行列と関数・微分

AVILEN

はじめに

Q なぜベクトルと行列の関数を考えるの？

はじめに

Q

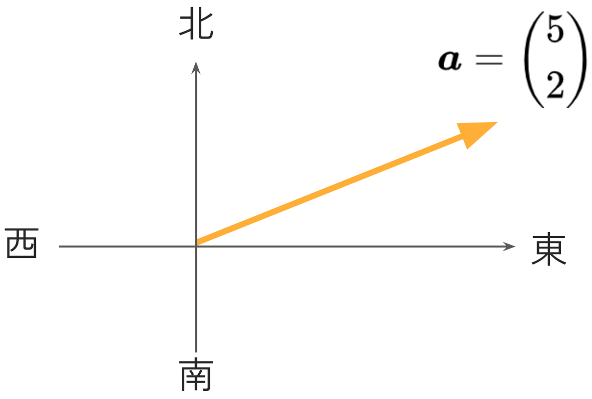
なぜベクトルと行列の関数を考えるの？

A

機械学習では、入力や出力が
スカラーとは限らないから！

ベクトルからベクトルを返す関数を考えることができる

風を例に考えてみよう



▲西南西・風速約5.4m/sの風

① 風は時間 t ごとに変わるので...

$$\boldsymbol{a}(t) = \begin{pmatrix} a_1(t) \\ a_2(t) \end{pmatrix}$$

$$t \longrightarrow \overset{\text{関数}}{\blacksquare} \longrightarrow \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

② 風は地点 $\boldsymbol{r} = (x, y)^\top$ によっても変わるので...

$$\boldsymbol{a}(t, \boldsymbol{r}) = \begin{pmatrix} a_1(t, \boldsymbol{r}) \\ a_2(t, \boldsymbol{r}) \end{pmatrix}$$

$$t, \begin{pmatrix} x \\ y \end{pmatrix} \longrightarrow \overset{\text{関数}}{\blacksquare} \longrightarrow \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

引数がベクトルになる関数の例

多変数関数 $z = f(x, y)$ を考える

引数を単に多変数とみる

関数

$$x, y \longrightarrow \blacksquare \longrightarrow z$$

引数をベクトルとみる

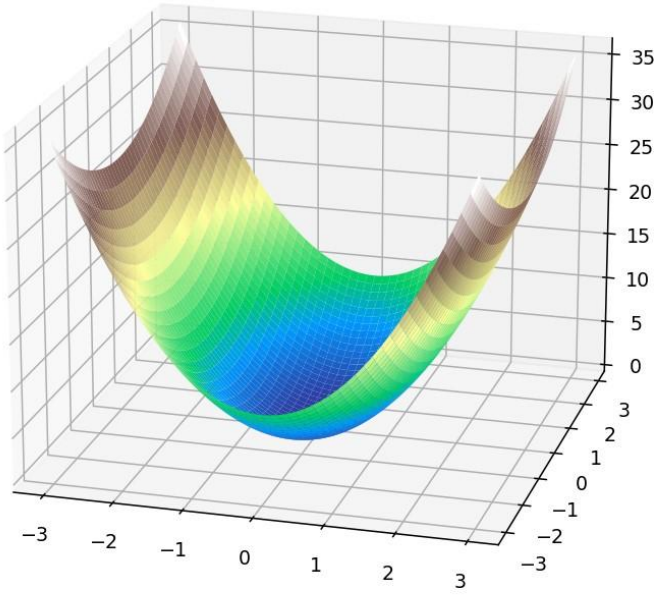
関数

$$\begin{pmatrix} x \\ y \end{pmatrix} \longrightarrow \blacksquare \longrightarrow z$$

そこで、$\boldsymbol{r} = (x, y)^\top$とおくと…

$$z = f(\boldsymbol{r})$$

のように書くことができる！



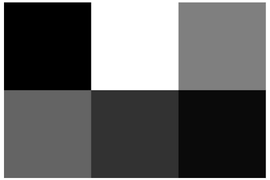
$f(x, y) = 3x^2 + y^2$ のグラフ

行列の場合も同様に考えることができる

白黒画像データ

各ピクセルの濃淡の度合いを
数値で表したもの

→ 行列とみなせる！

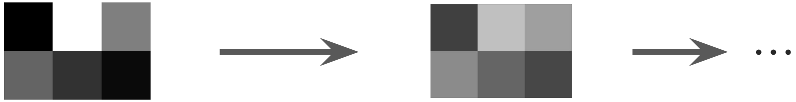


$$= \begin{pmatrix} 0 & 255 & 127 \\ 100 & 50 & 10 \end{pmatrix}$$

白黒動画データ

＝白黒画像の値が時間とともに変化

$$A(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & a_{13}(t) \\ a_{21}(t) & a_{22}(t) & a_{23}(t) \end{pmatrix}$$



