微分

Python 中級
鈴木 敬彦

「微分」とは、変化の割合を測るための方法です。これまでに「微分」そのものを習った事がなかったとしても、恐らく、義務教育課程のどこかで目にしているのではないかと思います。

例えば、次の様な類の問題を見掛けたことはありませんか?

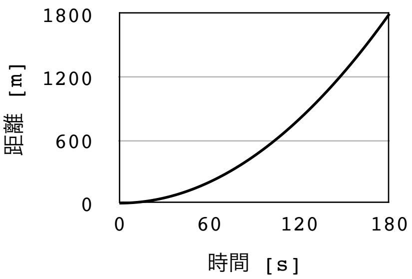
> 平地のまっすぐな道路に信号が二つあります。二つの信号間の距離は丁度 1800m です。
>
> 1. ある車が一定の速度で進み、最初の信号を通過してから次の信号を通過するまで 120 秒掛かりました。その車の速度は時速何 km/h ですか?
>
> 2. ある車が最初の信号で止まっています。青信号で走り始め、一定の割合で速度を増して 180 秒後に次の信号を通過しました。二つ目の信号を通過した時の車の速度は時速何 km/h ですか?

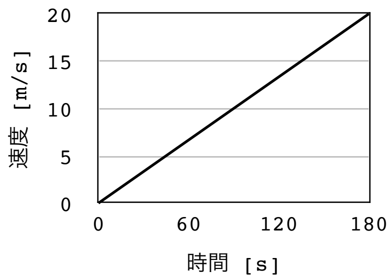
t:時間、d:距離、v:速度、v_0:初速度、a:加速度として、

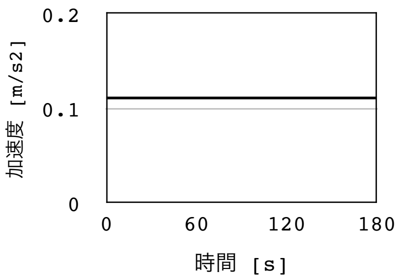
1. $d = vt = 120v = 1800\ [m]$
   $v = 1800/120\ [{}^{m}/_{s}] = 15\ [{}^{m}/_{s}] = 15 \times 3600 \div 1000\ [{}^{km}/_{h}] = 54\ [{}^{km}/_{h}]$
2. $d = v_0 t + \dfrac{1}{2}at^2 = 0 \cdot 180 + \dfrac{180^2}{2}a = 1800\ [m]$, $a = \dfrac{3600}{180^2}[{}^{m}/_{s^2}]$
   $v = v_0 + at = 0 + \dfrac{3600}{180^2}180 = 20\ [{}^{m}/_{s}] = 20 \times 3600 \div 1000\ [{}^{km}/_{h}] = 72\ [{}^{km}/_{h}]$

とても簡単な算数の問題ですが、この問題の中にも微分積分(微分と積分はとても深い関係にあります)の概念が入っています。加速度は速度が単位時間当たりに変化する割合で、速度は距離が単位時間当たりに変化する割合です。つまり、距離の変化を時間に関して微分したものが速度で、速度を時間に関して微分したものが加速度ということです。

$$d = v_0 t + \frac{1}{2}at^2 \xrightarrow{t\text{ で微分}} \dot{d} = v = v_0 + at \xrightarrow{t\text{ で微分}} \ddot{d} = \dot{v} = a$$







※ 逆に、加速度を時間で積分したものが速度で、速度を時間で積分したものが距離です。

定義

ある関数 $f(x)$ を変数 x に関して微分した関数 $f'(x)$ を $f(x)$ の導関数と呼び、導関数は以下の式で定義されます。導関数の変数に具体的な値を入力して得られる結果を微分係数と呼びます。

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

ここで lim は「極限」を意味し、lim の下の $h \to 0$ は h が限りなく 0 に近づく時（ $h = 0$ になるわけではないので注意）と言う意味です。例えば、次の様な簡単な例であればわかりやすいかと思います。

$$\lim_{x \to 0} x = 0 \qquad \lim_{x \to \infty} x = \infty \qquad \lim_{x \to 0} \frac{1}{x} = \infty \qquad \lim_{x \to \infty} \frac{1}{x} = 0$$

