・微分の概要について素人にも分かるように簡潔に説明せよ。

微分とは細かくすること。微分の考え方としては、細かくしたうえで、性質を見抜いて、何かの処理をする。

数式的に考えると、微分とは $f(x)$ から $f'(x)$ を作る手法である。

ここで $f'(x)$ とは、曲線上の x の場所の傾きを表す。

式で表すと、$f'(x)$ は曲線 $y=f(x)$ 上の点 (x, y) での傾きであるから、

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

ここで $\lim$ とは、英語の limit（極限）から作られた記号で、

$\lim_{h \to 0}$ は「h を極限まで 0 に近づける」という意味。

傾きを与える式を「導関数」という。

極限の概念から、h を 0 に近づけると、接線の傾き $f'(x)$ を得ることができる。

微分の計算法則は以下の通り。

- $(x^n)' = nx^{n-1}$, 特に $(\text{定数})' = 0$, $x' = 1$
- $(f(x) + g(x))' = f'(x) + g'(x)$
- $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$
- $\{f(g(x))\}' = f'(g(x))g'(x)$
- $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)gx - f(x)g'(x)}{\{g(x)\}^2}$

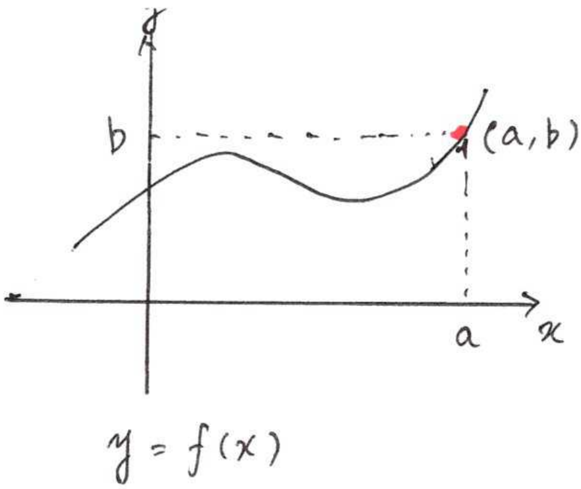
・偏微分の概要について素人にも分かるように簡潔に説明せよ。

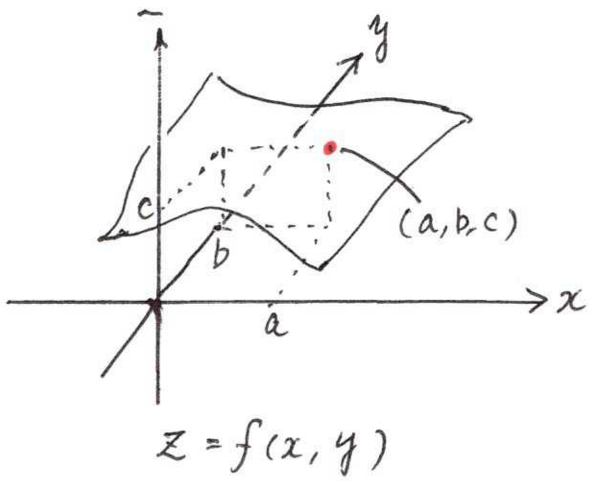
偏微分とは、微分する変数以外を定数としたときの微分をいう。すなわち、多変数関数に対して1つの変数のみに関する微分である。

