微分の概要について素人にも分かるように説明せよ

結論からいうと、以下のようにある関数$f(x)$ 上の点を通る接線の傾きを得る操作を「微分」と言います。

例えば関数 $y = x^2$ 上の $x = 1$ の点Aとそこからx との距離がh 離れた$1 + h$との2点間を通る直線の傾きを求めていきます。

$$
\begin{aligned}
\text{直線の傾き} &= \frac{(1+h)^2 - 1^2}{(1+h) - 1} \\
&= \frac{(1 + 2h + h^2) - 1}{h} \\
&= \frac{2h + h^2}{h} \\
&= 2 + h
\end{aligned}
$$

となります。 xの2点間の距離hを限りなく0に近づけていくと2点間を結ぶ直線の傾きは、$x + h$のhを０にしたときの値、つまり２になります。これは$y = x^2$の１点A(1,1)を通る直線すなわち接線の傾きを求めたことになります。 今度は$x = 1$以外の全ての点における傾きを求めてみます。 x= 1 の点Aでの接線の傾きを求めた方法と同様に、ある点xからhだけ離れた点$x + h$との2点間を通る直線の傾きは

$$
\begin{aligned}
\text{変化の割合} &= \frac{y\text{の増加量}}{x\text{の増加量}} \\
&= \frac{(x+h)^2 - x^2}{(x+h) - x} \\
\\
&= \frac{(x^2 + 2hx + h^2) - x^2}{(x+h) - x} \\
&= \frac{2hx + h^2}{h} \\
&= 2x + h
\end{aligned}
$$

2点間の x 座標の距離 h を限りなく 0 にすると、この式は $2x$ となります。したがって、関数 $y = x^2$ の各 x に対して、その点における傾きは $2x$ となります。
さらに一般化すると、$y = f(x)$ において上記のように各xの接線の傾きを求める式は以下のように表します。

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

この$f'(x)$ を導関数といい、さらにこの導関数を求めることを微分すると言います。

偏微分の概要について素人にも分かるように説明せよ

２つの変数 x, y で表された関数を $f(x, y)$ と書きます。このような関数を２変数関数といいます。$f(x, y)$ は以下のように定義されているとします。

$$f(x, y) = x^2 + y^2 - xy$$

上記関数は x, y が同時に動きますが、どちらか一つの変数「だけ」に「偏って」見るとどうなるかを考えるのが偏微分の考え方です。 例えば、xの動きだけに偏って見るとyは動かず定数と同様にみなすことができます。
すると、

$$f(x, y) = x^2 + y^2 - xy$$

をxで微分すると，yは定数として扱って

$$2x + 0 - y = 2x - y$$

となります。これがxで「偏微分」するということになり,

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

のように表します。 同様に

$$f(x,y) = x^2 + y^2 - xy$$

をyで「偏微分」するとは、xを定数のように扱って以下のように表します。

$$\frac{\partial f(x,y)}{\partial y} = 0 + 2y - x$$
$$= 2y - x$$

> 参照 https://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1125129201 (https://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1125129201)

微分/偏微分の機械学習、深層学習における使用について

機械学習、深層学習において、与えられたデータから今後未来に起こることを予測したい場合を考えます。 与えられたデータをプロットしたグラフで、各点を通る関数がわかれば今後ある時点での予測ができるようになります。 しかし関数が全ての点を通るわけではありません。そのときに関数を

$$f_\theta(x) = \theta_0 + \theta_1 x$$

とすると、データの値をyとすると誤差は $y - f_\theta(x)$ で表されますが、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、$(y - f_\theta(x))^2$ とします。これはある一点のみに限定されているため、全部の値を足すと

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

ここで便宜上この値に1/2をかけた値を二乗誤差といいます。

$$E_\theta = \frac{1}{2}\sum_{i=0}^{n}(y_n - f_\theta(x_n))^2 - (1)$$

この式の値を小さくして行くアプローチを最小２乗法といいます。E_θ のことを目的関数ともいいます。このE_θの値を小さくして行くときに微分を用います。たとえば二次関数$g(x) = (x-1)^2$ の最小値は$x = 1$ のときの$g(x) = 0$ になります。そのとき、$x = 1$ の周辺で$g(x)$ の導関数の符号が変化しています。x<1のときはマイナス、x<1のときはプラスです。そのことを利用し導関数の符号によって与えられたxの値をずらしてx=1に近づけていけば値が最小値に近づいていきます。 これは最急降下法、または勾配降下法ともよばれています。xの更新式は以下のように表されます。

$$x := x - \eta\frac{d}{dx}g(x) - (2)$$

と式で表します。「:」とは左辺を右辺で定義するという意味です。η は学習率と呼ばれる正の定数です。 また(1)の目的関数をみると、E_θ は$g(x)$ と同じように下に凸になった二次曲線であるから、(2)の$g(x)$ のように当てはめることができるが、$f_\theta(x)$ はθ_0 とθ_1 の２つのパラメータをもっているため、このE_θ は２つの変数を持つ二変数関数です。ですからこの場合の微分は偏微分をつかわければなりません。 ちなみに更新式は以下のようになります。

$$\theta_0 := \theta_0 - \eta\frac{d}{d\theta_0}(E)$$

$$\theta_1 := \theta_1 - \eta\frac{d}{d\theta_1}g(E)$$

合成関数の微分

例題１

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

と定義すると$y = g(f(x))$ であるから

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$
$$= g'(f(x))f'(x)$$
$$= 4(x^2 + 3x + 1)(2x + 3)$$

例題２

例題１と同様に

$$y = h(t) = logt$$
$$t = g(u) = \sin(u)$$
$$u = f(x) = x^3 - 2$$

と定義すると $y = h(g(f(x)))$ であるから

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{du}\frac{du}{dx}$$
$$= h'(t)g'(u)f'(x)$$
$$= \frac{1}{t} \times cos(u) \times 3x^2$$
$$= \frac{1}{sin(u)} \times cos(x^3 - 2) \times 3x^2$$
$$= \frac{3x^2 cos(x^3 - 2)}{sin(x^3 - 2)}$$
$$= \frac{3x^2}{tan(x^3 - 2)}$$

合成関数の偏微分

$$h(x, y) = (x^2 + y^2)$$
$$g(x, y) = sinxy$$

とすると$f = hg$となり、両辺をxで偏微分すると

$$\frac{\partial f}{\partial x} = \frac{\partial h}{\partial x}g + \frac{\partial g}{\partial x}f = 2xsinxy + (x^2 + y^2)ycosxy$$
$$= x^2 ycosxy + 2xsinxy + y^3 cosxy$$