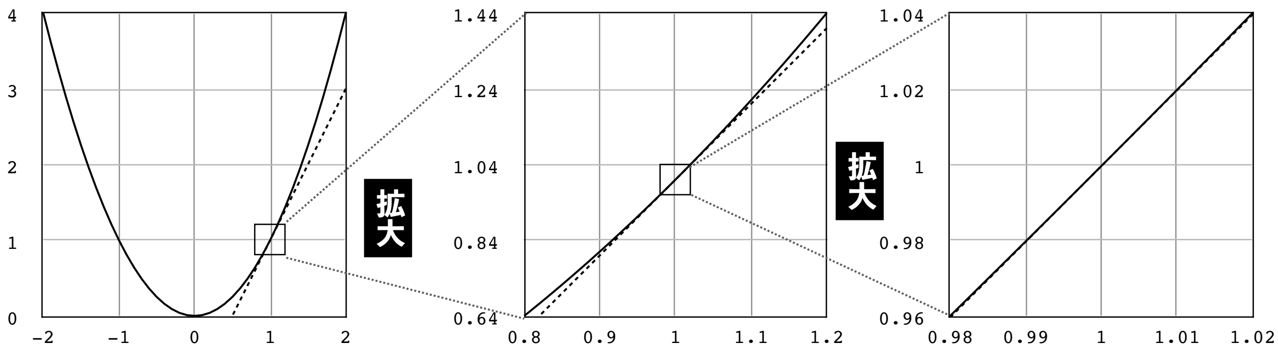
先程の式に戻ります。$f(x)$ は x の関数 $f(x)$ にある数 x を入力した時に得られる結果を、$f(x+h)$ は先程の数 x に限りなく 0 に近い数 h を加えた数 $x+h$ を入力した時に得られる結果を意味します。$f(x+h) - f(x)$ は、その差分を意味し、それを h で割るので、単に x が h 増加した時の $f(x)$ の変化の割合を表すものです。

前出の式を次の様に書き換えてみるとわかりやすいかもしれません。

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

例えば、関数が直線（1次関数）であれば、変化の割合は一定です。$y = f(x) = ax$ であれば、この直線上の任意の点付近での変化の割合は常に a で、これは初等数学では直線の傾きと呼ばれています。同様に関数が曲線であっても、その曲線上の任意の点付近での変化の割合は、その点で接する接線の傾きと等しくなります。

参考までに $y = f(x) = x^2$ を例として、(1,1) 近辺の変化の割合を見てみます。接線の傾きは 2 で (1,1) を通過するため、その直線の式は $y = 2x - 1$ となります。接線は点線で描画します。



全ての関数上の全ての点で微分できるわけではなく、微分できる場合は微分可能と言い、微分できない場合は微分不可能と言います。例えば、$y = |x|$ は $x = 0$ で微分不可能です。不連続であったり、尖点であったり、接線が垂直である様な点では微分不可能になります。

記法

微分（導関数）の記法にはいくつかあります。先程の $\ddot{d}$ や $\dot{v}$ も、 $f'(x)$ も微分を表しています。変数の上に ・ が付く記法は極初期の記法（ニュートン記法）で現在では滅多に使用されません。ちなみに 2 つの点 ¨ は、2 階微分と言って微分の微分（導関数の導関数）を意味します。現在、一般的に使用される記法はラグランジュ記法とライプニッツ記法です。$f'(x)$ はラグランジュ記法でダッシュの数が微分の階数を表します。例えば、元の関数が $y=f(x)=x^2$ であれば、その微分は、以下の様に記述されます。いずれも同じ意味で、分数形式のものがライプニッツ記法です。ライプニッツ記法は、あくまでも記法であって、分数ではないので間違えることのない様に注意してください。また、ライプニッツ記法を音読する場合は、分子から「でぃーわいでぃーえっくす」の様に読みます。

$$y' \qquad f'(x) \qquad \frac{dy}{dx} \qquad \frac{df}{dx} \qquad \frac{df(x)}{dx} \qquad \frac{d}{dx}f(x) \qquad \frac{dx^2}{dx} \qquad \frac{d}{dx}x^2$$

また、多変数関数の微分（偏微分、後述）はライプニッツ記法で記述される事が一般的です。例えば、元の関数を $z=f(x,y)=x^2+y^2$ として$\mathbf{x}$で偏微分する場合は次の様に書きます。

$$\frac{\partial z}{\partial x} \qquad \frac{\partial f}{\partial x} \qquad \frac{\partial f(x,y)}{\partial x} \qquad \frac{\partial}{\partial x}f(x,y) \qquad \frac{\partial(x^2+y^2)}{\partial x} \qquad \frac{\partial}{\partial x}(x^2+y^2)$$

導関数の求め方

ある関数の導関数は、定義に従って計算するしかないのですが、同じ関数であれば、誰が計算しても同じです。ここでは基本的な法則と多項式の導関数の求め方といくつか初等関数の微分公式を紹介するだけに留めます。その他の関数に関しては微分の公式集を参考にしてください。

基本的な法則

f 、g が共に微分可能な x の関数、a が x と関係のない数の時、線型性が成立します。

$$(af)' = af'$$

$$(f+g)' = f' + g'$$

多項式の導関数

多項式の微分にはわかりやすい法則性があるので、是非覚えておくべきでしょう。

$$y = f(x) = ax^2 + bx + c$$

$$\frac{dy}{dx} = \frac{d}{dx}ax^2 + \frac{d}{dx}bx + \frac{d}{dx}c = 2ax + b$$

元の関数の各項の次数が係数に掛かり、次数は 1 減じられます。1次の項は元の係数に1が掛かると共に次数が0となるため、定数となります。0 次の項（定数）は元の数に 0 が掛かるため、無くなります。

$$y = f(x) = ax^{\frac{3}{2}} + bx^{\frac{1}{2}} + c$$

$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}ax^{\frac{3}{2}} + \frac{d}{dx}bx^{\frac{1}{2}} + \frac{d}{dx}cx^0 = \frac{3}{2}ax^{\frac{3}{2}-1} + \frac{1}{2}bx^{\frac{1}{2}-1} + 0cx^{0-1} \\ &= \frac{3}{2}ax^{\frac{1}{2}} + \frac{1}{2}bx^{-\frac{1}{2}}\end{aligned}$$

