

微分入門 (2)

前田利之（まえだ としゆき）
maechan@hannan-u.ac.jp

阪南大学 経営情報学部

(2024.10.23)

なぜ微分を勉強するか？（再掲）

☆… だから

- **回帰分析**（最小二乗法）
- **シグモイド** 関数：ニューラルネットの重要な要素
- **ハイパボリック・タンジェント** 関数：ニューラルネット、特に LSTM 等で重要な要素

なぜ微分を勉強するか？（再掲）

☆… **必要**だから

- **回帰分析**（最小二乗法）
- **シグモイド** 関数：ニューラルネットの重要な要素
- **ハイパボリック・タンジェント** 関数：ニューラルネット、特に LSTM 等で重要な要素

そもそも「微分」って何？（復習）

- ある瞬間の
- 関数（のグラフ）の場合は、各点における傾き =
- のところが
（最大／最小の可能性）

そもそも「微分」って何？（復習）

- ある瞬間の **変化率**
- 関数（のグラフ）の場合は、各点における傾き =
- のところが
（最大／最小の可能性）

そもそも「微分」って何？（復習）

- ある瞬間の **変化率**
- 関数（のグラフ）の場合は、各点における傾き = **変化の割合**
- のところが
（最大／最小の可能性）

そもそも「微分」って何？（復習）

- ある瞬間の **変化率**
- 関数（のグラフ）の場合は、各点における傾き = **変化の割合**
- **微分係数が0** のところが
（最大／最小の可能性）

そもそも「微分」って何？（復習）

- ある瞬間の **変化率**
- 関数（のグラフ）の場合は、各点における傾き = **変化の割合**
- **微分係数が0** のところが **極大／極小**
（最大／最小の可能性）

そもそも「微分」って何？（復習その2）

性質【公式】

- $\{a \cdot f(x)\}' = a \cdot f(x)'$

... 関数の定数倍は、微分の外に出せる

- $\{f(x) + g(x)\}' = f(x)' + g(x)'$

... 多項式の場合は各項を微分すればいい

- $(x^n)' = nx^{n-1}$... 整式の単項の微分（基本中の基本）

合成関数の微分

- 合成関数とは：2つ以上の関数を合成したもの（そのまま）
- 例えば $f(x) = (x^3 + 1)^2$ の場合、

- とおくと、

- と書ける→

合成関数の微分

- 合成関数とは：2つ以上の関数を合成したもの（そのまま）
- 例えば $f(x) = (x^3 + 1)^2$ の場合、
 - $g(x) = x^2, h(x) = x^3 + 1$ とおくと、
 - と書ける→

合成関数の微分

- 合成関数とは：2つ以上の関数を合成したもの（そのまま）
- 例えば $f(x) = (x^3 + 1)^2$ の場合、
 - $g(x) = x^2, h(x) = x^3 + 1$ とおくと、
 - $f(x) = g(h(x))$ と書ける→

合成関数の微分

- 合成関数とは：2つ以上の関数を合成したもの（そのまま）
- 例えば $f(x) = (x^3 + 1)^2$ の場合、
 - $g(x) = x^2, h(x) = x^3 + 1$ とおくと、
 - $f(x) = g(h(x))$ と書ける → g, h の合成

合成関数の微分（続き）

- $f(x) = g(h(x))$ の微分は、

【重要公式】

... 残念ながら証明略, ちなみに $(f(x))'$ は簡単に $f'(x)$ と書ける (以下同様)

- 前の例 $f(x) = (x^3 + 1)^2$ だと、

なので、

と

なる

合成関数の微分（続き）

- $f(x) = g(h(x))$ の微分は、

$$f'(x) = g'(h(x))h'(x) \quad \text{【重要公式】}$$

... 残念ながら証明略, ちなみに $(f(x))'$ は簡単に $f'(x)$ と書ける (以下同様)

- 前の例 $f(x) = (x^3 + 1)^2$ だと、

なので、

と

なる

合成関数の微分（続き）

- $f(x) = g(h(x))$ の微分は、

$$f'(x) = g'(h(x))h'(x) \quad \text{【重要公式】}$$

... 残念ながら証明略, ちなみに $(f(x))'$ は簡単に $f'(x)$ と書ける (以下同様)

