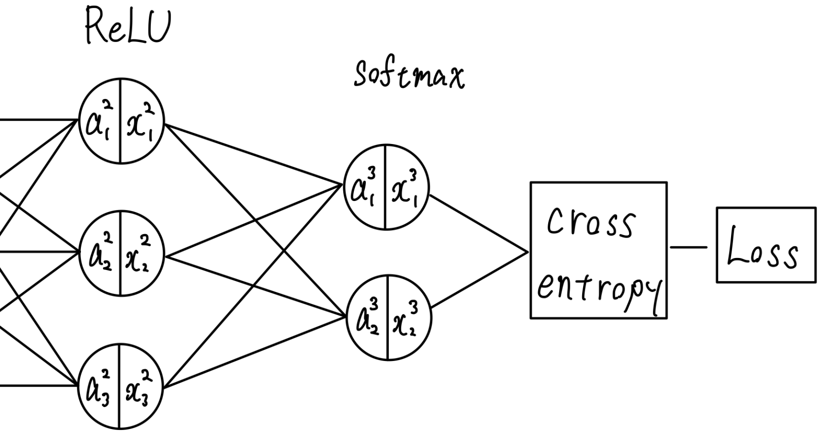
ニューラルネットワーク
の神髄を完全に理解した

誤差逆伝播-パラメータの更新　行列表現



誤差逆伝播-行列表現 出力層 重み

$$\frac{\partial L}{\partial {w_{ij}}^K} = {\delta_j}^K \frac{\partial {a_j}^K}{\partial {w_{ij}}^K} = {\delta_j}^K {x_i}^{K-1}$$

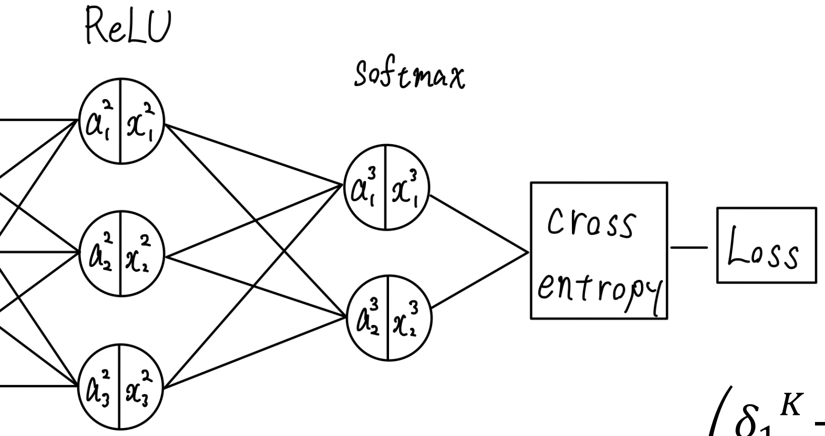


$$W^K = \begin{pmatrix} {w_{11}}^K & {w_{12}}^K \\ {w_{21}}^K & {w_{22}}^K \\ {w_{31}}^K & {w_{32}}^K \end{pmatrix}$$

$$\frac{\partial L}{\partial W^K} = \begin{pmatrix} \frac{\partial L}{\partial {w_{11}}^K} & \frac{\partial L}{\partial {w_{12}}^K} \\ \frac{\partial L}{\partial {w_{21}}^K} & \frac{\partial L}{\partial {w_{22}}^K} \\ \frac{\partial L}{\partial {w_{31}}^K} & \frac{\partial L}{\partial {w_{32}}^K} \end{pmatrix}$$

誤差逆伝播-行列表現 出力層 重み

$$\frac{\partial L}{\partial w_{ij}{}^K} = \delta_j{}^K \frac{\partial a_j{}^K}{\partial w_{ij}{}^K} = \delta_j{}^K x_i{}^{K-1}$$



