

# PRML の 3 章のための数学

サイボウズ・ラボ 光成滋生

2011 年 4 月 15 日

## 1 概要

この文章は『パターン認識と機械学習』(以下 PRML) の 3 章を理解するために必要な数学をまとめてみたものです。

いくつかの定理は証明せずに認めますが, 可能な限り self-contained であることを目指してみました. 概ね PRML に従っていますが, 違う方法をとっているところもあります.

間違い, 質問などございましたら, [herumi@nifty.com](mailto:herumi@nifty.com) または twitterID:herumi までご連絡ください.

## 2 最小二乗法

### 2.1 微分の復習

$x, y$  を縦ベクトルとして

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

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

ここで  $\frac{\partial}{\partial x}$  は  $\frac{\partial}{\partial x_i}$  を縦に並べた縦ベクトルとする. 2 章でも述べたが  $\frac{\partial}{\partial x}$  を  $\nabla$  と書くこともあるが PRML では場所によって縦ベクトル (3.22) だったり, 横ベクトル (3.13) だったりする. 常に縦ベクトルとしたほうが混乱は少ない.

### 2.2 誤差関数の最小化

$$f(w) = \sum_{n=1}^N \{t_n - w^T \phi(x_n)\}^2 + \lambda w^T w$$

とする. ここで  $w$  と  $\phi(x_n)$  は  $M$  次元縦ベクトルである.

$$\Phi^T = (\phi(x_1) \cdots \phi(x_N))$$

とおく.  $\Phi$  は  $N$  行  $M$  列の行列である.  $f(w)$  を  $w$  で微分しよう.

$$\frac{\partial}{\partial \mathbf{w}} f(\mathbf{w}) = 2 \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\} (-\phi(\mathbf{x}_n)) + 2\lambda \mathbf{w}.$$

一般に縦ベクトル  $\mathbf{x}$ ,  $\mathbf{y}$  に対して

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

だから  $\mathbf{t} = (t_1, \dots, t_N)^T$  とおくと

$$\begin{aligned} \frac{1}{2} \frac{\partial}{\partial \mathbf{w}} f(\mathbf{w}) &= - \sum_n t_n \phi(\mathbf{x}_n) + \sum_n (\phi(\mathbf{x}_n) \phi(\mathbf{x}_n)^T) \mathbf{w} + \lambda \mathbf{w} \\ &= -\Phi^T \mathbf{t} + \Phi^T \Phi \mathbf{w} + \lambda \mathbf{w} \\ &= -\Phi^T \mathbf{t} + (\Phi^T \Phi + \lambda I) \mathbf{w} = 0. \end{aligned}$$

よって  $\det(\lambda I + \Phi^T \Phi) \neq 0$  のとき

$$\mathbf{w}_{\text{ML}} = (\lambda I + \Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$

が最尤解.  $\mathbf{y} = \Phi \mathbf{w}$  が予測値である.

## 2.3 正射影

前節で  $\lambda = 0$  のときを考える.

$$\mathbf{y} = \Phi (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$

となる. ここでこの式の幾何学的な解釈を考えてみよう.

$\Phi = (a_1 \cdots a_M)$  と縦ベクトルの集まりで表す.  $N - M$  個のベクトル  $b_1, \dots, b_{N-M}$  を追加して,  $\{a_1, \dots, a_M, b_1, \dots, b_{N-M}\}$  全体で  $N$  次元ベクトル空間の基底であるようにとる. その際  $b_i$  を  $a_j$  と直交するようにとれる.

$$a_i^T b_j = 0.$$

さて  $X = \Phi (\Phi^T \Phi)^{-1} \Phi^T$  とおくと,  $X \Phi = \Phi$ . これは  $X a_i = a_i$  を意味する. つまり  $X$  は  $a_1, \dots, a_M$  で生成される部分空間  $V = \langle a_1, \dots, a_M \rangle$  の点を動かさない. また  $b_j$  のとりかたから  $X b_j = 0$  も成り立つ. つまり  $X$  は部分空間  $\langle b_1, \dots, b_{N-M} \rangle$  の点を 0 につづす.

二つ合わせると,  $X$  は任意の点を部分空間  $V$  方向につづす写像, つまり  $V$  への正射影写像と解釈できる.

式で書くと任意の点  $\mathbf{t}$  を  $\mathbf{t} = \sum_i s_i a_i + \sum_i t_i b_j$  と表したとすると,

$$\mathbf{y} = X \mathbf{t} = \sum_i s_i a_i$$

となる.  $\mathbf{t}$  から  $\mathbf{y}$  への変換を係数だけを使って書いてみると

$$X : (s_1, \dots, s_N, t_1, \dots, t_{N-M}) \rightarrow (s_1, \dots, s_N, 0, \dots, 0).$$

これを見ると正射影のニュアンスがより明確になる.

## 2.4 行列での微分

$x$  を  $n$  次元ベクトル,  $A$  を  $m$  行  $n$  列として  $y = Ax$  とおく.

$$f(A) = \|y\|^2 = (Ax)^T Ax$$

を  $A$  で微分してみよう.

$$(Ax)^T Ax = \sum_s (Ax)_s (Ax)_s = \sum_s \left( \sum_t a_{st} x_t \right) \left( \sum_u a_{su} x_u \right) = \sum_{s,t,u} x_t x_u a_{st} a_{su}.$$

よって

$$\begin{aligned} \frac{\partial}{\partial a_{ij}} f(A) &= \sum_{s,t,u} x_t x_u \left( \left( \frac{\partial}{\partial a_{ij}} a_{st} \right) a_{su} + a_{st} \frac{\partial}{\partial a_{ij}} a_{su} \right) \\ &= \sum_{s,t,u} x_t x_u (\delta_{is} \delta_{jt} a_{su} + a_{st} \delta_{is} \delta_{ju}) \\ &= \left( \sum_u x_j x_u a_{iu} \right) + \left( \sum_t x_t x_j a_{it} \right) \\ &= 2 \sum_u x_j x_u a_{iu} \\ &= 2x_j (Ax)_i \\ &= 2(Axx^T)_{ij}. \end{aligned}$$

よって

$$\frac{\partial}{\partial A} \|Ax\|^2 = 2Axx^T.$$

## 2.5 Woodbury の逆行列の公式

行列  $A, B, C, D$  について

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

が成り立つ.