- 前の例 $f(x) = (x^3 + 1)^2$ だと、

$$g'(x) = 2x, h'(x) = 3x^2 \quad \text{なので、}$$

と

なる

合成関数の微分（続き）

- $f(x) = g(h(x))$ の微分は、

$$f'(x) = g'(h(x))h'(x) \quad \text{【重要公式】}$$

... 残念ながら証明略, ちなみに $(f(x))'$ は簡単に $f'(x)$ と書ける (以下同様)

- 前の例 $f(x) = (x^3 + 1)^2$ だと、

$$g'(x) = 2x, h'(x) = 3x^2 \quad \text{なので、}$$

$$f'(x) = 2(x^3 + 1) \cdot 3x^2 = 6x^2(x^3 + 1) \quad \text{と}$$

なる

合成関数の微分（続き）

- 実際、 $f(x) = (x^3 + 1)^2 = x^6 + 2x^3 + 1$ なので、これを微分すると

とな

り一致する（！）

- この例ではたまたま（？）展開できたので簡単に検証できた

合成関数の微分（続き）

- 実際、 $f(x) = (x^3 + 1)^2 = x^6 + 2x^3 + 1$ なので、これを微分すると

$$f'(x) = 6x^5 + 2 \cdot 3x^2 = 6x^2(x^3 + 1) \text{ となり一致する（！）}$$

- この例ではたまたま（？）展開できたので簡単に検証できた

合成関数の微分（続き）

- 展開できない関数も、もちろん微分可能

- $f(x) = \frac{1}{x^3 + 1}$ の場合：

- $\frac{1}{x}$ の微分は、 $\frac{1}{x} = x^{-1}$ なので $\left(\frac{1}{x}\right)' = -x^{-2}$

- よって $f'(x) = -(x^3 + 1)^{-2} \cdot 3x^2$



合成関数の微分（続き）

- 展開できない関数も、もちろん微分可能

- $f(x) = \frac{1}{x^3 + 1}$ の場合：

- $\frac{1}{x}$ の微分は、 $\frac{1}{x} = x^{-1}$ なので $\left(\frac{1}{x}\right)' = -x^{-2}$

- よって $f'(x) = -(x^3 + 1)^{-2} \cdot 3x^2$

$$= \frac{-3x^2}{(x^3 + 1)^2}$$

合成関数の微分（分数関数）

発展として、分数関数の微分の公式は

(証明略)

■ これを適用しても $f(x) = \frac{1}{x^3 + 1}$ は

$$f'(x) = \frac{0 \cdot (x^3 + 1) - 1 \cdot 3x^2}{(x^3 + 1)^2} = \frac{-3x^2}{(x^3 + 1)^2}$$

合成関数の微分（分数関数）

発展として、分数関数の微分の公式は

$$\left(\frac{f(x)}{g(x)} \right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \quad (\text{証明略})$$

■ これを適用しても $f(x) = \frac{1}{x^3 + 1}$ は

$$f'(x) = \frac{0 \cdot (x^3 + 1) - 1 \cdot 3x^2}{(x^3 + 1)^2} = \frac{-3x^2}{(x^3 + 1)^2}$$

ネイピア数 e

- ネイピア数 e : 微積分で非常に重要、かつ頻繁に出てくる定数
- $e = 2.718...$ という無理数
- , すなわち
 - 何回微分しても、自分自身に戻る、あるいは、 e^x の値がそのまま微分係数になっている、とも捉えられる
 - たとえば $e^0 = (e^0)' = 1 \leftarrow x = 0$ のとき

ネイピア数 e

- ネイピア数 e : 微積分で非常に重要、かつ頻繁に出てくる定数
- $e = 2.718...$ という無理数
- $(e^x)' = e^x$, すなわち
 - 何回微分しても、自分自身に戻る、あるいは、 e^x の値がそのまま微分係数になっている、とも捉えられる
 - たとえば $e^0 = (e^0)' = 1 \leftarrow x = 0$ のとき

