

ディープラーニングの仕組みを知ろう！

第1回 人工知能勉強会 数学編

Shion MORISHITA

June 15, 2024

目次

はじめに

ニューラルネットワークの考え方

ニューロン

ニューロンの働きの数理的解釈

ユニット

ニューラルネットワーク

数学の基礎

ニューラルネットワークのための数学

数列と漸化式

Σ 記号

ベクトル

行列

はじめに

- ニューラルネットワークの構造を数理的に理解する
- ニューラルネットワークや、その学習方法について理解するために必要な数学の基礎を復習・理解する

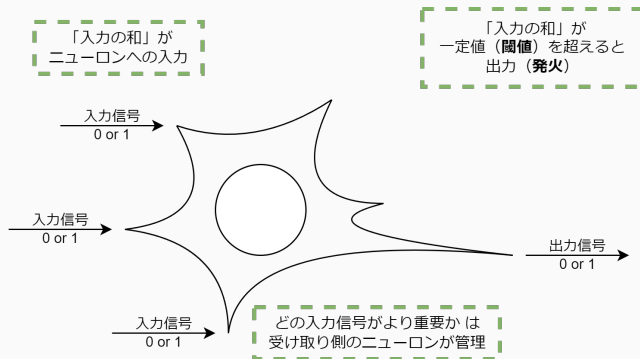
ニューラルネットワークの考え方

ニューラルネットワークの考え方

ニューロン

ニューロン = 神経細胞

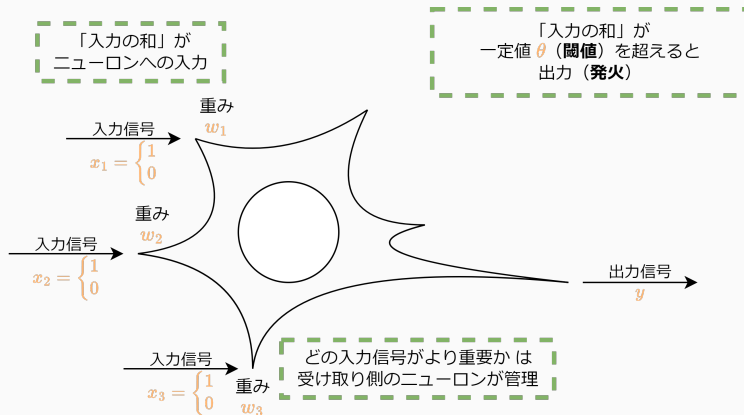
- 互いに結びついてネットワークを構築することで、
さまざまな処理を行なっている