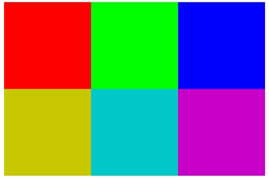
$$\begin{pmatrix} 0 & 255 & 127 \\ 100 & 50 & 10 \end{pmatrix} \qquad \begin{pmatrix} 64 & 191 & 159 \\ 139 & 101 & 71 \end{pmatrix}$$

テンソルの場合も同様に考えることができる

カラー画像データ

赤/緑/青の濃淡の度合いを
数値で表したもの

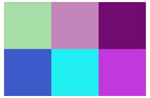
→ 3階のテンソルとみなせる！



$$= \begin{pmatrix} 255 & 0 & 0 \\ 200 & 0 & 200 \end{pmatrix} \begin{pmatrix} 0 & 255 & 0 \\ 200 & 200 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 255 \\ 0 & 200 & 200 \end{pmatrix}$$

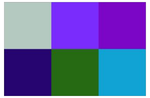
カラー動画データ

＝カラー画像の値が時間とともに変化



$$\begin{pmatrix} 167 & 196 & 114 \\ 62 & 34 & 193 \end{pmatrix} \begin{pmatrix} 222 & 133 & 11 \\ 90 & 239 & 56 \end{pmatrix} \begin{pmatrix} 167 & 186 & 111 \\ 202 & 241 & 221 \end{pmatrix}$$

→



$$\begin{pmatrix} 180 & 121 & 124 \\ 39 & 38 & 16 \end{pmatrix} \begin{pmatrix} 202 & 44 & 6 \\ 5 & 107 & 163 \end{pmatrix} \begin{pmatrix} 193 & 251 & 198 \\ 113 & 19 & 211 \end{pmatrix}$$

→ ⋯

まとめ

01 ベクトルを返す関数や、
ベクトルを引数とする関数を考えることができる

02 ベクトルだけでなく、
行列・テンソルでも同様に考えることができる

はじめに

Q

ベクトルや行列の関数を
微分すると何の役に立つの？

はじめに

Q

ベクトルや行列の関数を
微分すると何の役に立つの？

A

ベクトルや行列の関数の最小値を
求めることができる！

ベクトルを返す関数の微分

風の時間変化を調べたい

→ベクトルを返す関数 $\boldsymbol{a}(t)$ を t で微分

$$\boldsymbol{a}(t) = \begin{pmatrix} a_1(t) \\ a_2(t) \end{pmatrix}$$

$$\longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \begin{pmatrix} ? \\ ? \end{pmatrix}$$

ベクトルを返す関数をスカラーで微分するには

下のように各要素を微分すればよい

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}a_1(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}a_2(t) \end{pmatrix}$$

行列・テンソルを返す関数の微分も同様

行列やテンソルを返す関数の場合も同様に、各要素を微分すればよい。

$$A(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{pmatrix}$$

$$\longrightarrow \frac{\mathrm{d}}{\mathrm{d}t}A(t) = \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}a_{11}(t) & \frac{\mathrm{d}}{\mathrm{d}t}a_{12}(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}a_{21}(t) & \frac{\mathrm{d}}{\mathrm{d}t}a_{22}(t) \end{pmatrix}$$

ベクトルを返す関数の微分の図形的意味

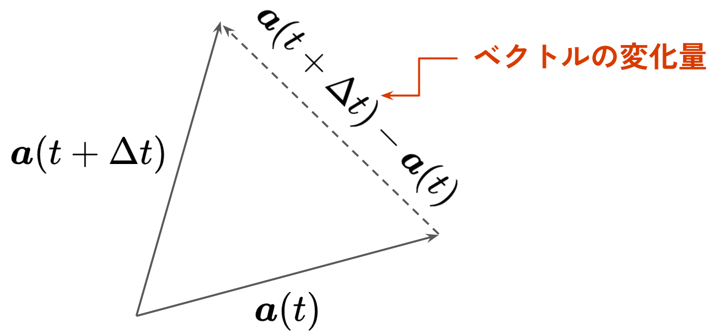
微分の定義から、
ベクトルを返す関数の微分は…

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \lim_{\Delta t \to 0} \frac{\boldsymbol{a}(t + \Delta t) - \boldsymbol{a}(t)}{\Delta t}$$

