# 微分とは

関数のある時点の傾き、つまり変化の割合を求めることを**微分する**という。

関数 $y = f(x)$があったとき、ある点 $x_1$ での$x$軸の単位あたりの$y$の増加量になる。

具体例としてある乗り物の時間（秒）と走行距離（m）の関数 $y = x^2$ を考えてみる。

$x_1$ 秒時点から$h$秒たったときの$x_1 + h$時点の距離の変化量$\Delta y$を計算すると

$$\Delta y = \frac{f(x_1 + h) - f(x_1)}{(x_1 - h) - x_1}$$
$$= \frac{(x_1 + h)^2 - x_1^2}{h}$$
$$= \frac{x_1^2 + 2x_1 h + h^2 - x_1^2}{h}$$
$$= 2x_1 + h$$

となる。これはどの時間x_iでも成り立つため$\Delta y = 2x + h$となる。

たとえば、２秒時点で次の１秒間で進む距離は$\Delta y = 2 * 2 + 1 = 5$なので5m進むことが分かる。

ではhを０に近づけていくとどうなるか。

$$f'(x) = \lim_{h \to 0} \Delta y$$
$$= \lim_{h \to 0} 2x + h$$
$$= 2x$$

となる。

この瞬間の増加量のことを $\frac{dy}{dx} = f'(x)$ と表し、今回の場合は $f'(x) = 2x$ となる。

これは**傾き**とも呼ばれるが、感覚的には**その瞬間の$x$にてどれだけ$\Delta y$が期待できるか**とか$\Delta y$**を増減させようとする意志**を表す量のイメージです。

このように$f'(x)$を求めることを**「関数$f(x)$を微分する。」**という。

# 偏微分とは

話を簡単にするために三次元空間の関数 $z = x^2 + y^3 + 2x + 3y$ のような関数を扱う。

**偏微分**とは **ある変数にだけ着目し、それ以外の変数を定数とみなして微分を行う**ことです。

今回の場合は$x$に着目する偏微分は$\frac{dz}{dx} = f_x'(x, y)$、

$y$に着目する偏微分は$\frac{dz}{dy} = f_y'(x, y)$と書きます。

実際に計算するとそれぞれ

$$\frac{dz}{dx} = 2x + 2$$
$$\frac{dz}{dy} = 3y^2 + 3$$

となります。

$x$で偏微分したときは$(z, x)$平面でのある$x_1$地点での$\Delta z$の期待値とみることができます。 当然偏微分したもののかに$y$に関する変数が残っていれば、その$\Delta z$の期待値は$y$にも左右されることになります。

# MLDLにおいて微分、偏微分はどのように使われるか？

機械学習の基本はn個の入力値$x_1, x_2, \ldots, x_n$それぞれに対し、 その入力値の重み$\theta_1, \theta_2, \ldots, \theta_n$をかけた

$f(x) = \sum_{i=0}^{n} x_i\theta_i$

を最小、もしくは最大化することを求める問題と言えます。

機械学習では$f(x)$は目的関数、コスト関数と呼びます。

例をあげるなら、家の大きさ、築年数、駅からの近さなどの入力値から家の売値を予測するモデルを作るとします。 それぞれの入力値の重み、つまり売値に寄与する入力値は重み$\theta$を大きくするのです。

データを使って、うまい$\theta$値を見つけることができれば新しい家を売りに出そうとしたときに売れる価格で市場に出すことができます。

この$\theta$を求めるのに微分、それも偏微分を使います。

偏微分はこの問題ではある$\theta_i$に着目して微分を行います。

$\frac{df(x)}{d\theta_i}$が０になったところがその最小値もしくは最大になります。

関数の形や問題の性質によってうまくいかないこともありますが、

**目的関数$f(x)$を最小、もしくは最大にするパラメータ$\theta$を求めるために$\theta$ごとの偏微分値を使う**こと

が機械学習の基本になります。

残りの問題は同じフォルダの下記ファイルに保存しております。

- 1-1-合成関数の微分.jpg
- 1-2-合成関数の偏微分.jpg