初等関数の微分公式

$$\frac{d}{dx}x^a = ax^{a-1}$$

$$\frac{d}{dx}e^x = e^x \qquad \frac{d}{dx}a^x = (\log_e a)\ a^x$$

$$\frac{d}{dx}\log_e x = \frac{1}{x} \qquad \frac{d}{dx}\log_a x = \frac{1}{x\log_e a}$$

$$\frac{d}{dx}\cos x = -\sin x \qquad \frac{d}{dx}\sin x = \cos x \qquad \frac{d}{dx}\tan x = \frac{1}{\cos^2 x} = 1 + \tan^2 x$$

連鎖律 / Cahin Rule

複雑な関数の導関数を得るのに役立ちます。特にニューラルネットワークの理解には必須です。
次の関数の導関数を得ることを考えます。

$$y = f(x) = (x-a)^2$$

二乗なので、展開してしまえば簡単に計算できることでしょう。

$$y = f(x) = x^2 - 2ax + a^2$$

$$\frac{dy}{dx} = \frac{d}{dx}x^2 - \frac{d}{dx}2ax + \frac{d}{dx}a^2 = 2x - 2a = 2(x-a)$$

しかし、これが百乗であったり、あるいは簡単には展開できない式だったらどうでしょうか。そんな時に連鎖律が役に立ちます。g が x の関数で、f が g である時、次の関係が成り立ちます。

$$\begin{aligned}y &= f(u) \\ u &= g(x)\end{aligned}$$

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

先程の例を連鎖律で計算します。

$$y = (x-a)^2$$
$$y = f(u) = u^2\text{，}\ \ u = g(x) = x - a$$
$$\frac{dy}{du} = \frac{d}{dx}u^2 = 2u\text{，}\ \ \frac{du}{dx} = \frac{d}{dx}(x-a) = 1$$
$$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx} = 2u\cdot 1 = 2u = 2(x-a)$$

ここまでの例で連鎖は 2 つまででしたが、3 つ以上でも同様に計算可能です。

$$y = \cos\left((e^x - ax)^3\right)$$
$$y = \cos(u)\text{，}\ \ u = v^3\text{，}\ \ v = e^x - ax$$
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx} = -\sin(u)\cdot 3v^2\cdot(e^x - a)$$
$$= -3\sin\left((e^x - ax)^3\right)(e^x - ax)^2(e^x - a)$$

偏微分

多変数関数を特定の 1 つの変数で微分することを偏微分と呼びます。その他の変数は定数として扱われるため、記号が変わること以外は、これまでの方法と変わりはありません。

$$z = f(x,y) = x^2 + y^2$$
$$\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}\left(x^2 + y^2\right) = 2x$$

ベクトル、行列の微分

ベクトルや行列、あるいはテンソルを伴う関数でも、微分の基本は同じです。

ベクトル入力、スカラー出力

ベクトルの入力からスカラーを出力する、つまりは多変数関数（多変数スカラー値関数とも）なので前出の偏微分と大きくは変わりませんが、入力が x, y, z の様に独立した変数ではなく、$\boldsymbol{x} = (x_1\ x_2\ \cdots\ x_n)$ の様なベクトルとなり、x, y, z のどれか 1 つで微分するのではなく、$\boldsymbol{x}$ の各成分 $x_1, x_2, \cdots, x_n$ で微分します。

$$\boldsymbol{x} = (x_1 \quad x_2 \quad \cdots \quad x_n),\quad \boldsymbol{v} = (v_1 \quad v_2 \quad \cdots \quad v_n)^\intercal$$

$$y = f(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{v} = x_1v_1 + x_2v_2 + \cdots + x_nv_n$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \left(\frac{\partial y}{\partial x_1} \quad \frac{\partial y}{\partial x_2} \quad \cdots \quad \frac{\partial y}{\partial x_n}\right)$$

分母側のベクトルの形状に合わせます。詳細は後述。

$$\frac{\partial y}{\partial x_1} = v_1,\quad \frac{\partial y}{\partial x_2} = v_2,\quad \cdots \quad ,\quad \frac{\partial y}{\partial x_n} = v_n$$

$$\therefore\ \frac{\partial y}{\partial \boldsymbol{x}} = (v_1 \quad v_2 \quad \cdots \quad v_n) = \boldsymbol{v}^\intercal$$

