機械学習のための線形代数学

山下 智樹

2022 年 6 月 29 日

目次

1 はじめに

このノートでは，機械学習を理解するための線形代数の基礎をまとめておく．はじめに，ここで使用する数学記号の表記方法を記しておく．ベクトルは太字の小文字で $\mathbf{x}$ のように表し，N 次元ベクトルは $N \times 1$ 行列を用いて縦ベクトルで表現する．

$$\mathbf{x} = (x_1, x_2, \cdots, x_N)^{\mathrm{T}} \tag{1}$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}. \tag{2}$$

行列は太字の大文字で $\mathbf{A}$ のように表記する．$N \times M$ 行列であれば

$$
\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{pmatrix}. \tag{3}
$$

2　ベクトルの内積と直積

2.1　内積

ベクトルの内積はスカラー.

$$
\begin{aligned}
\mathbf{x} \cdot \mathbf{y} &= \mathbf{x}^{\mathrm{T}} \mathbf{y} && (4) \\
&= (x_1, x_2, \cdots, x_N) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} && (5) \\
&= \sum_{i=1}^{N} x_i y_i. && (6)
\end{aligned}
$$

2.2　直積

直積は行列. $\mathbf{x}$ が N 次元, $\mathbf{y}$ が M 次元とすると, $N \times M$ 行列になる.

$$
\begin{aligned}
\mathbf{x} \otimes \mathbf{y} &= \mathbf{x} \mathbf{y}^{\mathrm{T}} && (7) \\
&= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} (y_1, y_2, \cdots, y_M) && (8) \\
&= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_M \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_M \\ \vdots & \vdots & \ddots & \vdots \\ x_N y_1 & x_N y_2 & \cdots & x_N y_M \end{pmatrix}. && (9)
\end{aligned}
$$

3　行列とベクトルの積 $\mathbf{Ax}$

$\mathbf{x}$ を N 次元ベクトル, $\mathbf{A}$ を N 次正方行列とすると, $\mathbf{Ax}$ は N 次元ベクトルになり, 第 i 成分は

$$
(\mathbf{Ax})_i = \sum_{j=1}^{N} a_{ij} x_j. \tag{10}
$$

また, $\mathbf{A}^{\mathrm{T}}\mathbf{x}$ の第 i 成分は $\mathbf{A}$ を転置させるだけなので簡単で,

$$
(\mathbf{A}^{\mathrm{T}}\mathbf{x})_i = \sum_{j=1}^{N} a_{ji} x_j. \tag{11}
$$

4　対称行列と反対称行列 (交代行列)

$\mathbf{A}^{\mathrm{T}} = \mathbf{A}$ となる正方行列 $\mathbf{A}$ を対称行列，$\mathbf{A}^{\mathrm{T}} = -\mathbf{A}$ となる正方行列 $\mathbf{A}$ を反対称行列 (交代行列) という．交代行列と呼ぶ方が普通.

あらためて，$\mathbf{A}$ を任意の N 次正方行列とすると，これはいつでも対称行列と反対称行列に分解できる．

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}}) \tag{12}$$

$$= \mathbf{A}^{\mathrm{S}} + \mathbf{A}^{\mathrm{A}}. \tag{13}$$

ここで，$\mathbf{A}^{\mathrm{S}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$ は対称行列，$\mathbf{A}^{\mathrm{A}} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}})$ は反対称行列である．

5　2 次形式

$\mathbf{x}$ を N 次元ベクトル，$\mathbf{A}$ を N 次正方行列とすると，$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ を 2 次形式といい，

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}(\mathbf{A}\mathbf{x}) \tag{14}$$

$$= \sum_{i=1}^{N} x_i (\mathbf{A}\mathbf{x})_i \tag{15}$$

$$= \sum_{i=1}^{N} x_i \sum_{j=1}^{N} a_{ij} x_j \tag{16}$$

$$= \sum_{i,j=1}^{N} a_{ij} x_i x_j \tag{17}$$

と表される．ここでは式 (6) と式 (10) の成分表示を用いた．

ここで反対称行列の 2 次形式 $\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{A}}\mathbf{x}$ を考えると，$\mathbf{A}^{\mathrm{A}}$ の ij 成分と ji 成分は $a_{ij} = -a_{ji}$ の関係にあるので，和をとるとキャンセルして 0 になる．任意の正方行列は $\mathbf{A} = \mathbf{A}^{\mathrm{S}} + \mathbf{A}^{\mathrm{A}}$ のように対称行列と反対称行列に分解できることを思い出すと，結局 2 次形式をとると対称行列の部分しか残らないので，$\mathbf{A}$ を対称行列とみなしてよい．この性質を知っておくと，多変量ガウス分布を理解するのに役に立つ．

6　ベクトルによる微分

機械学習ではベクトル表示したパラメータ $\mathbf{w}$ で何かを微分するような計算が頻繁に出てくるので，いくつかまとめておく．ベクトルで微分するというのは，簡単に言えば勾配をとればいいので結果もベクトルになる．$f(\mathbf{w})$ を $\mathbf{w} = (w_0, w_1, \cdots, w_{M-1})^{\mathrm{T}}$ で微分すると下記のようになる．

$$\nabla f(\mathbf{w}) = \frac{\partial f}{\partial \mathbf{w}} \tag{18}$$

$$= \begin{pmatrix} \frac{\partial f}{\partial w_0} \\ \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_{M-1}} \end{pmatrix}. \tag{19}$$

6.1　内積の微分

内積は式 (6) によると，$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{N} x_i y_i$ のように書けるので，

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) = \begin{pmatrix} \frac{\partial}{\partial x_1}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) \\ \frac{\partial}{\partial x_2}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) \\ \vdots \\ \frac{\partial}{\partial x_N}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) \end{pmatrix} \tag{20}$$

$$= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \tag{21}$$

$$= \mathbf{y}. \tag{22}$$

同様に，

$$\frac{\partial}{\partial \mathbf{y}}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) = \mathbf{x}. \tag{23}$$

6.2　2 次形式の微分

$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ を $\mathbf{x}$ で微分することを考える．勾配の第 k 成分は式 (17) を使うと，

$$\frac{\partial}{\partial x_k}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \frac{\partial}{\partial x_k}\left(\sum_{i,j=1}^{N} a_{ij}x_i x_j\right) \tag{24}$$

$$= \sum_{i,j=1}^{N}\left(a_{ij}\frac{\partial x_i}{\partial x_k}x_j + a_{ij}x_i\frac{\partial x_j}{\partial x_k}\right). \tag{25}$$

ここで，上式右辺括弧の中の第 1 項目は $i = k$ 以外は 0，第 2 項は $j = k$ 以外は 0 になるので，

$$\frac{\partial}{\partial x_k}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \sum_{j=1}^{N} a_{kj}x_j + \sum_{i=1}^{N} a_{ik}x_i \tag{26}$$

$$= (\mathbf{A}\mathbf{x})_k + (\mathbf{A}^{\mathrm{T}}\mathbf{x})_k \tag{27}$$

$$= \left\{(\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{x}\right\}_k. \tag{28}$$

途中の式変形には式 (10) と式 (11) を用いた．従ってベクトル表記では

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right) = \left(\mathbf{A} + \mathbf{A}^{\mathrm{T}}\right)\mathbf{x}. \tag{29}$$

特に $\mathbf{A}$ が対称行列の場合，

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right) = 2\mathbf{A}\mathbf{x}. \tag{30}$$