ニューラルネットワークの考え方

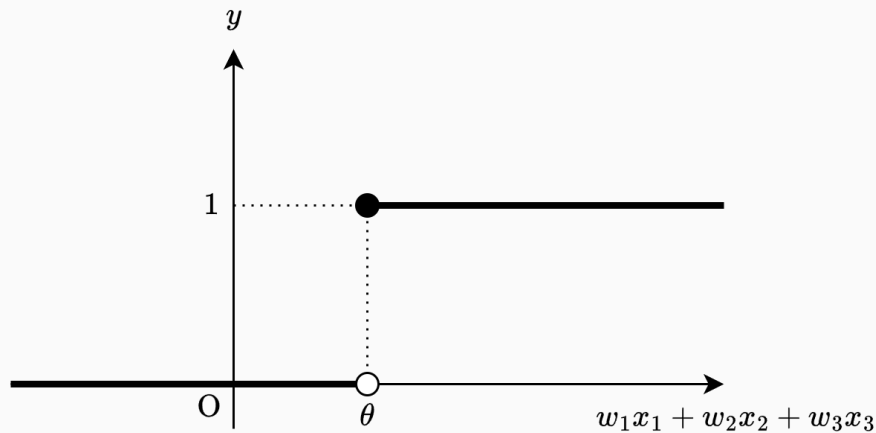
ニューロンの働きの数理的解釈

ニューロンの働きの数理解釈



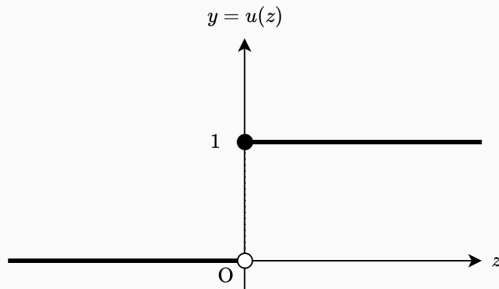
- 出力信号なし ($y = 0$) : $w_1x_1 + w_2x_2 + w_3x_3 < \theta$
- 出力信号あり ($y = 1$) : $w_1x_1 + w_2x_2 + w_3x_3 \geq \theta$

発火の条件のグラフ表現



発火の式

- 単位ステップ関数



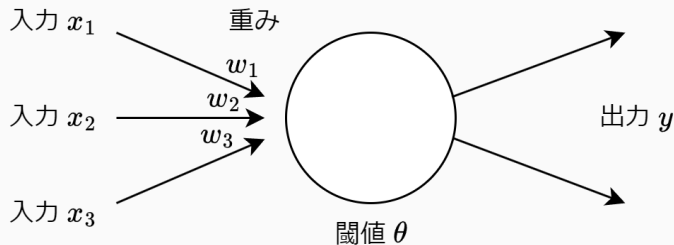
- 発火の式： $y = u(z) = u(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$
 - $z = w_1x_1 + w_2x_2 + w_3x_3 - \theta$ を、そのニューロンに対する重み付き入力という

ニューラルネットワークの考え方

ユニット

ユニット

- 簡略され抽象化されたニューロンを、生物学的なニューロンと区別して**ユニット** (unit) とよぶ

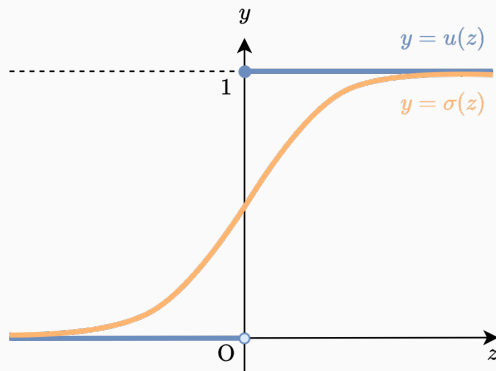


- 発火の式（旧）： $y = u(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$
 - 単位ステップ関数 u に限定する必要はない
- 発火の式（新）： $y = a(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$
 - 関数 a を活性化関数（activation function）という
 - この関数 a はモデル作成者がさまざまに定義可能

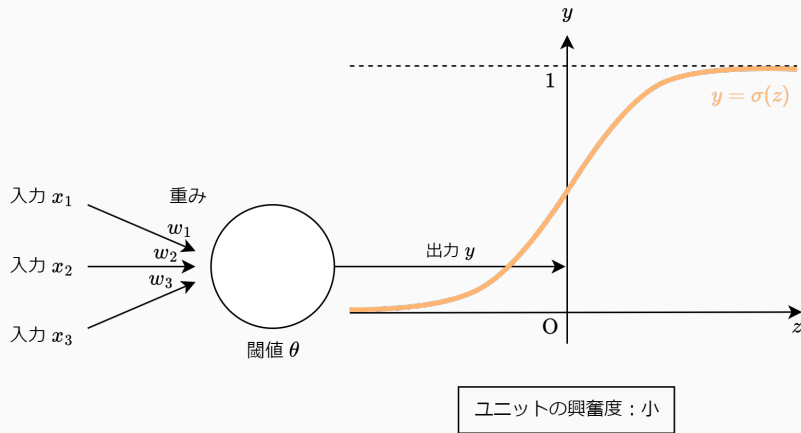
活性化関数の代表例

シグモイド関数 (Sigmoid function)