と書ける！

図形的意味

ベクトルを返す関数の微分
||
ベクトル $\boldsymbol{a}(t)$ が微小時間 Δt の間に
(Δt に対して)どれだけ変化したかを表すベクトル



例題 1. ベクトルを返す関数の微分

次の関数を t で微分せよ。

(1) $$\boldsymbol{a}(t) = \begin{pmatrix} 1 - t \\ t^2 - 2t \end{pmatrix}$$

(2) $$\boldsymbol{b}(t) = \begin{pmatrix} \cos t \\ e^t \\ t^{10} \end{pmatrix}$$

例題 1 の答え

次の関数を t で微分せよ。

(1) $\boldsymbol{a}(t) = \begin{pmatrix} 1-t \\ t^2-2t \end{pmatrix}$

答え: $\dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{a}(t) = \begin{pmatrix} (1-t)' \\ (t^2-2t)' \end{pmatrix} = \begin{pmatrix} -1 \\ 2t-2 \end{pmatrix}$

(2) $\boldsymbol{b}(t) = \begin{pmatrix} \cos t \\ e^t \\ t^{10} \end{pmatrix}$

答え: $\dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{b}(t) = \begin{pmatrix} (\cos t)' \\ (e^t)' \\ (t^{10})' \end{pmatrix} = \begin{pmatrix} -\sin t \\ e^t \\ 10t^9 \end{pmatrix}$

例題２.行列を返す関数の微分

次の関数をtで微分せよ。

$$A(t) = \begin{pmatrix} t^3 + 1 & \log t \\ e^t & \sin 2t \end{pmatrix} \quad (t > 0)$$

例題 2 の答え

次の関数をtで微分せよ。

$$A(t) = \begin{pmatrix} t^3 + 1 & \log t \\ e^t & \sin 2t \end{pmatrix} \quad (t > 0)$$

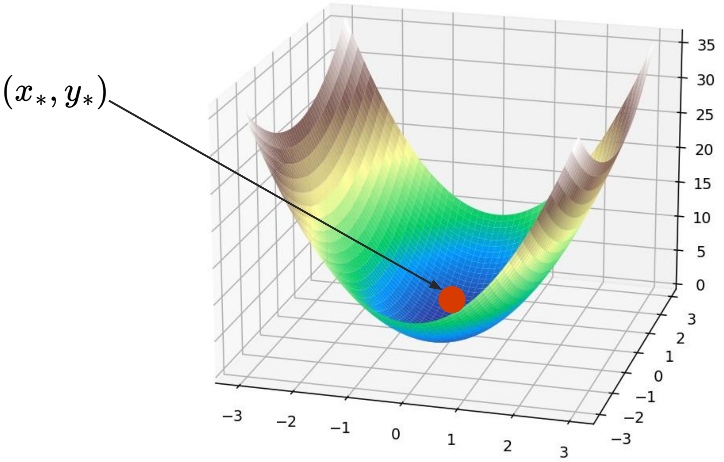
答え：

$$\frac{\mathrm{d}}{\mathrm{dt}} A(t) = \begin{pmatrix} (t^3 + 1)' & (\log t)' \\ (e^t)' & (\sin 2t)' \end{pmatrix} = \begin{pmatrix} 3t^2 & \frac{1}{t} \\ e^t & 2\cos 2t \end{pmatrix}$$

引数がベクトルになる関数の微分

2つのパラメータを持つ損失関数 $f(x, y)$

→極小点 (x_*, y_*) を**勾配法**で求めよう！



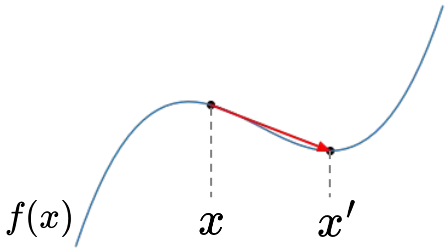
※イメージしやすくするために今回は2変数の場合のみを考えるが、実際の活用では3変数以上であることが多い。