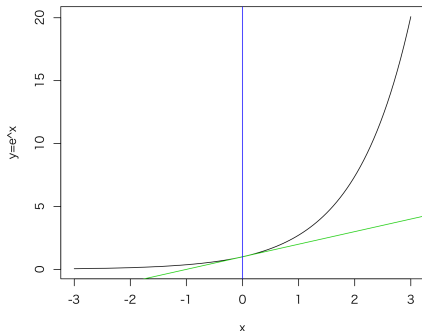


ネイピア数 e

- ネイピア数 e : 微積分で非常に重要、かつ頻繁に出てくる定数
- $e = 2.718...$ という無理数
- $(e^x)' = e^x$, すなわち
 - 何回微分しても、自分自身に戻る、あるいは、 e^x の値がそのまま微分係数になっている、とも捉えられる
 - たとえば $e^0 = (e^0)' = 1 \leftarrow x = 0$ のとき

$y = e^0$ の値も接線の傾きも 1

ネイピア数 e



☆縦横のスケールが違うので分りにくい
が
$$e^0 = (e^0)' = 1$$

ネイピア数（続き）

☆ちょっと（めちゃめちゃ？）ややこしいのですが雰囲気だけでも掴んでください

- e^x のマクローリン展開（詳細略）は

$$1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{i=0}^{\infty} \frac{x^i}{i!} \quad (\text{無限級数})$$

- $1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$ を微分すると、

$$0 + \frac{1}{1!} + \frac{2x}{2!} + \frac{3x^2}{3!} + \frac{4x^3}{4!} \dots = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

となり、確かに同じになる（！）

ネイピア数（続き）

- ちなみに (i は虚数単位、 π は円周率) は、 という、数学で最も美しい式のうちの1つ
- ぱっと見、 i で冪乗 って意味不明、で、本当はこれだけで 1 年授業できる内容 (詳細略)

ネイピア数（続き）

- ちなみに $e^{i\pi} + 1 = 0$ (i は虚数単位、 π は円周率) は、 という、数学で最も美しい式のうちの1つ
- ぱっと見、 i で冪乗 って意味不明、で、本当はこれだけで 1 年授業できる内容 (詳細略)

ネイピア数（続き）

- ちなみに $e^{i\pi} + 1 = 0$ (i は虚数単位、 π は円周率) は、**オイラーの等式** という、数学で**最も美しい式**のうちの1つ
- ぱっと見、 i で冪乗 って意味不明、で、本当はこれだけで 1 年授業できる内容 (詳細略)

シグモイド関数

- $\text{sigmoid}_a(x) = \frac{1}{1 + e^{-ax}}$
- $(\varsigma_a(x))$ と書かれることが多い)
- ディープラーニングでパラメータ学習するときの するための代表的な関数
- 値域（計算結果の値の範囲）は

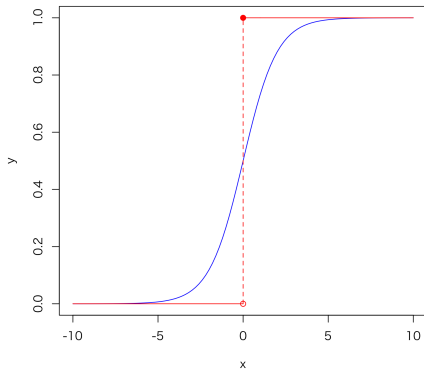
シグモイド関数

- $\text{sigmoid}_a(x) = \frac{1}{1 + e^{-ax}}$
- $(\varsigma_a(x))$ と書かれることが多い)
- ディープラーニングでパラメータ学習するときの **閾値判定** するための代表的な関数
- 値域（計算結果の値の範囲）は

シグモイド関数

- $\text{sigmoid}_a(x) = \frac{1}{1 + e^{-ax}}$
- $(\varsigma_a(x))$ と書かれることが多い)
- ディープラーニングでパラメータ学習するときの **閾値判定** するための代表的な関数
- 値域（計算結果の値の範囲）は $0 < \text{sigmoid}_a(x) < 1$