$$\sigma(z) \triangleq \frac{1}{1 + e^{-z}} \quad (e = 2.71828\cdots)$$

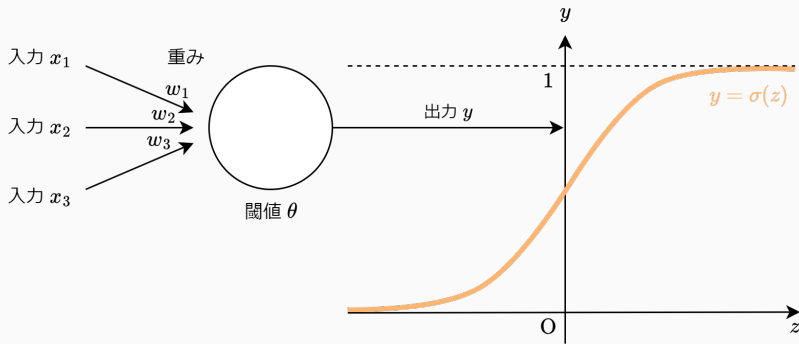


「発火の有無」から「興奮度」へ



$$y = \sigma(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$$

「発火の有無」から「興奮度」へ



ユニットの興奮度：大

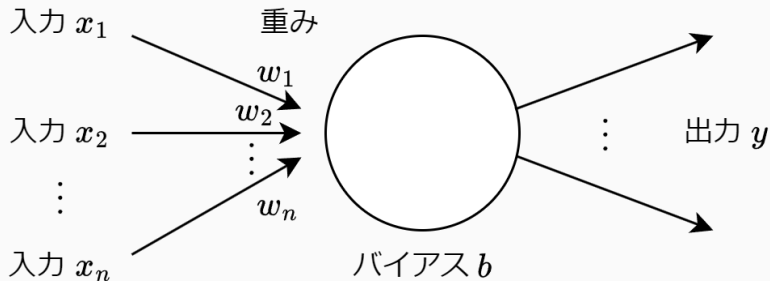
$$y = \sigma(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$$

$$y = a(w_1x_1 + w_2x_2 + w_3x_3 - \theta)$$

$$y = a(w_1x_1 + w_2x_2 + w_3x_3 + b)$$

- $-\theta \rightarrow +b$ に表記を変更
- すべて足し算に統一することで計算しやすくなる

ユニットのまとめ



重み付き入力： $z = w_1x_1 + w_2x_2 \cdots + w_nx_n + b$

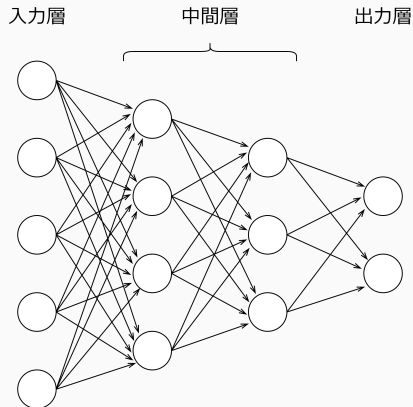
出力： $y = \sigma(z)$

ニューラルネットワークの考え方

ニューラルネットワーク

ニューラルネットワーク (Neural Network; NN)

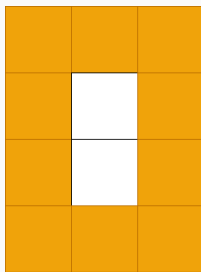
- ユニットをネットワーク状に結合したもの



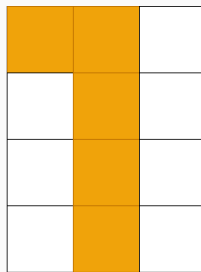
ニューラルネットワークを用いた問題の具体例

例題

4 × 3 画素からなる画像で読み取られた手書きの数字「0」「1」を識別するニューラルネットワークを作成せよ。ただし、学習データは 64 枚の画像とし、画素はモノクロ 2 階調とする。



