クラシックな機械学習の入門

付録1. 数学の復習

行列の微分
行列式のlogの微分
対称行列の2次形式のtraceへの置き換え
ブロック行列の逆行列（Woodbury)

by 中川裕志（東京大学）

行列の微分

$$\mathbf{x}=\begin{pmatrix}x_1\\ \vdots\\ x_k\end{pmatrix}\quad f(\mathbf{x})\text{はスカラー}\quad \mathbf{f}(\mathbf{x})=\begin{pmatrix}f_1(\mathbf{x})\\ \vdots\\ f_m(\mathbf{x})\end{pmatrix}\quad \mathbf{a}=\begin{pmatrix}a_1\\ \vdots\\ a_k\end{pmatrix}\quad \text{とする。}$$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=\begin{pmatrix}\dfrac{\partial f(\mathbf{x})}{\partial x_1}\\ \vdots\\ \dfrac{\partial f(\mathbf{x})}{\partial x_k}\end{pmatrix}\qquad \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}=\begin{pmatrix}\dfrac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \dfrac{\partial f_m(\mathbf{x})}{\partial x_1}\\ \vdots & \ddots & \vdots\\ \dfrac{\partial f_1(\mathbf{x})}{\partial x_k} & \cdots & \dfrac{\partial f_m(\mathbf{x})}{\partial x_k}\end{pmatrix}\qquad \frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}}=\frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}}=\mathbf{a}$$

$\mathbf{A},\mathbf{B}$を$matrix$とする。

$$\frac{\partial Tr(\mathbf{AB})}{\partial A_{ij}}=B_{ji}\quad\rightarrow\quad\frac{\partial Tr(\mathbf{AB})}{\partial \mathbf{A}}=\mathbf{B}^T$$

行列で微分する場合のchain rule

$$\frac{\partial f(g(\mathbf{x}))}{\partial \mathbf{x}}=\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}\frac{\partial f(g(\mathbf{x}))}{\partial g(\mathbf{x})}$$

行列で微分

$$A=\begin{bmatrix} a_{11} & & a_{1n} \\ & \ddots & \\ a_{m1} & & a_{mn} \end{bmatrix} \Rightarrow \frac{\partial f(A)}{\partial A}=\begin{bmatrix} \dfrac{\partial f(A)}{\partial a_{11}} & & \dfrac{\partial f(A)}{\partial a_{1n}} \\ & \ddots & \\ \dfrac{\partial f(A)}{\partial a_{m1}} & & \dfrac{\partial f(A)}{\partial a_{mn}} \end{bmatrix}$$

行列の積の微分、逆行列の微分

$$\frac{\partial(AB)}{\partial x} = \frac{\partial A}{\partial x} B + A \frac{\partial B}{\partial x}$$

$$\frac{\partial(AB)}{\partial A} = B + A \frac{\partial B}{\partial A} = B(\text{if } A \text{ is independent on } B)$$

$$A^{-1} A = I$$

これをxで微分すると

$$\frac{\partial A^{-1}}{\partial x} A + A^{-1} \frac{\partial A}{\partial x} = 0 \quad \Rightarrow \quad \frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}$$

$$A^{-1} A = I$$

これをAで微分すると

$$\frac{\partial A^{-1}}{\partial A} A + A^{-1} \frac{\partial A}{\partial A} = 0 \quad \Rightarrow \quad \frac{\partial A^{-1}}{\partial A} = -A^{-1} \frac{\partial A}{\partial A} A^{-1}$$

行列式のlogの微分

$$\boxed{\frac{\partial}{\partial x}\log|\mathbf{A}| = Tr\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right)}$$

$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$の場合の例は以下の通り

$$example: \quad \frac{\partial}{\partial x}\log\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = \frac{\partial}{\partial x}\log(ad-bc) = \frac{1}{(ad-bc)}\left(\frac{\partial a}{\partial x}d + \frac{\partial d}{\partial x}a - \frac{\partial b}{\partial x}c - \frac{\partial c}{\partial x}b\right)$$

$$= Tr\left[\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} \frac{\partial a}{\partial x} & \frac{\partial b}{\partial x} \\ \frac{\partial c}{\partial x} & \frac{\partial d}{\partial x} \end{pmatrix}\right] = Tr\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right)$$

$$\boxed{\frac{\partial}{\partial \mathbf{A}}\log|\mathbf{A}| = \left(\mathbf{A}^{-1}\right)^T}$$

$$example: \quad \frac{\partial}{\partial \mathbf{A}}\log|\mathbf{A}| = \frac{1}{(ad-bc)}\begin{pmatrix} \frac{\partial}{\partial a}\log(ad-bc) & \frac{\partial}{\partial b}\log(ad-bc) \\ \frac{\partial}{\partial c}\log(ad-bc) & \frac{\partial}{\partial d}\log(ad-bc) \end{pmatrix}$$

$$= \frac{1}{(ad-bc)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}\right)^T$$

線形代数学の役立つ公式

$$trace \qquad tr \qquad Tr$$

$$trace(AB) = trace(BA)$$

Aが対称行列なら　$\mathbf{x}^T A\mathbf{x} = trace(A\mathbf{x}\mathbf{x}^T)$

共分散行列Σは対象で、正規分布では、$x^T\Sigma x$の計算をすることが多く、そのときには必須。
AICやBICなどの情報量基準の計算ではよく使う。

線形代数学の役立つ公式1

$$|AB| = |A||B| \quad \Rightarrow \quad |A^{-1}| = \frac{1}{|A|} \qquad (A^{-1})^T = (A^T)^{-1}$$

A, Bは$N \times M$行列のとき

$$|I_{N\times N} + AB^T| = |I_{M\times M} + A^T B|$$

$N \times 1$すなわち列ベクトルa, bのとき Matrix Inversion Lemma

$$|I_{N\times N} + ab^T| = 1 + a^T b$$

逆行列を求めるとき役立つ公式

$$(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} = PB^T (BPB^T + R)^{-1}$$

P-1の計算が大変なとき役立つ

special case

$$(I + AB)^{-1} A = A(I + BA)^{-1}$$

D-1の計算が大変なとき役立つ

逆行列を求めるとき役立つ公式：Woodbury identity

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1} CA^{-1}$$

線形代数学の役立つ公式

$$|A| = \sum (\pm 1) A_{1,i1} A_{2,i2} \cdots A_{N,iN}$$

$$\begin{aligned} &if \quad permutaion(i1,..iN) = odd \quad then \quad -1 \\ &\qquad\qquad\qquad\qquad\qquad\quad = \quad even \quad then \quad +1 \end{aligned}$$

$$|AB| = |A||B| \quad \Rightarrow \quad |A^{-1}| = \frac{1}{|A|}$$

線形代数学の役立つ公式
ブロック行列の逆行列

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}+D^{-1}CMBD^{-1} \end{pmatrix}$$

$$where \quad M = \left(A - BD^{-1}C\right)^{-1} \qquad \text{式}(Matrix^{-1})$$

例えば、多次元正規分布の共分散行列やその逆行列（精度行列）を求めるときに必須