シグモイド関数（グラフ）



☆ 赤：階段関数, 青：シグモイド関数

シグモイド関数（続き）

- 単純な判定は、たとえば「パラメータ（の合計）が0未満ならば0、0以上ならば1」
 のように出来るが、 $x = 0$ で不連続なので微分できない（詳細略）
- 微分できないと代数的に極小（最小）値を求められないのでシグモイド関数を用いられる
- つまり、不連続だがよく使う関数を
 関数で近似している
- a は (x の係数: グラフの傾き)
以下は計算の簡単のため $a = 1$ の場合について考える

シグモイド関数（続き）

- 単純な判定は、たとえば「パラメータ（の合計）が0未満ならば0、0以上ならば1」
【階段関数】のように出来るが、 $x = 0$ で不連続なので微分できない（詳細略）
- 微分できないと代数的に極小（最小）値を求められないのでシグモイド関数を用いられる
- つまり、不連続だがよく使う関数を
関数で近似している
- a は (x の係数: グラフの傾き)
以下は計算の簡単のため $a = 1$ の場合について考える

シグモイド関数（続き）

- 単純な判定は、たとえば「パラメータ（の合計）が0未満ならば0、0以上ならば1」
【階段関数】のように出来るが、 $x = 0$ で不連続なので微分できない（詳細略）
- 微分できないと代数的に極小（最小）値を求められないのでシグモイド関数を用いられる
- つまり、不連続だがよく使う関数を
なめらかな関数で近似している
- a は (x の係数: グラフの傾き)
以下は計算の簡単のため $a = 1$ の場合について考える

シグモイド関数（続き）

- 単純な判定は、たとえば「パラメータ（の合計）が0未満ならば0、0以上ならば1」
【階段関数】のように出来るが、 $x = 0$ で不連続なので微分できない（詳細略）
- 微分できないと代数的に極小（最小）値を求められないのでシグモイド関数を用いられる
- つまり、不連続だがよく使う関数を
なめらかな関数で近似している
- a は ゲイン（ x の係数:グラフの傾き）
以下は計算の簡単のため $a = 1$ の場合について考える

シグモイド関数（続き）

$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$ を微分すると、

$$\begin{aligned}\text{sigmoid}'(x) &= \frac{-1}{(1 + e^{-x})^2} \cdot (1 + e^{-x})' = \frac{-1 \cdot -e^{-x}}{(1 + e^{-x})^2} \\ &= \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{1 + e^{-x}} \frac{e^{-x}}{1 + e^{-x}} \\ &= \frac{1}{1 + e^{-x}} \cdot \frac{(1 + e^{-x}) - 1}{1 + e^{-x}} \\ &\quad (\text{次頁へ続く})\end{aligned}$$

シグモイド関数（続き）

（前頁 $\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$ の微分の続き…）

$$\begin{aligned}\text{sigmoid}'(x) &= \frac{1}{1 + e^{-x}} \cdot \left(1 - \frac{1}{1 + e^{-x}}\right) \\ &= \boxed{\phantom{\frac{e^{-x}}{(1 + e^{-x})^2}}}\end{aligned}$$

（ゲイン a があるときは、微分の頭に a が係数としてつく）

☆微分関数が元の関数から計算できることに注意（！）

シグモイド関数（続き）

（前頁 $\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$ の微分の続き…）

$$\begin{aligned}\text{sigmoid}'(x) &= \frac{1}{1 + e^{-x}} \cdot \left(1 - \frac{1}{1 + e^{-x}}\right) \\ &= \boxed{\text{sigmoid}(x) \cdot (1 - \text{sigmoid}(x))}\end{aligned}$$

（ゲイン a があるときは、微分の頭に a が係数としてつく）

☆微分関数が元の関数から計算できることに注意（！）

シグモイド関数（練習）

1 R を立ち上げ、以下を実行する

シグモイド関数

```
sg <- function(x) {1/(1+exp(-x))}
```

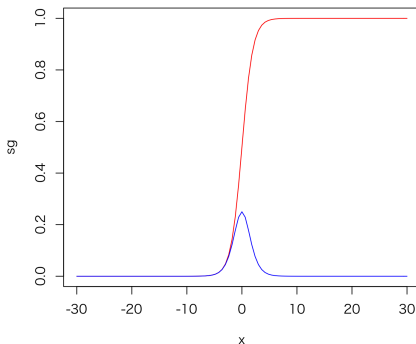
↑の微分が↓

```
sgd <- function(x) {sg(x)*(1-sg(x))}
```

```
plot(sg,-30,30, col=2) # プロット（↓も同様）
```

```
plot(sgd,-30,30, col=4, add=TRUE) #
```