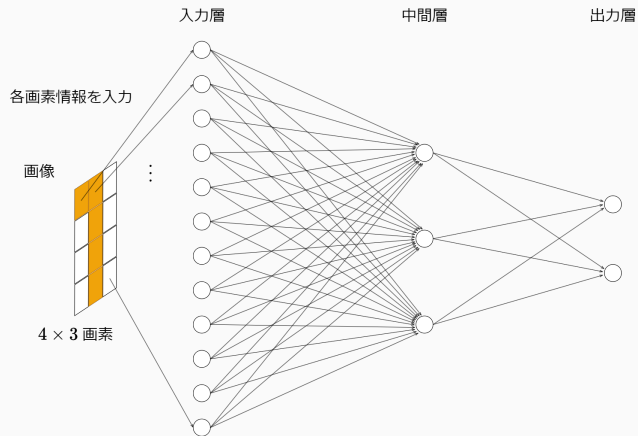
「0」の画像の例



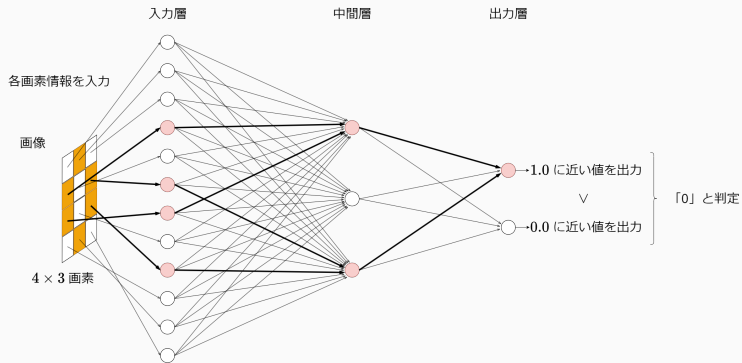
「1」の画像の例

ニューラルネットワークを用いた問題の具体例

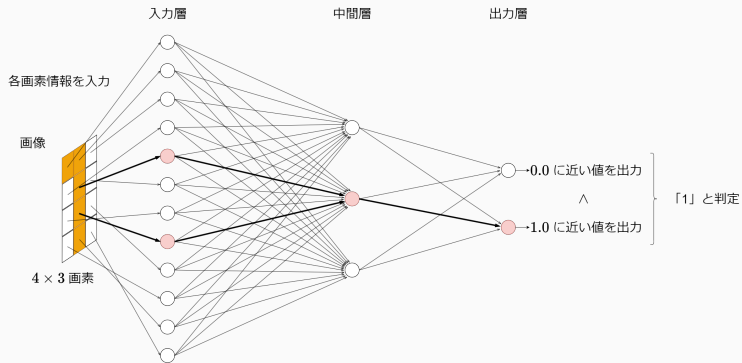
解答例



ニューラルネットワークの反応例



ニューラルネットワークの反応例



数学の基礎

数学の基礎

ニューラルネットワークのための数学

数列の意味

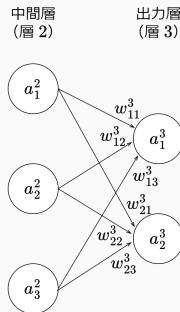
- **数列**：数の列
 - **項**：並べられたひとつひとつの数
 - **初項**：1 番目の項
 - **第 n 項**： n 番目の項
 - **有限数列**：有限個の項数を持つ数列
 - **末項**：有限数列において、数列の最後の項
 - **一般項**：数列の n 番目にある数で、通常 a_n などと表現する

ニューラルネットワークにおいて、ユニットの重み付き入力やその出力は数列と考えられる

例 a_j^l ： l 層 j 番目のユニットの出力値

数列と漸化式

- **漸化式**：初項 a_1 と、隣り合う 2 つの項 a_n, a_{n+1} の関係式
- **連立漸化式**：複数の数列がいくつかの関係式で結び付けられているもの
 - ニューラルネットワークにおいて、ユニットの入力と出力は連立漸化式で結ばれている



$$a_1^3 = \sigma(w_{11}^3 a_1^2 + w_{12}^3 a_2^2 + w_{13}^3 a_3^2 + b_1^3)$$
$$a_2^3 = \sigma(w_{21}^3 a_1^2 + w_{22}^3 a_2^2 + w_{23}^3 a_3^2 + b_2^3)$$