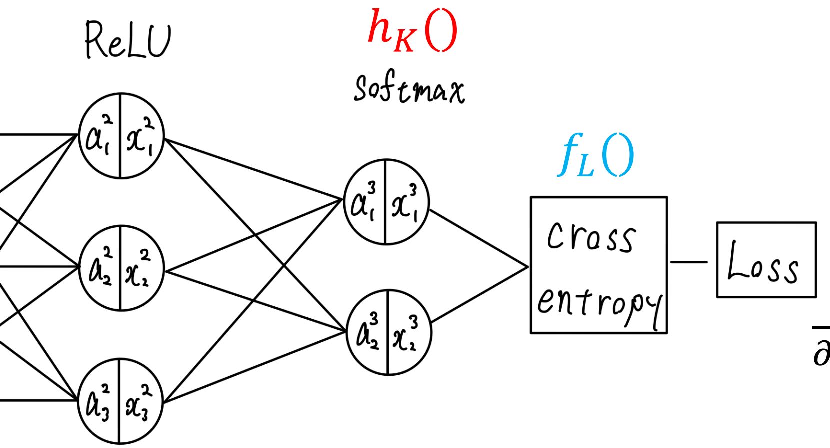
$$\frac{\partial L}{\partial W^K} = \begin{pmatrix} \delta_1{}^K \frac{\partial a_1{}^K}{\partial w_{11}{}^K} & \delta_2{}^K \frac{\partial a_2{}^K}{\partial w_{12}{}^K} \\ \delta_1{}^K \frac{\partial a_1{}^K}{\partial w_{21}{}^K} & \delta_2{}^K \frac{\partial a_2{}^K}{\partial w_{22}{}^K} \\ \delta_1{}^K \frac{\partial a_1{}^K}{\partial w_{31}{}^K} & \delta_2{}^K \frac{\partial a_2{}^K}{\partial w_{32}{}^K} \end{pmatrix} = \begin{pmatrix} \delta_1{}^K x_1{}^{K-1} & \delta_2{}^K x_1{}^{K-1} \\ \delta_1{}^K x_2{}^{K-1} & \delta_2{}^K x_2{}^{K-1} \\ \delta_1{}^K x_3{}^{K-1} & \delta_2{}^K x_3{}^{K-1} \end{pmatrix} = \begin{pmatrix} x_1{}^{K-1} \\ x_2{}^{K-1} \\ x_3{}^{K-1} \end{pmatrix} \begin{pmatrix} \delta_1{}^K & \delta_2{}^K \end{pmatrix}$$

$$= X^{2T} \Delta^3$$

$$\Delta^3 = \begin{pmatrix} \delta_1{}^K & \delta_2{}^K \end{pmatrix} = \left(\frac{\partial L}{\partial x_j{}^K} \frac{\partial x_j{}^K}{\partial a_j{}^K}\right)$$

誤差逆伝播-行列表現 出力層 重み

$$\frac{\partial L}{\partial {w_{ij}}^K} = {\delta_j}^K \frac{\partial {a_j}^K}{\partial {w_{ij}}^K} = {\delta_j}^K {x_i}^{K-1}$$



$$\frac{\partial L}{\partial W^K} = \begin{pmatrix} {\delta_1}^K \frac{\partial {a_1}^K}{\partial {w_{11}}^K} & {\delta_2}^K \frac{\partial {a_2}^K}{\partial {w_{12}}^K} \\ {\delta_1}^K \frac{\partial {a_1}^K}{\partial {w_{21}}^K} & {\delta_2}^K \frac{\partial {a_2}^K}{\partial {w_{22}}^K} \\ {\delta_1}^K \frac{\partial {a_1}^K}{\partial {w_{31}}^K} & {\delta_2}^K \frac{\partial {a_2}^K}{\partial {w_{32}}^K} \end{pmatrix} = \begin{pmatrix} {\delta_1}^K {x_1}^{K-1} & {\delta_2}^K {x_1}^{K-1} \\ {\delta_1}^K {x_2}^{K-1} & {\delta_2}^K {x_2}^{K-1} \\ {\delta_1}^K {x_3}^{K-1} & {\delta_2}^K {x_3}^{K-1} \end{pmatrix} = \begin{pmatrix} {x_1}^{K-1} \\ {x_2}^{K-1} \\ {x_3}^{K-1} \end{pmatrix} \begin{pmatrix} {\delta_1}^K & {\delta_2}^K \end{pmatrix}$$

$$= X^{2T} \Delta^3$$

$$\Delta^3 = \begin{pmatrix} {\delta_1}^K & {\delta_2}^K \end{pmatrix} = \begin{pmatrix} \frac{\partial f_L}{\partial {x_1}^K} \frac{\partial h_K}{\partial {a_1}^K} & \frac{\partial f_L}{\partial {x_2}^K} \frac{\partial h_K}{\partial {a_2}^K} \end{pmatrix}$$