一つ一つを見れば、これまでの微分と大差無い事がわかります。実際に多変数関数の微分の定義も大きくは変わりません。ただし、多変数スカラー値関数の微分では、一変数スカラー値関数の微分で言うところの微分係数を「**勾配（gradient）**」と呼びます。

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x})}{\|\boldsymbol{h}\|} \qquad f : \mathbb{R}^N \to \mathbb{R},\ \ \boldsymbol{x} \in \mathbb{R}^N$$

左側が定義で、右側が前提条件です。
前提条件から説明します。$\mathbb{R}$ は実数空間を意味し、右肩に乗った数が次元を表します。$\mathbb{R}^N$ は N 次元の実数空間を意味し、その空間のベクトルは N 個の実数値の組で表されます。$\to$ は写像を意味します。写像はある集合の元と別の集合の元とを結びつける対応の事で、関数、変換と捉えても良いでしょう。$f:$ のコロンは定義を意味していて、f をその後に続く記述で定義する、という意味です。つまり、$f : \mathbb{R}^N \to \mathbb{R}$ は、fを実 N 次元空間から実スカラーに変換する関数と定義する、という意味になります。その隣の $\in$ は 'in' の意味で、$\boldsymbol{x} \in \mathbb{R}^N$ は、$\boldsymbol{x}$ は実 N 次元空間に含まれるという意味になります。つまり、$\boldsymbol{x}$ は N 次元の実数からなるベクトルという意味になります。
次に左側の定義です。∇ はナブラと（アルテッドとも）呼ばれる演算子で、勾配を意味します。$\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ は関数 $f(\boldsymbol{x})$ の $\boldsymbol{x}$ に関する勾配という意味です。$\|\boldsymbol{h}\|$ はノルムと呼ばれ、ベクトル $\boldsymbol{h}$ の長さを意味します。
それぞれの記号の意味がわかれば、式の意味も理解できることと思います。

ベクトル入力、ベクトル出力

ベクトルの入力でベクトルを出力する関数は、多変数ベクトル値関数と呼ばれます。多変数ベクトル値関数であっても、微分の手順は変わりません。多変数ベクトル値関数の微分で微分係数に相当するテンソルをヤコビアン (Jacobian) と呼びます。導関数は以下の手順で計算します。

$$\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}, \quad \boldsymbol{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix}$$

$$\boldsymbol{y} = f(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{W} = \begin{pmatrix} x_1 w_{11} + x_2 w_{21} + \cdots + x_n w_{n1} \\ x_1 w_{12} + x_2 w_{22} + \cdots + x_n w_{n2} \\ \vdots \\ x_1 w_{1m} + x_2 w_{2m} + \cdots + x_n w_{nm} \end{pmatrix}^\intercal$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{y}}{\partial x_1} & \frac{\partial \boldsymbol{y}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{y}}{\partial x_n} \end{pmatrix}$$

前回と同様に分母側の形状に合わせます。

$$\frac{\partial \boldsymbol{y}}{\partial x_1} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \end{pmatrix}^\intercal, \quad \frac{\partial \boldsymbol{y}}{\partial x_2} = \begin{pmatrix} w_{21} & w_{22} & \cdots & w_{2m} \end{pmatrix}^\intercal,$$

$$\cdots, \quad \frac{\partial \boldsymbol{y}}{\partial x_n} = \begin{pmatrix} w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix}^\intercal$$

分子側は転置

$$\therefore \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \left(\begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{1m} \end{bmatrix} \begin{bmatrix} w_{21} \\ w_{22} \\ \vdots \\ w_{2m} \end{bmatrix} \cdots \begin{bmatrix} w_{n1} \\ w_{n2} \\ \vdots \\ w_{nm} \end{bmatrix} \right) = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix}^\intercal = \boldsymbol{W}^\intercal$$

最後の行列の中の行列は一般的な記述法ではないので注意してください。ただし、上記の様にテンソルの各成分がテンソルで、結果の階数が更に上がる様な場合には、このようなイメージを持つことは役に立ちます。その場合、成分テンソルは展開前の形状から鉛直方向に展開させます。

上記の例では、$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ を $\boldsymbol{x}$ の形状に合わせて行ベクトルで記述したので、$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ の成分 $\frac{\partial \boldsymbol{y}}{\partial x_n}$ は列方向に展開させています。

ここまで、多変数スカラー値関数の微分、多変数ベクトル値関数の微分の例では全て変数を $\boldsymbol{x}$ のみとしてきましたが、$\boldsymbol{v}$ や $\boldsymbol{W}$ が変数である場合もあります。参考までに、先程の例をそれぞれ $\boldsymbol{v}$ と $\boldsymbol{W}$ での微分例も掲載しておきます。