公式

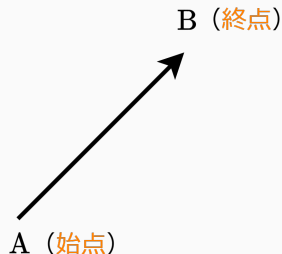
$$(I) \sum_{k=1}^n a_k = a_1 + a_2 + \cdots + a_{n-1} + a_n$$

$$(II) \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k$$

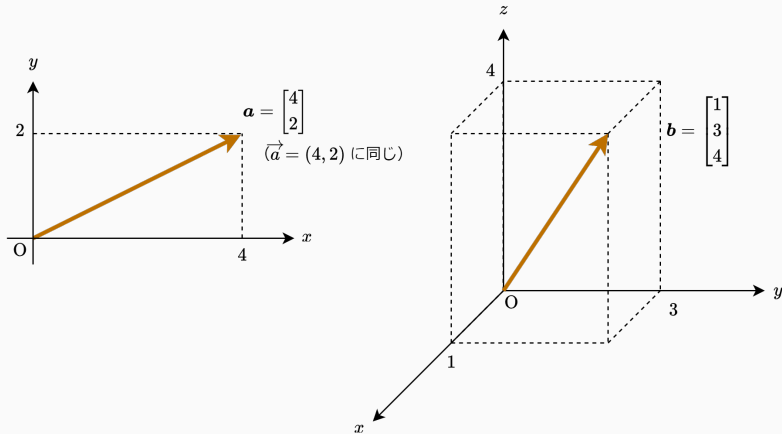
$$(III) \sum_{k=1}^n ca_k = c \sum_{k=1}^n a_k \quad (c: \text{定数})$$

有向線分とベクトル

- **有向線分**：向きを持つ線分
 - (始点の) **位置**
 - (線分の始点から終点への) **向き**
 - (線分の) **大きさ**
- **ベクトル**：有向線分から位置の属性を抜いたもの
 - (線分の始点から終点への) **向き**
 - (線分の) **大きさ**



ベクトルの成分表示



ベクトルの基本公式

ベクトル a の大きさ $\|a\|$

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \text{ のとき、 } \|a\| \triangleq \sqrt{a_1^2 + \cdots + a_n^2} = \sqrt{\sum_{k=1}^n a_k^2} \geq 0.$$

例 $a = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ のとき、 $\|a\| = \sqrt{(-1)^2 + 2^2} = 1 + 4 = 5.$

ベクトル a, b の内積 $\langle a, b \rangle$

- ベクトル a, b のなす角を θ とすると、 $\langle a, b \rangle \triangleq \|a\| \|b\| \cos \theta$.

- $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ とすると、 $\langle a, b \rangle \triangleq a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{k=1}^n a_k b_k$.

例 $a = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, b = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ のとき、 $\langle a, b \rangle = 2 \cdot 5 + 3 \cdot 1 = 13$.

コーシー・シュワルツの不等式

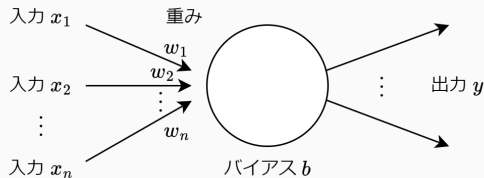
ベクトル a, b に対して、

$$-\|a\|\|b\| \leq \langle a, b \rangle \leq \|a\|\|b\|$$

が成り立つ。

コラム：ベクトルがどう役に立つ？

ニューラルネットワークの出力 y



$$y = \sigma(w_1 x_1 + \cdots + w_n x_n + b) = \sigma \left(\left\langle \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right\rangle + b \right) = \sigma(\langle \mathbf{w}, \mathbf{x} \rangle + b).$$

行列とは

数を縦と横に矩形上に配列したもの

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{bmatrix}$$

行

行

行

列

列

列

The diagram shows a 3x3 matrix A. The matrix is enclosed in large square brackets. The elements are arranged in three rows and three columns. To the right of the matrix, three horizontal orange arrows point from each row to the Chinese character '行' (row). Below the matrix, three vertical blue arrows point from each column to the Chinese character '列' (column).