(証明)

$$\begin{aligned} A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1} &= A^{-1}B((DB^{-1}A + C)A^{-1}B)^{-1}CA^{-1} \\ &= (DB^{-1}A + C)^{-1}CA^{-1} \\ &= (DB^{-1}(A + BD^{-1}C))^{-1}CA^{-1} \\ &= (A + BD^{-1}C)^{-1}BD^{-1}CA^{-1} \end{aligned}$$

よって

$$\begin{aligned} \text{左辺} &= (I - (A + BD^{-1}C)^{-1}BD^{-1}C)A^{-1} \\ &= (A + BD^{-1}C)^{-1}((A + BD^{-1}C) - BD^{-1}C)A^{-1} \\ &= \text{右辺}. \end{aligned}$$

特に,  $A$  が  $n$  次正方行列で  $B$  を  $n$  次縦ベクトル  $x$ ,  $C = x^T$ ,  $D$  を  $n$  次単位行列とすると

$$(A + xx^T)^{-1} = A^{-1} - \frac{(A^{-1}x)(x^T A^{-1})}{1 + x^T A^{-1}x} \quad (1)$$

が成り立つ.

## 2.6 正定値対称行列

$n$  次元実対称行列  $A$  はある直行列  $P$  を用いて常に対角化可能であった.

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n).$$

全ての固有値が正であるとき  $A$  を正定値といい,  $A > 0$  とかく. 全ての固有値が正または 0 であるとき, 半正定値といい,  $A \geq 0$  とかく.

任意の実ベクトル  $x$  について  $y = Px$  とおくと  $x$  が  $\mathbb{R}^n$  の全ての点をとるとき  $y$  も全ての点を渡る.

$$x^T Ax = \sum_i \lambda_i y_i^2$$

なので  $A \geq 0$  ならば  $x^T Ax \geq 0$ .  $A > 0$  のときは等号が成り立つのは  $x = 0$  のときのみである.

逆に任意の  $x$  について  $x^T Ax \geq 0$  とすると,  $y$  として単位ベクトル  $e_i$  を考えれば  $\lambda_i \geq 0$ . つまり  $A \geq 0$ .

更に等号は  $x = 0$  のときに限るためには  $\lambda_i > 0$ . つまり  $A > 0$  であることが分かる. まとめると

$$A \geq 0 \iff x^T Ax \geq 0. \text{ for } \forall x \in \mathbb{R}$$

$$A > 0 \iff \text{上記不等式が成り立ち, 更に等号は } x = 0 \text{ のときに限る.}$$

この同値性から  $A > 0$ ,  $B > 0$  のとき  $A + B > 0$ . また  $A^{-1} > 0$  も分かる.

また実ベクトル  $v$  に対して  $A = vv^T$  とおくと,  $A$  は実対称であり, 任意の  $x$  に対して

$$x^T Ax = (v^T x)^2 \geq 0$$

なので  $A \geq 0$ .

## 2.7 予測分布の分散

$S_N^{-1} = S_0^{-1} + \beta \Phi_N^T \Phi_N$  としたときの予測分布の分散

$$\sigma_N^2 = \frac{1}{\beta} + \phi^T S_N \phi$$

を考える.  $\beta > 0$  であり,  $S_0$  は共分散行列なので実正定値であることに注意する. まず計画行列  $\Phi_N$  は  $N$  が一つ増える毎に 1 行増える.  $v$  を  $M$  次元縦ベクトルとして

$$\Phi_{N+1}^T = (\Phi_N^T v)$$

としよう. すると

$$S_{N+1}^{-1} = S_0^{-1} + \beta(\Phi_N^T \Phi_N + vv^T) = S_N^{-1} + \beta vv^T.$$

行列  $A = \beta vv^T$  は正定値であり,  $S_N$  に関して帰納法を使うと全ての  $S_N$  は正定値であることが分かる.

公式 1 を使って

$$\begin{aligned}\sigma_{N+1}^2 &= \frac{1}{\beta} + \phi^T (S_N^{-1} + \beta v v^T)^{-1} \phi \\ &= \frac{1}{\beta} + \phi^T \left( S_N - \frac{(S_N v)(v^T S_N)}{1 + v^T S_N v} \right) \phi \\ &= \sigma_N^2 - z\end{aligned}$$

ここで  $S_N$  は対称なので

$$\begin{aligned}z &= \phi^T \frac{(S_N v)(v^T S_N)}{1 + v^T S_N v} \phi \\ &= \frac{1}{1 + v^T S_N v} (v^T S_N \phi)^2.\end{aligned}$$

$S_N$  は正定値なので任意の  $v$  に対して  $v^T S_N v \geq 0$ . よって  $z \geq 0$  となり

$$\sigma_{N+1}^2 \leq \sigma_N^2.$$

帰納法の流れを見ると,

$$\Phi_N^T = (\phi(x_1) \cdots \phi(x_N))$$

とおくと

$$S_N^{-1} = S_0^{-1} + \beta \sum_{i=1}^N \phi(x_i) \phi(x_i)^T$$

となることがわかる.