例えば2変数関数の微分を考えてみる。

2変数関数 $z=f(x, y)$ は、x と y の値を決めたときに、z の値が決まる法則を表す。この様子は、xyz 空間に曲面で表される。

一方で、1変数関数 $y=f(x)$ は曲線を表すため、次元が1次あがって、曲面になったと考えられる。



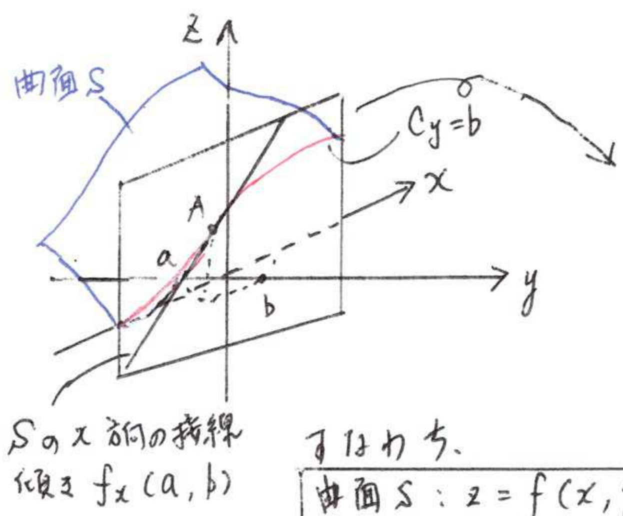


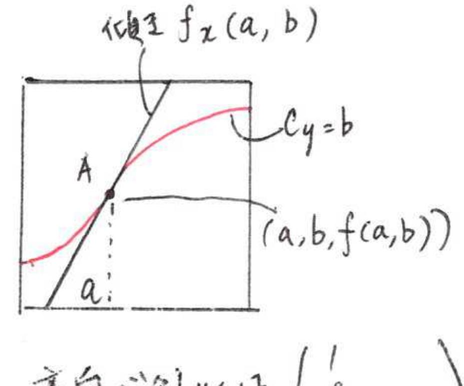
$y=f(x)$ 上の $x=a$ のとき、点 $A(a, f(a))$ での微分係数 $f'(a)$ は、接線の傾きを表す。

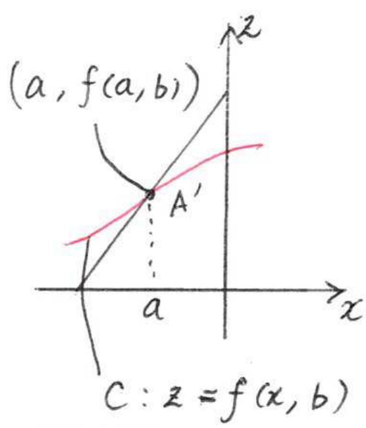
一方、$f(x, y)$ に関して、y を定数と見て x で微分することを $f(x, y)$ の x による偏微分といい、$f(x)$、$f_x(x, y)$、$\dfrac{\partial f}{\partial x}$ などと表す。このとき、$f_x(a, b)$ を $f(x, y)$ の (a, b) における x 方向の偏微分係数といい、接線の傾きを表している。

曲面は $(x, y, f(x, y))$ と表され、b を定数として $y=b$ とすると、$(x, b, f(x, b))$ となり、y 座標がつねに b であるため、$y=b$ という平面上の点と曲面との交わりになる。(曲線になる。)

ここで、$(a, b, f(a, b))$ を点 A とし、平面 $y=b$ 上で考えたとき、上記の曲線を $C_{y=b}$ とすると、$C_{y=b}$ の A での接線の傾きが $f_x(a, b)$ とな







方向ベクトルは $\begin{pmatrix} 1 \\ 0 \\ f_x(a, b) \end{pmatrix}$

すなわち、

曲面 $S: z=f(x, y)$ の $A(a, b, f(a, b))$ での
x 方向の接線の傾きは、$f_x(a, b)$
x 方向の接線の方向ベクトルは、$\begin{pmatrix} 1 \\ 0 \\ f_x(a, b) \end{pmatrix}$

- 微分/偏微分の機械学習、深層学習における使用について素人にも分かるように簡潔に説明せよ。

人工知能では、関数の値がどの地点で最小値をとるのかを調べることが多い。例えば、損失関数とは、正解値と予測値の誤差を表す関数だが、この関数を最小化するような値を求める方法を考える。

このとき、損失関数を「微分」すると、ある瞬間に損失関数がどちらにどれくらい傾いているのか知ることができる。傾きの絶対値の大きさを小さくする方向に徐々にずらし、損失関数の最小値を求める手法を勾配降下法といい、深層学習(ディープラーニング)で重要な役割を果たす手法の一つになっている。

また、最小2乗法は、誤差の2乗和が最小になるような関係式を求める方法で、線形回帰と呼ばれる最も基本的な人工知能のアルゴリズムにて使われる最適化手法である。$f(a,b)=$(誤差の2乗和)としたとき、$\frac{\partial f(a,b)}{\partial a}=0$ かつ $\frac{\partial f(a,b)}{\partial b}=0$ となるような方程式を解くことで、誤差の2乗和が最小になるような関係式を求めることができる。これは、2次関数を1階微分した関数の値が0のとき、極値(最小2乗法の場合は、必ず最小値)をとるという特性を使っている。
次に、ディープラーニングのアルゴリズムであるディープニューラルネットワーク(DNN)の特徴をみてみる。ニューラルネットワークとは人間の脳にある神経細胞(ニューロン)と、そのネットワークを模した数学モデルのことである。ニューラルネットワークはデータを読み込む入力層、最終的なデータを出力する出力層、そして入力層と出力層の間には中間層と呼ばれる人工のニューロンによって構成される。一般に、この中間層が2層以上あるものをDNNという。
ニューラルネットワークで学習するとき、正解データとニューラルネットワークの出力が合うように、ニューラルネットワークの重み(w)を調整する。このとき、ニューラルネットワークの重みの調整量は誤差の値を重みで偏微分した値を考慮したものになる。重みで偏微分した値を計算するために、このチェーンルールを使う。こうした手法を誤差逆伝播法という。