3	1	4
1	5	9
2	6	5

$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ は 3 行 1 列の行列とみなせる

行列同士の演算

行列の和

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} a+p & b+q \\ c+r & d+s \end{bmatrix}$$

行列の積

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}$$

$$\begin{array}{cc} \textcircled{1} & \textcircled{2} \\ \textcircled{1} & \begin{bmatrix} a & b \end{bmatrix} \\ \textcircled{2} & \begin{bmatrix} c & d \end{bmatrix} \end{array} \begin{array}{cc} \textcircled{1} & \textcircled{2} \\ \begin{bmatrix} p \\ r \end{bmatrix} & \begin{bmatrix} q \\ s \end{bmatrix} \end{array} = \begin{bmatrix} \textcircled{1} \cdot \textcircled{1} & \textcircled{1} \cdot \textcircled{2} \\ \textcircled{2} \cdot \textcircled{1} & \textcircled{2} \cdot \textcircled{2} \end{bmatrix} \\ = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}$$

行列とベクトルの演算

$$\begin{array}{l} \textcircled{1} \\ \textcircled{2} \\ \textcircled{3} \end{array} \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{bmatrix} \begin{array}{c} \textcircled{1} \\ x \\ y \\ z \end{array} = \begin{bmatrix} \textcircled{1} \cdot \textcircled{1} \\ \textcircled{2} \cdot \textcircled{1} \\ \textcircled{3} \cdot \textcircled{1} \end{bmatrix}$$
$$= \begin{bmatrix} 3x + 1y + 4z \\ 1x + 5y + 9z \\ 2x + 6y + 5z \end{bmatrix}$$

数学の基礎

勾配降下法のための数学

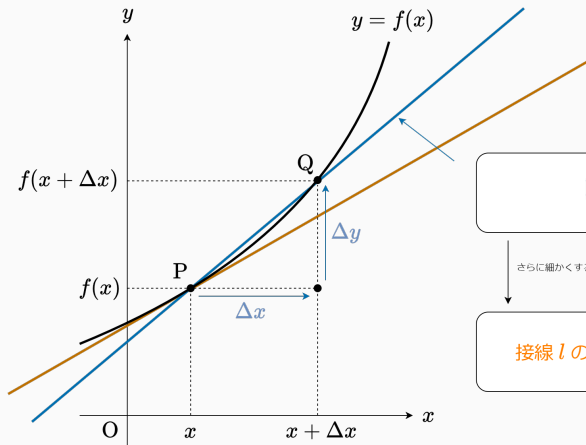
関数 $y = f(x)$ に対して導関数 $f'(x)$ は次のように定義される：

$$f'(x) \triangleq \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

※ 「 $\lim_{\Delta x \rightarrow 0} (\Delta x \text{ の式})$ 」は「数 Δx を限りなく 0 に近づけたとき、 $(\Delta x \text{ の式})$ の近づく値」を意味する

関数 $f(x)$ の導関数 $f'(x)$ を求めることを「関数 $f(x)$ を微分する」という

導関数の意味



$$\text{PQ の傾き } \frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

さらに細かくすると...

$$\text{接線 } l \text{ の傾き } f'(x) = \frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

関数の微分公式（抜粋）

x を変数、 c を定数とする。 e はネイピア数という定数で、 $e \simeq 2.71828\dots$

- $\frac{d(c)}{dx} = 0$
- $\frac{d(x)}{dx} = 1$
- $\frac{d(x^2)}{dx} = 2x$
- $\frac{d(e^x)}{dx} = e^x$
- $\frac{d(e^{-x})}{dx} = -e^{-x}$

- $\{f(x) + g(x)\}' = f'(x) + g'(x)$
- $\{cf(x)\}' = cf'(x)$ (c : 定数)

例

$C = (2 - y)^2$ (y : 変数) のとき、

$$\frac{dC}{dy} = \frac{d}{dy}C = \frac{d}{dy}(4 - 4y + y^2) = \frac{d(4)}{dy} - \frac{d(4y)}{dy} + \frac{d(y^2)}{dy} = 0 - 4 + 2y = -4 + 2y.$$