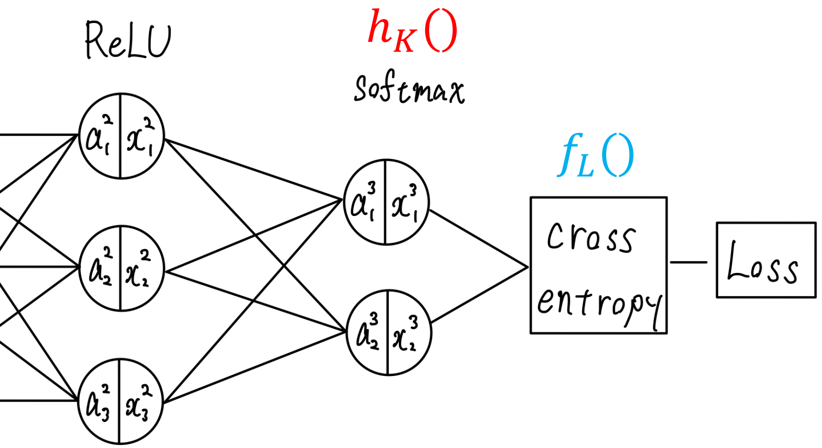
$$= \begin{pmatrix} \frac{\partial f_L}{\partial {x_1}^K} & \frac{\partial f_L}{\partial {x_2}^K} \end{pmatrix} \odot \begin{pmatrix} \frac{\partial h_K}{\partial {a_1}^K} & \frac{\partial h_K}{\partial {a_2}^K} \end{pmatrix}$$

$$= \frac{\partial f_L}{\partial X^K} \odot \frac{\partial h_K}{\partial A^K}$$

$${\delta_j}^K = \frac{\partial L}{\partial {x_j}^K} \frac{\partial {x_j}^K}{\partial {a_j}^K} = \frac{\partial f_L}{\partial {x_j}^K} \frac{\partial h_K}{\partial {a_j}^K}$$

$\odot$：アダマール積
サイズが等しい行列の各要素の積

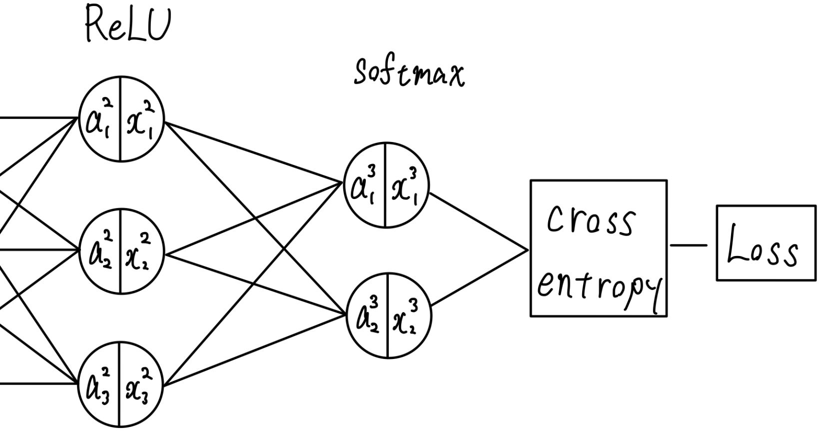
誤差逆伝播-行列表現 出力層 重み 公式まとめ



$$\frac{\partial L}{\partial W^K} = X^{(K-1)T} \Delta^K$$

$$\Delta^K = \frac{\partial f_L}{\partial X^K} \odot \frac{\partial h_K}{\partial A^K}$$