多変数スカラー値関数 $y = f(\boldsymbol{x}, \boldsymbol{v}) = \boldsymbol{x}\boldsymbol{v}$ を $\boldsymbol{v}$ で微分

$$\boldsymbol{x} = (x_1 \quad x_2 \quad \cdots \quad x_n), \quad \boldsymbol{v} = (v_1 \quad v_2 \quad \cdots \quad v_n)^\intercal$$

$$y = f(\boldsymbol{x}, \boldsymbol{v}) = \boldsymbol{x}\boldsymbol{v} = x_1v_1 + x_2v_2 + \cdots + x_nv_n$$

$$\frac{\partial y}{\partial \boldsymbol{v}} = \left(\frac{\partial y}{\partial v_1} \quad \frac{\partial y}{\partial v_2} \quad \cdots \quad \frac{\partial y}{\partial v_n}\right)^\intercal$$

$$\frac{\partial y}{\partial v_1} = x_1, \quad \frac{\partial y}{\partial v_2} = x_2, \quad \cdots \quad , \quad \frac{\partial y}{\partial v_n} = x_n$$

$$\therefore \ \frac{\partial y}{\partial \boldsymbol{v}} = (x_1 \quad x_2 \quad \cdots \quad x_n)^\intercal = \boldsymbol{x}^\intercal$$

多変数ベクトル値関数 $y = f(\boldsymbol{x}, \boldsymbol{W}) = \boldsymbol{x}\boldsymbol{W}$ を $\boldsymbol{W}$ で微分

$$\boldsymbol{x} = (x_1 \quad x_2 \quad \cdots \quad x_n), \quad \boldsymbol{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix}$$

$$\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{W}) = \boldsymbol{x}\boldsymbol{W} = \begin{pmatrix} x_1w_{11} + x_2w_{21} + \cdots + x_nw_{n1} \\ x_1w_{12} + x_2w_{22} + \cdots + x_nw_{n2} \\ \vdots \\ x_1w_{1m} + x_2w_{2m} + \cdots + x_nw_{nm} \end{pmatrix}^\intercal$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{W}} = \begin{pmatrix} \frac{\partial \boldsymbol{y}}{\partial w_{11}} & \frac{\partial \boldsymbol{y}}{\partial w_{12}} & \cdots & \frac{\partial \boldsymbol{y}}{\partial w_{1m}} \\ \frac{\partial \boldsymbol{y}}{\partial w_{21}} & \frac{\partial \boldsymbol{y}}{\partial w_{22}} & \cdots & \frac{\partial \boldsymbol{y}}{\partial w_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \boldsymbol{y}}{\partial w_{n1}} & \frac{\partial \boldsymbol{y}}{\partial w_{n2}} & \cdots & \frac{\partial \boldsymbol{y}}{\partial w_{nm}} \end{pmatrix}$$

$$\frac{\partial \boldsymbol{y}}{\partial w_{11}} = \left[\frac{\partial y_1}{\partial w_{11}} \quad \frac{\partial y_2}{\partial w_{11}} \quad \cdots \quad \frac{\partial y_m}{\partial w_{11}}\right]^\intercal = \left[\, x_1 \quad 0 \quad \cdots \quad 0 \,\right]^\intercal$$

$$\frac{\partial \boldsymbol{y}}{\partial w_{12}} = \left[\frac{\partial y_1}{\partial w_{12}} \quad \frac{\partial y_2}{\partial w_{12}} \quad \cdots \quad \frac{\partial y_m}{\partial w_{12}}\right]^\intercal = \left[\, 0 \quad x_1 \quad \cdots \quad 0 \,\right]^\intercal$$

$$\frac{\partial \boldsymbol{y}}{\partial w_{1m}} = \left[\frac{\partial y_1}{\partial w_{1m}} \quad \frac{\partial y_2}{\partial w_{1m}} \quad \cdots \quad \frac{\partial y_m}{\partial w_{1m}}\right]^\intercal = \left[\, 0 \quad 0 \quad \cdots \quad x_1 \,\right]^\intercal$$