また、ニューラルネットワークの重み(w)の調整量は、誤差の値を重みで偏微分した値が考慮されるのと同様に、活性化関数も微分される。

ここで、活性化関数とは、人工知能モデルの表現力を高めるために使われるクッションのようなものである。

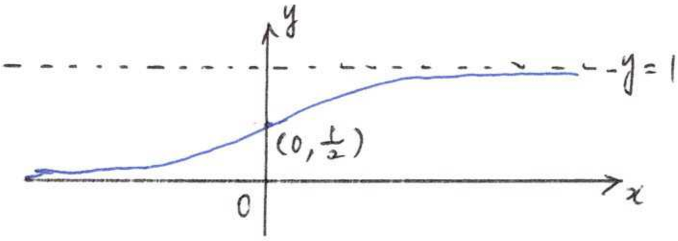
活性化関数を使うと、非線形分離（＝曲線で分離すること）が可能になるため、複雑な関係性も表現できるようになる。そのためニューラルネットワークなどの人工知能モデルでシグモイド関数などが使われる。

シグモイド関数は、人工知能で頻出する関数の一つで、定義は次の式で表せる。

$$\varsigma_a = \frac{1}{1+\exp(-ax)}$$

；このとき、aをゲインと呼び、特に$a=1$のときのシグモイド関数を標準シグモイド関数という。

（ςはギリシャ文字Σ（シグマ）の語末形）



（シグモイド関数のグラフ）

シグモイド関数の特徴としては、xが負の無限大に近づくと分母は正の無限大になるのでyは0に近づき、xが正の無限大に近づくと分母が1に近づくのでyは1に近づく。また常に$\varsigma_a(0)=\frac{1}{2}$となる。$a$の値が大きくなるほど変化の度合が大きくなる。

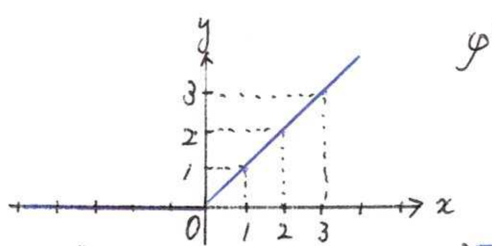
標準シグモイド関数は微分したとき、微分値の最大が0.25になる。そのため、ニューラルネットワークの層が深いと、誤差逆伝播法で誤差が伝播しなくなってしまう問題がある。これを勾配消失問題という。

その解決策として、微分値が0か1をとるReLU関数の利用が提唱されている。ここで、ReLU関数は、活性化関数の一つであり、ディープラーニングの一つであるDNN、CNNに使われる。
（CNNは、機械学習においての畳み込みニューラルネットワークで、最小限のデータ前処理しか必要としないように設計された、順伝播型のニューラルネットワークである。画像や動画認識に広く使われているモデルである。）

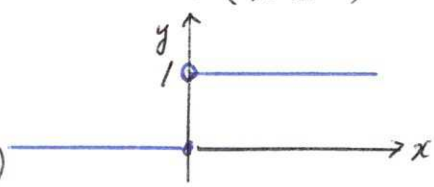
ReLU関数の公式．

$$\varphi(x) = \max(0, x) = \begin{cases} x & (x > 0) \\ 0 & (x \leqq 0) \end{cases}$$

$$\varphi'(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x \leqq 0) \end{cases}$$



（ReLU関数$\varphi(x)$のグラフ）



（微分したReLU関数$\varphi'(x)$のグラフ）

これにより、勾配消失問題が軽減されるため、現在のニューラルネットワークの活性化関数としてはReLU関数が多く利用されている。

合成関数の微分

例題 1 $y=(x^2+3x+1)^4$

$u=x^2+3x+1$ とおくと、

$$
\begin{aligned}
y' &= \left((x^2+3x+1)^4\right)' \\
&= (u^4)' \\
&= 4u^3 \cdot u' \\
&= \underline{\underline{4(x^2+3x+1)^3(2x+3)}}
\end{aligned}
$$

例題 2 $y=\log(\sin(x^3-2))$

まず、$\sin(x^3-2)$ を微分すると、

$$(\sin(x^3-2))' = \cos(x^3-2)\cdot 3x^2$$

したがって、

$$
\begin{aligned}
y' &= \frac{1}{\sin(x^3-2)}\cdot\cos(x^3-2)\cdot 3x^2 \\
&= \underline{\underline{\frac{3x^2}{\tan(x^3-2)}}}
\end{aligned}
$$

合成関数の偏微分

$f(x,y)=(x^2+y^2)\sin xy$ に対して、

偏導関数 $\dfrac{\partial f}{\partial x}$ を求めます。

$$\frac{\partial f}{\partial x} = \underline{\underline{2x\sin xy+(x^2+y^2)\cos xy}}$$