【重要】シグモイド関数 σ の微分公式

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (z: \text{変数}) \text{ のとき、}$$

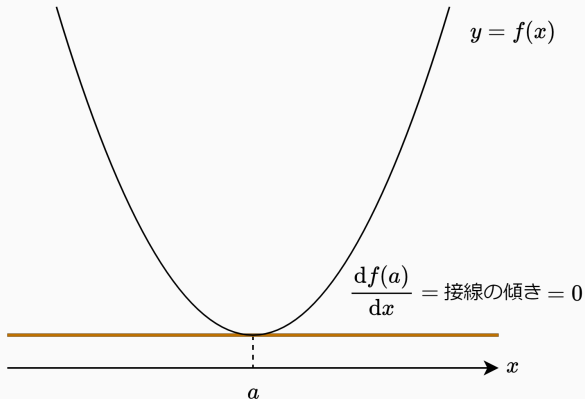
$$\frac{d\sigma(z)}{dz} = \sigma(z)(1 - \sigma(z))$$

ポイント

- 導関数の値が欲しいときに、微分計算をせずに済む！
 - つまり、計算機でも微分の値を容易に計算できる！

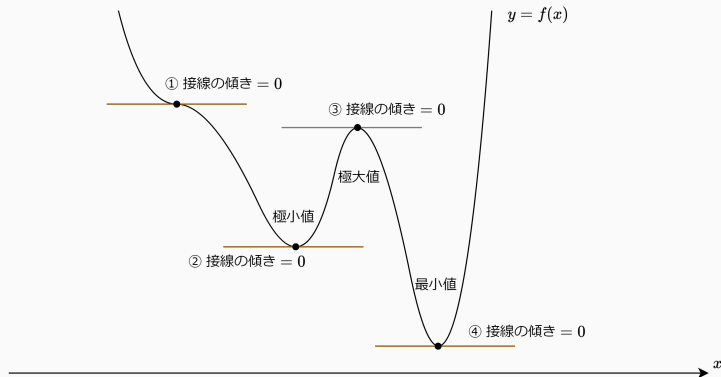
【重要】関数の最小値の条件

関数 $f(x)$ が $x = a$ で最小値をとるとき、 $\frac{df(a)}{dx} = 0$

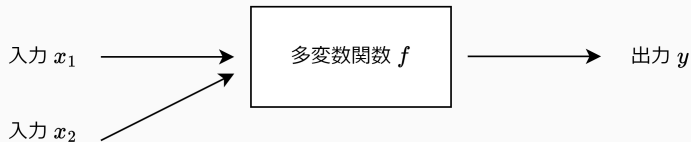
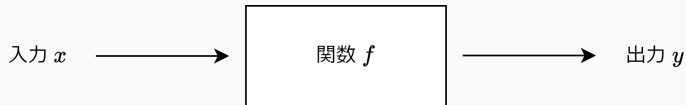


【重要】関数の最小値の条件

$f'(a) = 0$ は関数 $f(x)$ が $x = a$ で最小値になるための必要条件である。



入力が2つ以上ある関数を**多変数関数**という。



ある特定の変数について微分することを**偏微分** (partial derivative) という。

$y = f(x)$ を x について微分する場合：

- $\frac{df(x)}{dx}$

$z = f(x, y)$ を x について偏微分する場合：

- $\frac{\partial f(x, y)}{\partial x}$

例 $z = wx + b$ のとき、 $\frac{\partial z}{\partial x} = w$, $\frac{\partial z}{\partial w} = x$, $\frac{\partial z}{\partial b} = 1$

【重要】 多変数関数の最小条件

関数 $z = f(x, y)$ が最小になる必要条件は、 $\frac{\partial f}{\partial x} = 0$ かつ $\frac{\partial f}{\partial y} = 0$

ポイント

どの成分から見ても傾きが 0 なら、最小値の可能性あり！

関数の近似公式

1 変数関数の場合

$$f(x + \Delta x) \simeq f(x) + f'(x)\Delta x$$

2 変数関数の場合

$$f(x + \Delta x, y + \Delta y) \simeq f(x, y) + \frac{\partial f(x, y)}{\partial x} \Delta x + \frac{\partial f(x, y)}{\partial y} \Delta y$$

※ Δx や Δy は十分小さな数とする。

【重要】関数の近似公式 簡潔 ver.

