

機械学習のための線形代数学

山下 智樹

2025 年 9 月 9 日

目次

1	はじめに	1
2	ベクトルの内積と外積	2
2.1	内積	2
2.2	外積	2
3	行列とベクトルの積 Ax	2
4	対称行列と反対称行列 (交代行列)	3
5	2 次形式	3
6	ベクトルによる微分	3
6.1	内積の微分	4
6.2	2 次形式の微分	4

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。

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

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

行列は大文字で A のように表記する。 $N \times M$ 行列であれば

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

2 ベクトルの内積と外積

2.1 内積

ベクトルの内積はスカラー。

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} \quad (4)$$

$$= (x_1 \ x_2 \ \cdots \ x_N)^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad (5)$$

$$= \sum_{i=1}^N x_i y_i. \quad (6)$$

2.2 外積

3次元空間で定義されるクロス積とは異なるので注意。外積は行列になる。\$\mathbf{x}\$ が \$N\$ 次元、\$\mathbf{y}\$ が \$M\$ 次元とすると、\$N \times M\$ 行列になる。

$$\mathbf{xy}^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} (y_1, y_2, \cdots, y_M) \quad (7)$$

$$= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_M \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_M \\ \vdots & \vdots & \ddots & \vdots \\ x_N y_1 & x_N y_2 & \cdots & x_N y_M \end{pmatrix}. \quad (8)$$

3 行列とベクトルの積 \$A\mathbf{x}\$

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

$$(A\mathbf{x})_i = \sum_{j=1}^N a_{ij} x_j. \quad (9)$$

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

$$(A^T \mathbf{x})_i = \sum_{j=1}^N a_{ji} x_j. \quad (10)$$

4 対称行列と反対称行列 (交代行列)

$A^T = A$ となる正方行列 A を対称行列, $A^T = -A$ となる正方行列 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, 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.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) \quad (11)$$

$$= A_S + A_A. \quad (12)$$

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

5 2次形式

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

$$\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T (A \mathbf{x}) \quad (13)$$

$$= \sum_{i=1}^N x_i (A \mathbf{x})_i \quad (14)$$

$$= \sum_{i=1}^N x_i \sum_{j=1}^N a_{ij} x_j \quad (15)$$

$$= \sum_{i,j=1}^N a_{ij} x_i x_j \quad (16)$$

と表される. ここでは式 (6) と式 (9) の成分表示を用いた.

ここで反対称行列の2次形式 $\mathbf{x}^T A_A \mathbf{x}$ を考えると, A_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 $A = A_S + A_A$ 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, 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 $\mathbf{w} = (w_0 \ w_1 \ \cdots \ w_{M-1})^T$ で微分すると下記のようなになる.

$$\nabla f(\mathbf{w}) = \frac{\partial f}{\partial \mathbf{w}} \quad (17)$$

$$= \begin{pmatrix} \frac{\partial f}{\partial w_0} \\ \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_{M-1}} \end{pmatrix}. \quad (18)$$

6.1 内積の微分

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

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{y}) = \begin{pmatrix} \frac{\partial}{\partial x_1} (\mathbf{x}^T \mathbf{y}) \\ \frac{\partial}{\partial x_2} (\mathbf{x}^T \mathbf{y}) \\ \vdots \\ \frac{\partial}{\partial x_N} (\mathbf{x}^T \mathbf{y}) \end{pmatrix} \quad (19)$$

$$= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad (20)$$

$$= \mathbf{y}. \quad (21)$$

同様に,

$$\frac{\partial}{\partial \mathbf{y}} (\mathbf{x}^T \mathbf{y}) = \mathbf{x}. \quad (22)$$

6.2 2次形式の微分

$\mathbf{x}^T \mathbf{A} \mathbf{x}$ を \mathbf{x} で微分することを考える. 勾配の第 k 成分は式 (16) を使うと,

$$\frac{\partial}{\partial x_k} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \frac{\partial}{\partial x_k} \left(\sum_{i,j=1}^N a_{ij} x_i x_j \right) \quad (23)$$

$$= \sum_{i,j=1}^N \left(a_{ij} \frac{\partial x_i}{\partial x_k} x_j + a_{ij} x_i \frac{\partial x_j}{\partial x_k} \right). \quad (24)$$

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,

$$\frac{\partial}{\partial x_k} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \sum_{j=1}^N a_{kj} x_j + \sum_{i=1}^N a_{ik} x_i \quad (25)$$

$$= (\mathbf{A} \mathbf{x})_k + (A^T \mathbf{x})_k \quad (26)$$

$$= \{(A + A^T) \mathbf{x}\}_k. \quad (27)$$

途中の式変形には式 (9) と式 (10) を用いた. 従ってベクトル表記では

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = (A + A^T) \mathbf{x}. \quad (28)$$

特に A が対称行列の場合,

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2A \mathbf{x}. \quad (29)$$