$$\frac{\partial \boldsymbol{y}}{\partial w_{21}} = \left[\frac{\partial y_1}{\partial w_{21}} \quad \frac{\partial y_2}{\partial w_{21}} \quad \cdots \quad \frac{\partial y_m}{\partial w_{21}}\right]^\intercal = \left[\, x_2 \quad 0 \quad \cdots \quad 0 \,\right]^\intercal$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{y}}{\partial w_{22}} &= \begin{bmatrix} \frac{\partial y_1}{\partial w_{22}} & \frac{\partial y_2}{\partial w_{22}} & \cdots & \frac{\partial y_m}{\partial w_{22}} \end{bmatrix}^\intercal &= \begin{bmatrix} 0 & x_2 & \cdots & 0 \end{bmatrix}^\intercal \\
\frac{\partial \boldsymbol{y}}{\partial w_{2m}} &= \begin{bmatrix} \frac{\partial y_1}{\partial w_{2m}} & \frac{\partial y_2}{\partial w_{2m}} & \cdots & \frac{\partial y_m}{\partial w_{2m}} \end{bmatrix}^\intercal &= \begin{bmatrix} 0 & 0 & \cdots & x_2 \end{bmatrix}^\intercal
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{y}}{\partial w_{n1}} &= \begin{bmatrix} \frac{\partial y_1}{\partial w_{n1}} & \frac{\partial y_2}{\partial w_{n1}} & \cdots & \frac{\partial y_m}{\partial w_{n1}} \end{bmatrix}^\intercal &= \begin{bmatrix} x_n & 0 & \cdots & 0 \end{bmatrix}^\intercal \\
\frac{\partial \boldsymbol{y}}{\partial w_{n2}} &= \begin{bmatrix} \frac{\partial y_1}{\partial w_{n2}} & \frac{\partial y_2}{\partial w_{n2}} & \cdots & \frac{\partial y_m}{\partial w_{n2}} \end{bmatrix}^\intercal &= \begin{bmatrix} 0 & x_n & \cdots & 0 \end{bmatrix}^\intercal \\
\frac{\partial \boldsymbol{y}}{\partial w_{nm}} &= \begin{bmatrix} \frac{\partial y_1}{\partial w_{nm}} & \frac{\partial y_2}{\partial w_{nm}} & \cdots & \frac{\partial y_m}{\partial w_{nm}} \end{bmatrix}^\intercal &= \begin{bmatrix} 0 & 0 & \cdots & x_n \end{bmatrix}^\intercal
\end{aligned}
$$

$$
\therefore \ \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{W}} = \begin{pmatrix}
\begin{bmatrix} x_1 & 0 & \vdots & 0 \end{bmatrix}^\intercal & \begin{bmatrix} 0 & x_1 & \vdots & 0 \end{bmatrix}^\intercal & \cdots & \begin{bmatrix} 0 & 0 & \vdots & x_1 \end{bmatrix}^\intercal \\
\begin{bmatrix} x_2 & 0 & \vdots & 0 \end{bmatrix}^\intercal & \begin{bmatrix} 0 & x_2 & \vdots & 0 \end{bmatrix}^\intercal & \cdots & \begin{bmatrix} 0 & 0 & \vdots & x_2 \end{bmatrix}^\intercal \\
\vdots & \vdots & \ddots & \vdots \\
\begin{bmatrix} x_n & 0 & \vdots & 0 \end{bmatrix}^\intercal & \begin{bmatrix} 0 & x_n & \vdots & 0 \end{bmatrix}^\intercal & \cdots & \begin{bmatrix} 0 & 0 & \vdots & x_n \end{bmatrix}^\intercal
\end{pmatrix}
$$

結果は3階のテンソルとなります。（この書き方は、一般的な記述の仕方ではありません。）
少々複雑になってきたかもしれませんが、ベクトル入力のニューラルネットワークでは、上記のような 3 階のテンソルも、一時的には出てきても、最終的には消えるので安心してください。

クロネッカー積を使うと、もう少し簡単に書けますが、

$$
\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{W}} = \begin{pmatrix} x_1 \boldsymbol{I}_m & x_2 \boldsymbol{I}_m & \cdots & x_n \boldsymbol{I}_m \end{pmatrix}^\intercal = \boldsymbol{x}^\intercal \otimes \boldsymbol{I}_m
$$

上記の $\frac{\partial y}{\partial \boldsymbol{x}}$ は、より一般的には、以下のような添字を用いた書き方が多いでしょう。δ_j^k はクロネッカーのデルタと呼ばれる関数で 2 つの添字が等しい時だけ 1 を、それ以外は 0 となる関数です。

$$
\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{W}} \right)_{ijk} = \frac{\partial \boldsymbol{y}_k}{\partial \boldsymbol{W}_{ij}} = \delta_j^k x_i
$$

連鎖律

一変数スカラー値関数の微分で紹介した連鎖律は、多変数関数の微分でも成立します。

連鎖律を用いない計算例

$$y = \boldsymbol{x}\boldsymbol{W}\boldsymbol{v}$$

$$\boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix},\quad \boldsymbol{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix},\quad \boldsymbol{v} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}^\intercal$$

$$y = \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix} \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} x_1w_{11} + x_2w_{21} + \cdots + x_mw_{m1} \\ x_1w_{12} + x_2w_{22} + \cdots + x_mw_{m2} \\ \vdots \\ x_1w_{1n} + x_2w_{2n} + \cdots + x_mw_{mn} \end{pmatrix}^\intercal \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

$$\begin{aligned} = (x_1w_{11} + x_2w_{21} + \cdots + x_mw_{m1})\, v_1 &+ (x_1w_{12} + x_2w_{22} + \cdots + x_mw_{m2})\, v_2 \\ + \quad \cdots \quad &+ (x_1w_{1n} + x_2w_{2n} + \cdots + x_mw_{mn})\, v_n \end{aligned}$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_m} \end{pmatrix} = \begin{pmatrix} w_{11}v_1 + w_{12}v_2 + \cdots + w_{1n}v_n \\ w_{21}v_1 + w_{22}v_2 + \cdots + w_{2n}v_n \\ \vdots \\ w_{m1}v_1 + w_{m2}v_2 + \cdots + w_{mn}v_n \end{pmatrix}^\intercal$$

$$= \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{m1} \\ w_{12} & w_{22} & \cdots & w_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1n} & w_{2n} & \cdots & w_{mn} \end{pmatrix} = \boldsymbol{v}^\intercal \boldsymbol{W}^\intercal = (\boldsymbol{W}\boldsymbol{v})^\intercal$$