$\Delta z \triangleq f(x_1 + \Delta x_1, x_2 + \Delta x_2) - f(x_1, x_2)$ とすると、

$$\Delta z \simeq \frac{\partial z}{\partial x_1} \Delta x_1 + \frac{\partial z}{\partial x_2} \Delta x_2.$$

$$\nabla z \triangleq \begin{bmatrix} \frac{\partial z}{\partial x_1} \\ \frac{\partial z}{\partial x_2} \end{bmatrix}, \Delta \mathbf{x} \triangleq \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} \text{ とすると、}$$

$$\Delta z \simeq \langle \nabla z, \Delta \mathbf{x} \rangle.$$

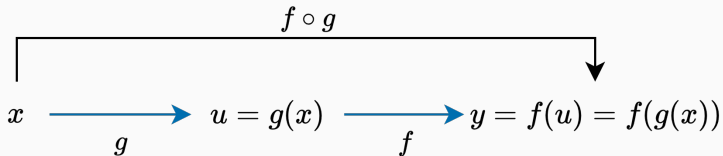
ポイント

関数の値の変化 Δz は、偏微分を集めたベクトル ∇z と、小さい値を集めたベクトル $\Delta \mathbf{x}$ で表される。

数学の基礎

誤差逆伝播法のための数学

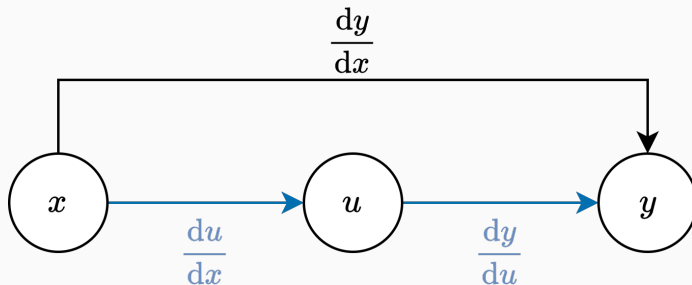
入れ子構造の関数 $f(g(x))$ を、関数 f と g の合成関数という。



1 変数関数の連鎖律

1 変数関数 $y = f(u)$ があり、その u が 1 変数関数 $u = g(x)$ と表されるとき、合成関数 $f(g(x))$ の導関数は次のように求められる：

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



1 変数関数の連鎖律の例

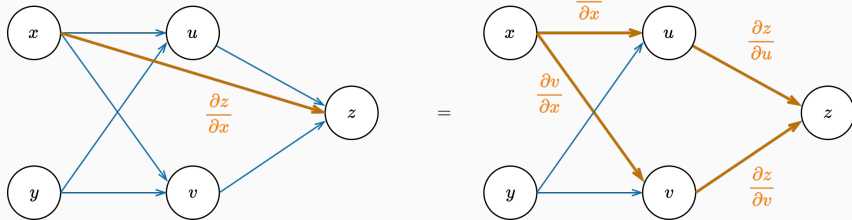
$y = u^2$, $u = 2 - x$ のとき、 y を x で微分すると、

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx} = \frac{d(u^2)}{du} \frac{d(2-x)}{dx} = 2u \cdot (-1) = -2u = -2(2-x) = -4 + 2x.$$

2 変数関数の連鎖律

2 変数関数 $z = f(u, v)$ があり、その u, v が 2 変数関数 $u = g_1(x, y), v = g_2(x, y)$ と表されるとき、

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



2 変数関数の連鎖律の例

$z = u^2 + v^2$, $u = ax + by$, $v = px + qy$ (a, b, p, q : 定数) のとき、

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial x} = 2u \cdot a + 2v \cdot p = 2a(ax + by) + 2p(px + qy).$$