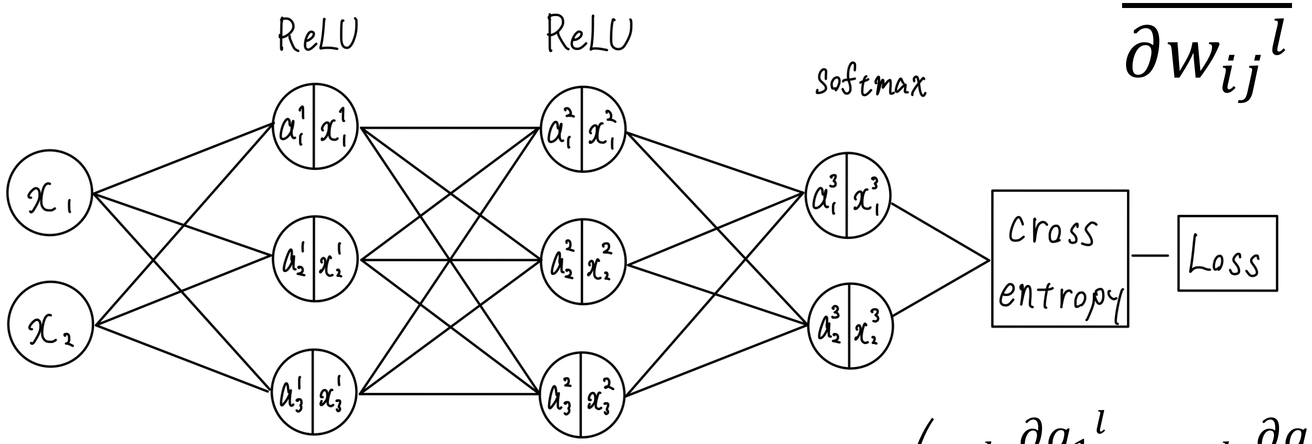
誤差逆伝播-行列表現 出力層 バイアス 公式



$$\frac{\partial L}{\partial {b_j}^K} = {\delta_j}^K \frac{\partial {a_j}^K}{\partial {b_j}^K} = {\delta_j}^K$$

$$\frac{\partial L}{\partial B^K} = \Delta^K$$

誤差逆伝播-行列表現 中間層 重み



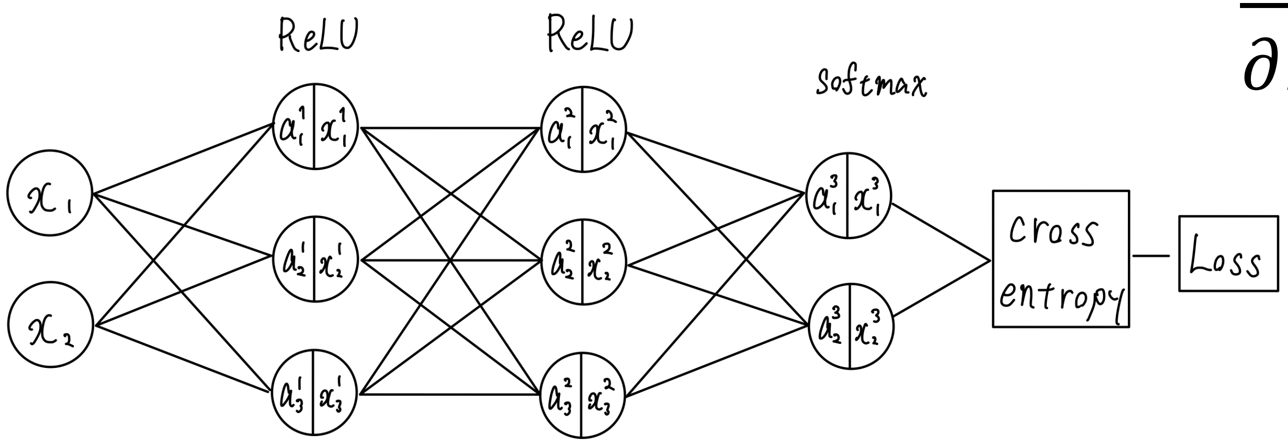
$$\frac{\partial L}{\partial {w_{ij}}^l} = {\delta_j}^l \frac{\partial {a_j}^l}{\partial {w_{ij}}^l} = {\delta_j}^l {x_i}^{l-1}$$

$$ {\delta_j}^l = \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{jk}}^{l+1} {h_l}'({a_j}^l)$$