連鎖律を用いた計算例

$$y = \boldsymbol{x}\boldsymbol{W}\boldsymbol{v} = \boldsymbol{z}\boldsymbol{v},\quad \boldsymbol{z} = \boldsymbol{x}\boldsymbol{W}$$

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

$$\boldsymbol{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} = \boldsymbol{x}\boldsymbol{W} = \begin{pmatrix} x_1w_{11} + x_2w_{21} + \cdots + x_mw_{m1} \\ x_1w_{12} + x_2w_{22} + \cdots + x_mw_{m2} \\ \vdots \\ x_1w_{1n} + x_2w_{2n} + \cdots + x_mw_{mn} \end{pmatrix}^\intercal$$

$$y = \boldsymbol{z}\boldsymbol{v} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}^\intercal = z_1v_1 + z_2v_2 + \cdots + z_nv_n$$

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{z}}{\partial x_1} & \frac{\partial \boldsymbol{z}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{z}}{\partial x_m} \end{pmatrix} = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{m1} \\ w_{12} & w_{22} & \cdots & w_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1n} & w_{2n} & \cdots & w_{mn} \end{pmatrix} = \boldsymbol{W}^\intercal$$

$$\frac{\partial y}{\partial \boldsymbol{z}} = \begin{pmatrix} \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2} & \cdots & \frac{\partial y}{\partial z_n} \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \boldsymbol{v}^\intercal$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial y}{\partial \boldsymbol{z}} \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \boldsymbol{v}^\intercal \boldsymbol{W}^\intercal = (\boldsymbol{W}\boldsymbol{v})^\intercal$$

上記の例で多変数関数の微分でも連鎖律が成立していることがわかるでしょう。
次は、先程の例（p.8）で 3 階テンソルが出てきたことを踏まえて、テンソルの計算や縮約を用いずに基本的な微分と行列演算だけで計算すべく、一手間加えた計算例です。
先ずは、前頁の「連鎖律を用いた計算例」を「一手間」加えた方法で計算します。

$$y = \boldsymbol{x}\boldsymbol{W}\boldsymbol{v} = z\boldsymbol{v},\quad z = \boldsymbol{x}\boldsymbol{W}$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial y}{\partial z}\frac{\partial z}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial z}\frac{\partial z}{\partial x_1} & \frac{\partial y}{\partial z}\frac{\partial z}{\partial x_2} & \cdots & \frac{\partial y}{\partial z}\frac{\partial z}{\partial x_m} \end{pmatrix}$$

$$\frac{\partial y}{\partial z} = (v_1 \quad v_2 \quad \cdots \quad v_n) \qquad\qquad \frac{\partial z}{\partial x_1} = (w_{11} \quad w_{12} \quad \cdots \quad w_{1n})^\intercal$$

$$\frac{\partial y}{\partial z}\frac{\partial z}{\partial x_1} = (v_1 \quad v_2 \quad \cdots \quad v_n)(w_{11} \quad w_{12} \quad \cdots \quad w_{1n})^\intercal = v_1 w_{11} + v_2 w_{12} + \cdots + v_n w_{1n}$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{pmatrix} v_1 w_{11} + v_2 w_{12} + \cdots + v_n w_{1n} \\ v_1 w_{21} + v_2 w_{22} + \cdots + v_n w_{2n} \\ \vdots \\ v_1 w_{m1} + v_2 w_{m2} + \cdots + v_n w_{mn} \end{pmatrix}^\intercal = (v_1 \quad v_2 \quad \cdots \quad v_n)\begin{pmatrix} w_{11} & w_{21} & \cdots & w_{m1} \\ w_{12} & w_{22} & \cdots & w_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1n} & w_{2n} & \cdots & w_{mn} \end{pmatrix}$$

$$= \boldsymbol{v}^\intercal \boldsymbol{W}^\intercal = (\boldsymbol{W}\boldsymbol{v})^\intercal$$