シグモイド関数の微分のグラフ



- 0 付近のみ値があり、それ以外はほぼ 0=微分前のグラフが平ら、ということを意味している（！）

ハイパボリック・タンジェント (tanh)

- 双曲線正接とも呼ぶ
- 三角関数 $\sin, \cos, \tan$ と似た (?) 双曲線関数 $\sinh, \cosh, \tanh$ のうちの1つ (詳細略)
- $$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
- ディープラーニングでパラメータ学習するときの、シグモイドと並ぶ代表的な関数 (特に LSTM などのリカレント系)
- シグモイドの値域が $0 < \text{sigmoid}(x) < 1$ なのに対して、 $\tanh$ の場合は であることに注意

ハイパボリック・タンジェント (tanh)

- 双曲線正接とも呼ぶ
- 三角関数 $\sin, \cos, \tan$ と似た (?) 双曲線関数 $\sinh, \cosh, \tanh$ のうちの1つ (詳細略)
- $$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
- ディープラーニングでパラメータ学習するときの、シグモイドと並ぶ代表的な関数 (特に LSTM などのリカレント系)
- シグモイドの値域が $0 < \text{sigmoid}(x) < 1$ なのに対して、 $\tanh$ の場合は $-1 < \tanh(x) < 1$ であることに注意

tanh (続き)

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \text{ を微分すると (分数関数)}$$

$$\begin{aligned}\tanh'(x) &= \frac{(e^x - e^{-x})'(e^x + e^{-x}) - (e^x - e^{-x})(e^x + e^{-x})'}{(e^x + e^{-x})^2} \\&= \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2} \\&= \frac{(e^x + e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} \\&= 1 - \tanh^2(x)\end{aligned}$$

☆シグモイド関数同様、微分関数が元の関数から計算できることに注意 (！)

tanh (練習)

1 R を立ち上げ、以下を実行する

tanh 関数

th <- function(x) # ↓と、本当は1行

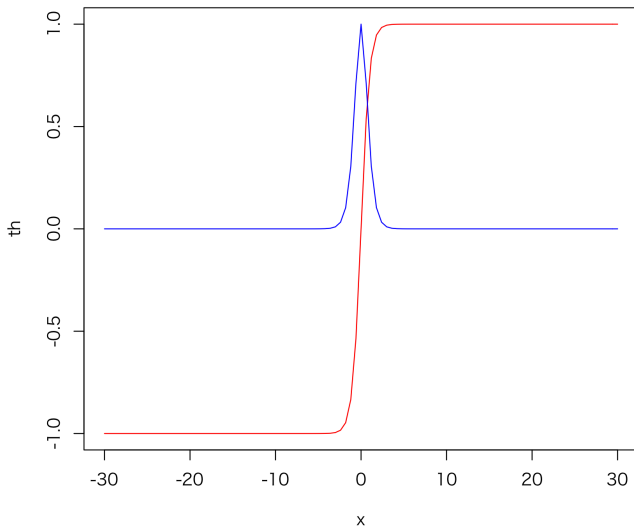
{ (exp(x)-exp(-x)) / (exp(x)+exp(-x)) }

thd <- function(x) {1-th(x)**2} # ↑の微分

plot(th,-30,30, col=2) # プロット (↓も同様)

plot(thd,-30,30, col=4, add=TRUE)

tanh (続き)



おまけ：偏微分とは

- 偏微分とは、2変数以上の関数を微分するときに、他の変数は無視してその変数を変化させたときの関数の変化の量をあらわしている
- $\Rightarrow$ 偏微分の値の逆の方向に結合荷重を変化させると、誤差を減らす（0の場合は増やさない） = (gradient descent method)
- 最適解は保証されない（局所的な最小値＝極小値で学習が終了する可能性がある）
- その回避のための もある（詳細略）

