# 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

微分は関数の接線の傾きを算出する方法である。 ある点の傾きがわかると、それを手がかりに最小の点を探すことができる。接線の傾きの符号によって関数の最小の位置を把握することができる。 つまり、接戦の傾きは関数の増加・減少の仕方を表す。 接戦の傾きが右肩上がりであれば、その関数はその点で増加していることを表す。 勾配降下法は接線の傾きを手がかりに計算していく方法である。 微分を理解するためには、まず平均変化率という概念を理解する必要がある。 平均変化率とは2点を結んだ直線の傾きである。

$y = f(x)$とした時に$f(a+h) - f(a) \leftarrow \frac{y\text{の増加量}}{x\text{の増加量}}$で傾きが求められる。

ここで極限の概念から、hを限りなくaに近づけることで、2点の直線は接線に近づいていく。 したがってhを０に近づけると、接戦の傾き(微分係数)f'(x)を得ることができる。

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

xに値を入れるとその点での接戦の傾きを返す関数である定数aを変数xに置き換えたものを導関数と呼ぶ。 このようにf(x)の導関数を求めることを微分するという

合成関数の微分は、ディープラーニングで最重要の内容である。 以下ではライプニッツの記法( $\frac{dy}{dx}$)のように,何で微分しているかを明確に示すようにする。

$f(x) = (x^2 + 2x + 5)^{10}$

$t = x^2 + 2x + 5$のようにかたまりをtと置くと

$(\text{与式}) = t^{10}, t = x^2 + 2x + 5$

fはまずtの関数となり、tはxの関数という関係となっている。 つまり合成関数の微分とはfがtの関数、tがxの関数であるとき $\frac{dy}{dx} = \frac{dy}{dx}\frac{dy}{dx}$となる

元を辿ればfはxの関数だから原理的にはxで微分できることになる。 また、合成関数の微分から導かれる事と、ひとかたまりが$ax+b$の形の場合に限り $\frac{df(ax+b)}{dx} = af'(ax+b)$と定義できる

また、高階微分としての2階微分の意味は関数の凹凸を表し

$\frac{d^2f}{dx^2} > 0$の場合は下に凸

$\frac{d^2f}{dx^2} = 0$の場合は変曲点

$\frac{d^2f}{dx^2} < 0$の場合は関数が上に凸

以上のような意味を表す。 １階微分は関数の増減を把握して、２階微分によって増減の詳細を得ることができる。極大極小とは周囲のどの点よりも大きい(小さい)点のことで極大or極小ならば $\frac{df(x)}{dx} = 0$

極大か極小かの判断については２階微分を計算して結果が>0の時極小、<0の時極大

# 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時の微分のことである。$f(x,y)=x^2+y^2$のように２つ以上の変数をもつ関数を多変数関数という。$f(x,y)$をxで偏微分する場合は着目する文字を１つに決めて、注目した文字以外を全て定数とみなして計算していく。仮に$f(x,y)$をxで偏微分する場合、$\frac{\partial f(x,y)}{\partial x}$という記号で偏微分を表す。

偏微分は高階微分を考えられる。２変数関数$f(x,y)$で考えてみたときに、微分する順番は下記の４パターンが考えられる。

1.x→y　$\leftrightarrows \frac{\partial^2 f}{\partial x^2}$

2.y→y　$\leftrightarrows \frac{\partial^2 f}{\theta y^2}$

3.x→y　$\leftrightarrows \frac{\partial^2 f}{\partial x \partial y}$

4.y→x　$\leftrightarrows \frac{\partial^2 f}{\partial y^2}$

合成関数の偏微分偏導関数 を求める場合の考え方として、xを入力し、xに定数wがかけられ,uとして出力される。uが関数fの引数となり、fにより変換される（これをyとする)出力yは元を辿れば、xの関数と考えることができる。

# 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

与えられたデータから未来に起こることを予測する時に機械学習や深層学習の手法が用いられる。

機械学習とはデータから関数を作らせる仕事である。具体的に言うとまず、作りたい関数の性質を考えて、関数を作らせるためのどのMLアルゴリズムを使うかを決めてデータをMLアルゴリズムに当たることで関数を作らせい出来上がった関数を使う。関数には調整用のパラメータ変数θがある$y=f(x,\theta)$

機械学習アルゴリズムはデータに合わせて、最も良いパラメータ(θ)を自動的に決定する。

このパラメータ(θ)に対する評価を数値化して判断する評価関数の結果によってθを決定する。

$E=h(\theta)$

目的関数Eが利得値ならば最大のものを選び、Eが損失値を選ぶならば最小のものを選ぶ。

つまり関数の最大、最小を求めることが機械学習アルゴリズムを作る上で欠かせないことになるので特にこの部分で微分法が使用される。

過去のデータ全てを描画したグラフにおいて、各点を通る関数を定義できれば未来のデータを予測できるが関数が全ての点を通る訳ではない。そこで過去のデータとの「誤差」を限りなく小さくするよう計算して未来を予測していく。

関数を$f_\theta(x) = \theta_0 + \theta_1 x$とすると、データの値をyとすると誤差は$y - f_\theta(x)$で表される。誤差には正負があるため誤差を２乗して正の値にすると$(y - f_\theta(x))^2$とする。全ての座標の値を足すと

$E_\theta = \sum_{i=0}^{n}(y - f_\theta(x))^2$

計算の都合上この値に1/2をかけたものを２乗誤差という。この目的関数の値を小さくしていく時に微分を使う。導関数の符号の変化で関数の状態を確認して、xの値を動かしながら最小値を導いていく方法を勾配降下法と呼ぶ。

xの更新式は　$x := x - \eta \frac{d}{dx} g(x)$　と表す

:とは左辺を右辺で定義するという意味で、η(エータ)は学習率と呼ばれる正の定数である

この時、複数のパラメータ($\theta_0 - \theta_n$)を持つ場合偏微分を用いて計算することになる。

# 合成関数の微分

## 例題１

$y = (x^2 + 3x + 1)^4$

$x^2 + 3x + 1$をtとおくと

(与式) $= t^4$

したがって $\frac{dx}{dt} = 4t^3, \frac{dt}{dx} = 2x + 3$

そこで、$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx}$

与式 =$4(x^2 + 3x + 1)^3(2x + 3)$

## 例題２

$y = log(sin(x^3 - 2))$

$sin(x^3 - 2) = t$、$x^3 - 2 = u$とおく

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

(与式) $= \frac{1}{t}cos(u)3x^2$

$= \frac{1}{sin(u)}cos(x^3 - 2)3x^2$

$= \frac{3x^2 cos(x^3-2}{sin(x^3-2)}$

$= \frac{3x^2}{tan(x^3-2}$

-

# 合成関数の偏微分

$f(x,y)=(x^2+y^2)sinxy$　に対して、偏導関数 $\frac{\partial f}{\partial x}$ を求めよ

$h(x,y)=(x^2+y^2)$、$g(x,y)=sinxy$

以上のように置き換えると $f=gh$　となり、両辺をxで偏微分すると

$\frac{\partial f}{\partial x}=\frac{\partial h}{\partial x}g+\frac{\partial g}{\partial x}f$

（与式）$=x^2ycosxy+2xsinxy+y^3cosxy$