$$\frac{\partial L}{\partial W^{l=2}} = \begin{pmatrix} {\delta_1}^l \frac{\partial {a_1}^l}{\partial {w_{11}}^l} & {\delta_2}^l \frac{\partial {a_2}^l}{\partial {w_{12}}^l} & {\delta_3}^l \frac{\partial {a_3}^l}{\partial {w_{13}}^l} \\ {\delta_1}^l \frac{\partial {a_1}^l}{\partial {w_{21}}^l} & {\delta_2}^l \frac{\partial {a_2}^l}{\partial {w_{22}}^l} & {\delta_3}^l \frac{\partial {a_3}^l}{\partial {w_{23}}^l} \\ {\delta_1}^l \frac{\partial {a_1}^l}{\partial {w_{31}}^l} & {\delta_2}^l \frac{\partial {a_2}^l}{\partial {w_{32}}^l} & {\delta_3}^l \frac{\partial {a_3}^l}{\partial {w_{33}}^l} \end{pmatrix} = \begin{pmatrix} {\delta_1}^l {x_1}^{l-1} & {\delta_2}^l {x_1}^{l-1} & {\delta_3}^l {x_1}^{l-1} \\ {\delta_1}^l {x_2}^{l-1} & {\delta_2}^l {x_2}^{l-1} & {\delta_3}^l {x_2}^{l-1} \\ {\delta_1}^l {x_3}^{l-1} & {\delta_2}^l {x_3}^{l-1} & {\delta_3}^l {x_3}^{l-1} \end{pmatrix}$$

$$= \begin{pmatrix} {x_1}^{l-1} \\ {x_2}^{l-1} \\ {x_3}^{l-1} \end{pmatrix} \begin{pmatrix} {\delta_1}^l & {\delta_2}^l & {\delta_3}^l \end{pmatrix} = X^{(l-1)T} \Delta^l$$

誤差逆伝播-行列表現 中間層 バイアス

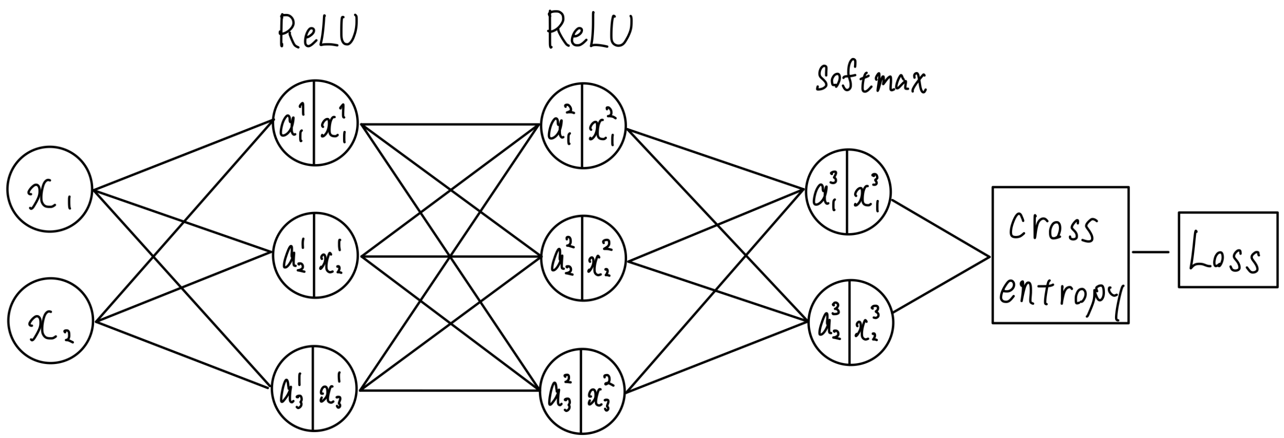


$$\frac{\partial L}{\partial b_j{}^l} = \delta_j{}^l \frac{\partial a_j{}^l}{\partial b_j{}^l} = \delta_j{}^l$$

$$\delta_j{}^l = \sum_{k=1}^{n} \delta_k{}^{l+1} w_{jk}{}^{l+1} h_l{}'(a_j{}^l)$$

$$\frac{\partial L}{\partial B^{l=2}} = \cancel{X^{(l-1)T}} \Delta^l = \Delta^l$$

誤差逆伝播-行列表現 中間層 誤差



$$\delta_j{}^l = \sum_{k=1}^{n} \frac{\partial L}{\partial a_k{}^{l+1}} \frac{\partial a_k{}^{l+1}}{\partial x_j{}^l} \frac{\partial x_j{}^l}{\partial a_j{}^l}$$

$$= \sum_{k=1}^{n} \delta_k{}^{l+1} w_{jk}{}^{l+1} h_l{}'(a_j{}^l)$$

$$\Delta^l = \begin{pmatrix} \delta_1{}^l & \delta_2{}^l & \delta_3{}^l \end{pmatrix} = ?$$