同じ結果となることが確認できました。次は、p.8　の例と同様に行列での微分が出てくる場合です。ベクトルの行列での微分の結果が　3　階テンソルになることは既に確認した通りですが、スカラーの行列での微分は行列です。「一手間」加えていない方法では正しい結果に辿り着けません。

$$\frac{\partial y}{\partial \boldsymbol{W}} = \frac{\partial y}{\partial z}\frac{\partial z}{\partial \boldsymbol{W}}$$

$$\frac{\partial z}{\partial \boldsymbol{W}} = \begin{pmatrix} [x_1 \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_1 \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_1]^\intercal \\ [x_2 \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_2 \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_2]^\intercal \\ \vdots & \vdots & \ddots & \vdots \\ [x_m \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_m \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_m]^\intercal \end{pmatrix}$$

$$\frac{\partial y}{\partial \boldsymbol{W}} = (v_1 \quad v_2 \quad \cdots \quad v_n)\begin{pmatrix} [x_1 \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_1 \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_1]^\intercal \\ [x_2 \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_2 \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_2]^\intercal \\ \vdots & \vdots & \ddots & \vdots \\ [x_m \quad 0 \quad \vdots \quad 0]^\intercal & [0 \quad x_m \quad \vdots \quad 0]^\intercal & \cdots & [0 \quad 0 \quad \vdots \quad x_m]^\intercal \end{pmatrix}$$

これを今までの行列演算の枠組みで計算しようとするのは無理がありますが、「一手間」加えた方法であれば、今までの行列演算の枠組みで計算可能です。

$$\frac{\partial y}{\partial \boldsymbol{W}} = \frac{\partial y}{\partial z}\frac{\partial z}{\partial \boldsymbol{W}} = \begin{pmatrix} \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{11}} & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{12}} & \cdots & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{1n}} \\ \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{21}} & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{22}} & \cdots & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{m1}} & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{m2}} & \cdots & \frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{mn}} \end{pmatrix}$$

$$\frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{11}} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} x_1 & 0 & \cdots & 0 \end{pmatrix}^\intercal = v_1 x_1$$

$$\frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{12}} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} 0 & x_1 & \cdots & 0 \end{pmatrix}^\intercal = v_2 x_1$$

$$\frac{\partial y}{\partial z}\frac{\partial z}{\partial w_{1n}} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & x_1 \end{pmatrix}^\intercal = v_n x_1 \qquad \cdots$$

$$\frac{\partial y}{\partial \boldsymbol{W}} = \begin{pmatrix} v_1 x_1 & v_2 x_1 & \cdots & v_n x_1 \\ v_1 x_2 & v_2 x_2 & \cdots & v_n x_2 \\ \vdots & \vdots & \ddots & \vdots \\ v_1 x_m & v_2 x_m & \cdots & v_n x_m \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \boldsymbol{x}^\intercal \boldsymbol{v}^\intercal = (\boldsymbol{v}\boldsymbol{x})^\intercal$$

Numerator Layout / Denominator Layout

日本ではあまり気にされないようですが、とても重要なポイントです。訳語は「分子レイアウト」、「分母レイアウト」でも良いのですが、一般的な訳語がないため標題は訳していません。これまでのベクトル、行列の微分計算例では、常に分母側の形状に合わせて計算を進めてきました。これは Denominator Layout での表記です。Numerator Layout は分子側の形状に合わせます。二つの表記法に特に優劣はなく、どちらを使ったのでも構わないのですが、どちらに合わせたのかは把握しておくべきです。Denominator Layout では、分子側の形状は転置されます。計算例で「分子側の形状は転置」としたのはそのためです。Numerator Layout の場合はその逆で、分母側の形状が転置されます。Denominator Layout と Numerator Layout では結果の形状が異なるため、その後の計算にも違いが出てきます。

Numerator Layout での計算例

$$y = \boldsymbol{x}\boldsymbol{W}\boldsymbol{v} = \boldsymbol{z}\boldsymbol{v},\quad \boldsymbol{z} = \boldsymbol{x}\boldsymbol{W}$$

$$\boldsymbol{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} = \boldsymbol{x}\boldsymbol{W} = \begin{pmatrix} x_1 w_{11} + x_2 w_{21} + \cdots + x_m w_{m1} \\ x_1 w_{12} + x_2 w_{22} + \cdots + x_m w_{m2} \\ \vdots \\ x_1 w_{1n} + x_2 w_{2n} + \cdots + x_m w_{mn} \end{pmatrix}^\mathsf{T}$$

$$y = \boldsymbol{z}\boldsymbol{v} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}^\mathsf{T} = z_1 v_1 + z_2 v_2 + \cdots + z_n v_n$$

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{z}}{\partial x_1} & \frac{\partial \boldsymbol{z}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{z}}{\partial x_m} \end{pmatrix}^\mathsf{T} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix} = \boldsymbol{W}$$

$$\frac{\partial y}{\partial \boldsymbol{z}} = \begin{pmatrix} \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2} & \cdots & \frac{\partial y}{\partial z_n} \end{pmatrix}^\mathsf{T} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}^\mathsf{T} = \boldsymbol{v}$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial y}{\partial \boldsymbol{z}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \boldsymbol{v}\boldsymbol{W} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix}$$

WRONG

Numerator Layout では、連鎖律の順番も逆になります。

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\frac{\partial y}{\partial \boldsymbol{z}} = \boldsymbol{W}\boldsymbol{v} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$