おまけ：偏微分とは

- 偏微分とは、2変数以上の関数を微分するときに、他の変数は無視してその変数を変化させたときの関数の変化の量をあらわしている
- $\Rightarrow$ 偏微分の値の逆の方向に結合荷重を変化させると、誤差を減らす（0の場合は増やさない） = **最急降下法** (gradient descent method)
- 最適解は保証されない（局所的な最小値＝極小値で学習が終了する可能性がある）
- その回避のための もある（詳細略）

おまけ：偏微分とは

- 偏微分とは、2変数以上の関数を微分するときに、他の変数は無視してその変数を変化させたときの関数の変化の量をあらわしている
- $\Rightarrow$ 偏微分の値の逆の方向に結合荷重を変化させると、誤差を減らす（0の場合は増やさない） = **最急降下法** (gradient descent method)
- 最適解は保証されない（局所的な最小値＝極小値で学習が終了する可能性がある）
- その回避のための **確率的勾配降下法** もある（詳細略）

おまけ：偏微分

- $f(x, y) = x^2 + y^2$ のような、2つ以上変数がある関数を微分する場合を考える
- これを、ある1つの変数で微分する場合、どの変数に対しての微分かを区別する必要がある
- これが ... のように書く
 - ちなみに ∂ はいろんな読み方がある が、個人的には「でる」に1票(?)
 - さらにちなみに 全微分というものもある（詳細略）

おまけ：偏微分

- $f(x, y) = x^2 + y^2$ のような、2つ以上変数がある関数を微分する場合を考える
- これを、ある1つの変数で微分する場合、どの変数に対しての微分かを区別する必要がある
- これが **偏微分** ...  のように書く
 - ちなみに ∂ はいろんな読み方がある が、個人的には「でる」に1票(?)
 - さらにちなみに 全微分 というものもある (詳細略)

おまけ：偏微分

- $f(x, y) = x^2 + y^2$ のような、2つ以上変数がある関数を微分する場合を考える
- これを、ある1つの変数で微分する場合、どの変数に対しての微分かを区別する必要がある
- これが **偏微分** ... $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ のように書く
 - ちなみに ∂ はいろんな読み方があるが、個人的には「でる」に1票(?)
 - さらにちなみに 全微分 というものもある (詳細略)

おまけ：偏微分

- その場合、微分しないほうの変数は として扱う（固定化）
- たとえば関数 $f(x, y) = x^2 + y^2$ の $(x, y) = (3, 4)$ での偏微分は
 - $\frac{\partial f}{\partial x} = 2x = 6$ (y^2 は定数なので無くなる)
 - $\frac{\partial f}{\partial y} = 2y = 8$ (x^2 は定数なので無くなる)

おまけ：偏微分

- その場合、微分しないほうの変数は **定数** として扱う（固定化）
- たとえば関数 $f(x, y) = x^2 + y^2$ の $(x, y) = (3, 4)$ での偏微分は
 - $\frac{\partial f}{\partial x} = 2x = 6$ (y^2 は定数なので無くなる)
 - $\frac{\partial f}{\partial y} = 2y = 8$ (x^2 は定数なので無くなる)

まとめ

- とりあえず、線形代数（特に行列計算）と今回の微分が（最悪、完全に理解出来ていなくても）
 ようになれば、ディープラーニングに使われる数式は扱えます
- AI・データサイエンスを学ぶ以上、最低限の数式には慣れないと先に進めないので心してください（！）
- かなり詳細を略しているので、興味のある人は是非調べたり参考書を紐解いたりしてください

まとめ

- とりあえず、線形代数（特に行列計算）と今回の微分が（最悪、完全に理解出来ていなくても）
使える ようになれば、ディープラーニングに使われる数式は扱えます
- AI・データサイエンスを学ぶ以上、最低限の数式には慣れないと先に進めないので心してください（！）
- かなり詳細を略しているので、興味のある人は是非調べたり参考書を紐解いたりしてください

おしまい

質問があれば、お気軽に！