誤差逆伝播-行列表現 中間層 誤差

$$\Delta^l = \begin{pmatrix} {\delta_1}^l & {\delta_2}^l & {\delta_3}^l \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n=2} {\delta_k}^{l+1} {w_{1k}}^{l+1} h_l'({a_1}^l) & \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{2k}}^{l+1} h_l'({a_2}^l) & \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{3k}}^{l+1} h_l'({a_3}^l) \end{pmatrix}$$

$$\Delta^{l+1} = \begin{pmatrix} {\delta_1}^{l+1} & {\delta_2}^{l+1} \end{pmatrix} \qquad W^{l+1} = \begin{pmatrix} {w_{11}}^{l+1} & {w_{12}}^{l+1} \\ {w_{21}}^{l+1} & {w_{22}}^{l+1} \\ {w_{31}}^{l+1} & {w_{32}}^{l+1} \end{pmatrix} \qquad \frac{\partial h_l}{\partial A^l} = \begin{pmatrix} \frac{\partial h_l}{\partial {a_1}^K} & \frac{\partial h_l}{\partial {a_2}^K} & \frac{\partial h_l}{\partial {a_3}^K} \end{pmatrix}$$

を使ってどう表せるか

誤差逆伝播-行列表現 中間層 誤差

$$\Delta^l = \begin{pmatrix} {\delta_1}^l & {\delta_2}^l & {\delta_3}^l \end{pmatrix}$$

$$= \left(\sum_{k=1}^{n=2} {\delta_k}^{l+1} {w_{1k}}^{l+1} h_l{}'({a_1}^l) \quad \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{2k}}^{l+1} h_l{}'({a_2}^l) \quad \sum_{k=1}^{n} {\delta_k}^{l+1} {w_{3k}}^{l+1} h_l{}'({a_3}^l) \right)$$

$$\Delta^{l+1} = \begin{pmatrix} {\delta_1}^{l+1} & {\delta_2}^{l+1} \end{pmatrix} \qquad W^{l+1} = \begin{pmatrix} {w_{11}}^{l+1} & {w_{12}}^{l+1} \\ {w_{21}}^{l+1} & {w_{22}}^{l+1} \\ {w_{31}}^{l+1} & {w_{32}}^{l+1} \end{pmatrix} \qquad \frac{\partial h_l}{\partial A^l} = \begin{pmatrix} \frac{\partial h_l}{\partial {a_1}^l} & \frac{\partial h_l}{\partial {a_2}^l} & \frac{\partial h_l}{\partial {a_3}^l} \end{pmatrix}$$

$$\Delta^l = \Delta^{l+1} W^{(l+1)T} \odot \frac{\partial h_l}{\partial A^l}$$

$$= \begin{pmatrix} {\delta_1}^{l+1} & {\delta_2}^{l+1} \end{pmatrix} \begin{pmatrix} {w_{11}}^{l+1} & {w_{21}}^{l+1} & {w_{31}}^{l+1} \\ {w_{12}}^{l+1} & {w_{22}}^{l+1} & {w_{32}}^{l+1} \end{pmatrix} \odot \begin{pmatrix} \frac{\partial h_l}{\partial {a_1}^l} & \frac{\partial h_l}{\partial {a_2}^l} & \frac{\partial h_l}{\partial {a_3}^l} \end{pmatrix}$$

誤差逆伝播-公式 まとめ 行列表現

出力層

$$\frac{\partial L}{\partial W^K} = X^{(K-1)T}\Delta^K$$

$$\frac{\partial L}{\partial B^K} = \Delta^K$$

$$\Delta^K = \frac{\partial f_L}{\partial X^K} \odot \frac{\partial h_K}{\partial A^K}$$

中間層

$$\frac{\partial L}{\partial W^l} = X^{(l-1)T}\Delta^l$$

$$\frac{\partial L}{\partial B^l} = \Delta^l$$

$$\Delta^l = \Delta^{l+1}W^{(l+1)T} \odot \frac{\partial h_l}{\partial A^l}$$

誤差逆伝播-公式 まとめ 誤差/活性化関数の微分

誤差関数(cross-entropy)

$$\frac{\partial f_L}{\partial {x_j}^K} = \left(-\frac{t_j}{{x_j}^K} + \sum_{i=1 \wedge i \neq j}^{n} \frac{t_i}{{x_i}^K} \right)$$