勾配法を2変数に拡張する

1変数の場合の勾配法

$$x' = x - \eta \frac{\mathrm{d}f}{\mathrm{d}x}$$

上式に従って x を更新する。

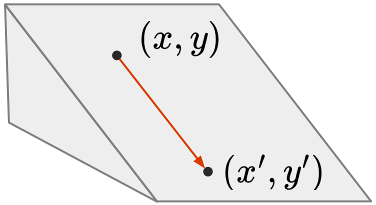


拡張

2変数の場合の勾配法

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

上式に従って (x, y) を更新する。



勾配ベクトルを用いた表記

$\boldsymbol{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ $\boldsymbol{r}' = \begin{pmatrix} x' \\ y' \end{pmatrix}$ $\nabla_{\boldsymbol{r}} f(\boldsymbol{r}) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$ とおくと、次のように書き換えられる

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

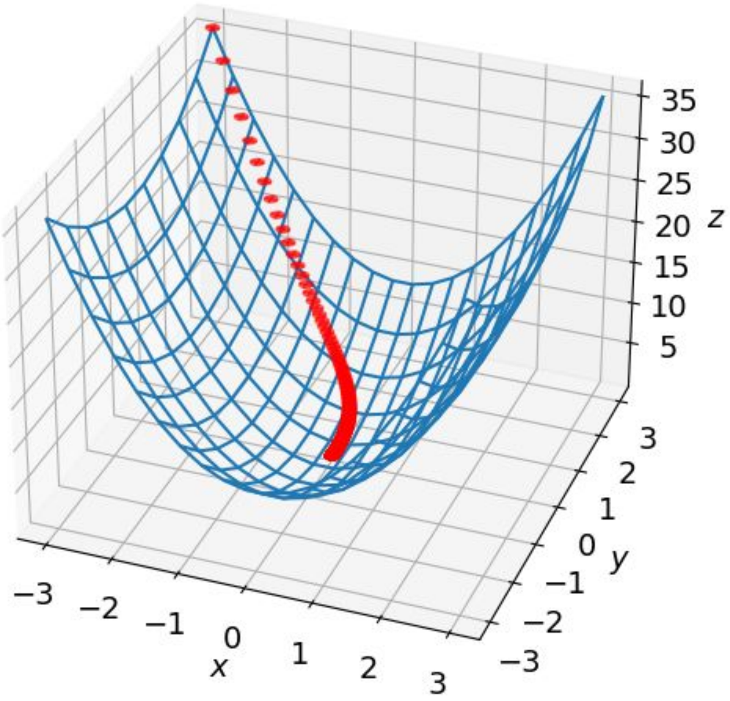
$$\boldsymbol{r}' = \boldsymbol{r} - \eta \nabla_{\boldsymbol{r}} f(\boldsymbol{r})$$

$\nabla_{\boldsymbol{r}} f$ のことを**勾配ベクトル**といい、$\frac{\partial f}{\partial \boldsymbol{r}}$ や $\mathrm{grad} f$ などと書くこともある

→これが「ベクトル**で**微分する」ということ

2変数の勾配法を可視化してみると…

2変数関数で勾配法を適用すると、
図のような軌跡をたどって極小点を得る

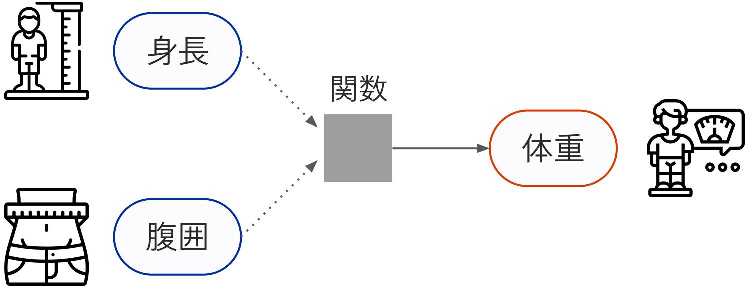


損失関数を用いる例：重回帰分析

身長と腹囲から体重を予測したい

名前	身長	腹囲	体重
児童A	140	57	35
児童B	135	56	32
児童C	142	57	36
児童D	129	55	28
⋮	⋮	⋮	⋮

予測



関数の形

(体重)＝ α×(身長) ＋ β×(腹囲)

αとβの値はわからない

→ どう求めれば良いだろうか？

“誤差”を最小にすることを考える

予測誤差が小さくなるようにしたい

→ (予測した体重) - (実際の体重) の和を
最小化すればよいのでは？？

しかし…右表のような場合では
予測は全く当たっていないのに
誤差の和は **0** になってしまっている！

名前	予測した体重	実際の体重	**誤差**
児童A	0	35	**-35**
児童B	100	32	**+68**
児童C	20	36	**-16**
児童D	11	28	**-17**

合計 : ±0

＋と-で相殺された

誤差の２乗を考えてみると...