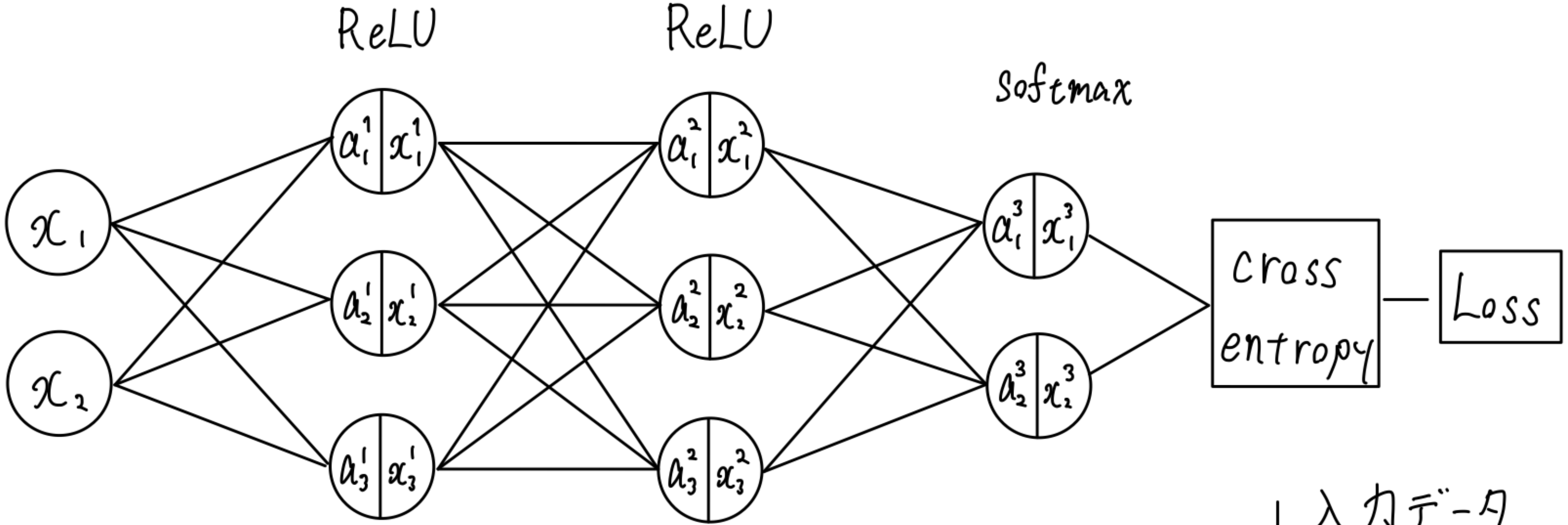
活性化関数(ReLU)

$$\frac{\partial h_l}{\partial {a_j}^l} = \begin{cases} 1({a_j}^l > 0) \\ 0({a_j}^l \leq 0) \end{cases}$$

活性化関数(soft-max)

$$\frac{\partial h_K}{\partial {a_j}^K} = {x_j}^K(1 - {x_j}^K)$$

誤差逆伝播-演習③手計算でNNを学習せよ



$$W^1 = \begin{pmatrix} 1 & -2 & 3 \\ -1 & 2 & -3 \end{pmatrix}, \quad W^2 = \begin{pmatrix} -1 & 2 & -3 \\ 1 & -2 & 3 \\ -1 & 2 & -3 \end{pmatrix}, \quad W^3 = \begin{pmatrix} -1 & 2 \\ -3 & 1 \\ -2 & 3 \end{pmatrix}$$

$$B^1 = \begin{pmatrix} 1 & -2 & 3 \end{pmatrix}, \quad B^2 = \begin{pmatrix} 1 & -2 & 3 \end{pmatrix}, \quad B^3 = \begin{pmatrix} 34 & -54 \end{pmatrix}$$

入力データ

$X = \begin{pmatrix} -1 & 2 \end{pmatrix}$

教師ラベル

$T = \begin{pmatrix} 1 & 0 \end{pmatrix}$

学習率

$\eta = 0.1$

左のNNを学習せよ

＝すべてのパラメータを1回更新せよ

Lossはいくら改善するか？

誤差逆伝播法
オンライン完

次回
誤差逆伝播法
ミニバッチor実装