正負で相殺されないようにするには...

→ {(予測した体重) - (実際の体重)} ^{2}
の和を考える

これは損失関数の一種
＝ **残差平方和(RSS: Residual Sum of Squares)**

名前	予測した体重	実際の体重	**(誤差)^2**
児童A	0	35	**1225**
児童B	100	32	**4624**
児童C	20	36	**256**
児童D	11	28	**289**

合計 : 6394

残差平方和を最小にするαとβを勾配法で求める

残差平方和(RSS: Residual Sum of Squares)

$$f(\alpha, \beta) = \sum_{i=0}^{N} \{(\alpha x_{\text{児童}i\text{の身長}} + \beta x_{\text{児童}i\text{の腹囲}}) - y_{\text{児童}i\text{の体重}}\}^2$$

→ これを最小にする点(α, β)を勾配法で求めればよい！

$$\begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \eta \begin{pmatrix} \frac{\partial f}{\partial \alpha} \\ \frac{\partial f}{\partial \beta} \end{pmatrix}$$

行列で微分することも考えることができる

関数を行列で微分することもある。

関数 f を行列の各要素で偏微分した $\frac{\partial f}{\partial a_{ij}}$ を並べた行列となる。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \frac{\partial f}{\partial A} = \begin{pmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \frac{\partial f}{\partial a_{13}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \frac{\partial f}{\partial a_{23}} \\ \frac{\partial f}{\partial a_{31}} & \frac{\partial f}{\partial a_{32}} & \frac{\partial f}{\partial a_{33}} \end{pmatrix}$$

例題３.勾配

次の関数の $\boldsymbol{r}$ における勾配を求めよ。

（1） $f(\boldsymbol{r}) = f(x, y) = 3x^2 + y^2$

（2） $g(\boldsymbol{r}) = g(x, y, z) = (x-2)^2 + (y-1)^2 + z^2$

例題３の答え

次の関数の $\boldsymbol{r}$ における勾配を求めよ。

(1) $f(\boldsymbol{r}) = f(x, y) = 3x^2 + y^2$

答え：

$$\nabla_{\boldsymbol{r}} f(\boldsymbol{r}) = \begin{pmatrix} \frac{\partial}{\partial x}(3x^2 + y^2) \\ \frac{\partial}{\partial y}(3x^2 + y^2) \end{pmatrix} = \begin{pmatrix} 6x \\ 2y \end{pmatrix}$$

(2) $g(\boldsymbol{r}) = g(x, y, z) = (x-2)^2 + (y-1)^2 + z^2$

答え：

$$\nabla_{\boldsymbol{r}} g(\boldsymbol{r}) = \begin{pmatrix} \frac{\partial}{\partial x}((x-2)^2 + (y-1)^2 + z^2) \\ \frac{\partial}{\partial y}((x-2)^2 + (y-1)^2 + z^2) \\ \frac{\partial}{\partial z}((x-2)^2 + (y-1)^2 + z^2) \end{pmatrix} = \begin{pmatrix} 2(x-2) \\ 2(y-1) \\ 2z \end{pmatrix}$$

例題４.行列で微分

次の関数 f と行列 A について、$\dfrac{\partial f}{\partial A}$ を求めよ。

$$f(x, y) = 3x^2 + y^2$$

$$A = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$$

例題 4 の答え

次の関数 f と行列 A について、$\dfrac{\partial f}{\partial A}$ を求めよ。

$$f(x, y) = 3x^2 + y^2$$

$$A = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$$

答え：

$$\frac{\partial f}{\partial A} = \begin{pmatrix} \frac{\partial}{\partial x}(3x^2 + y^2) & \frac{\partial}{\partial y}(3x^2 + y^2) \\ \frac{\partial}{\partial y}(3x^2 + y^2) & \frac{\partial}{\partial x}(3x^2 + y^2) \end{pmatrix} = \begin{pmatrix} 6x & 2y \\ 2y & 6x \end{pmatrix}$$

まとめ

01 ベクトルや行列の関数でも、微分を考えられる

02 多変数版の勾配法で、複数のパラメータをもつ関数でも最小値を求められる