7.18

ノートのタイトル 2016/03/03

(7.109) $\ln p(\mathbf{w}|\mathbf{t},\alpha) = \sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\} - \frac{1}{2}\mathbf{w}^T A \mathbf{w} + \text{const}$

を $\mathbf{w}$ で微分する。

ここで

$$\nabla_w y_n = \nabla_w \sigma(\mathbf{w}^T\phi(x_n)) = \sigma(\mathbf{w}^T\phi(x_n))\{1-\sigma(\mathbf{w}^T\phi(x_n))\}\phi(x_n) = y_n(1-y_n)\phi(x_n)$$

(7.108)　(4.88) $\sigma(a)' = \sigma(a)(1-\sigma(a))$　(C.19)より $\nabla_w \mathbf{w}^T\phi(x_n) = \phi(x_n)$

$\nabla_w = \frac{\partial}{\partial \mathbf{w}} = \begin{pmatrix}\frac{\partial}{\partial w_1}\\ \frac{\partial}{\partial w_2}\end{pmatrix}$ である。

これを使って

$$\nabla_w \ln y_n = \frac{1}{y_n}\nabla_w y_n = (1-y_n)\phi(x_n)$$

また

$$\nabla_w \ln(1-y_n) = \frac{1}{1-y_n}(-1)\nabla_w y_n = -y_n\phi(x_n)$$

さらに

$$\nabla_w \mathbf{w}^T A \mathbf{w} = 2A\mathbf{w}$$

成分計算で確認できる

$\frac{\partial}{\partial x} x^T A x = (A + A^T)x = 2Ax$

Aが対称のとき

よって

$$\nabla_w \ln p(\mathbf{w}|\mathbf{t},\alpha) = \sum_{n=1}^{N}\{t_n(1-y_n)\phi(x_n) + (1-t_n)(-y_n)\phi(x_n)\} - A\mathbf{w}$$

$$= \sum_{n=1}^{N}(t_n - y_n)\phi(x_n) - A\mathbf{w}$$

$$= \Phi^T \mathbf{y} - A\mathbf{w} \quad \cdots (7.110)$$

$\Phi = \begin{pmatrix}\phi(x_1)^T\\ \phi(x_2)^T\end{pmatrix}$　Block matrixの積

$\because \Phi^T(\mathbf{t}-\mathbf{y}) = (\phi(x_1)\ \phi(x_2))\begin{pmatrix}t_1-y_1\\ t_2-y_2\end{pmatrix} = \phi(x_1)(t_1-y_1) + \phi(x_2)(t_2-y_2)$

$= \sum \phi(x_n)(t_n - y_n)$

を得る

ここで　ベクトルをベクトルで微分の定義

$$\nabla_w y_n\phi(x_n) = \begin{pmatrix}\frac{\partial}{\partial w_1}y_n\phi_1(x_n) & \frac{\partial}{\partial w_1}y_n\phi_2(x_n)\\ \frac{\partial}{\partial w_2}y_n\phi_1(x_n) & \frac{\partial}{\partial w_2}y_n\phi_2(x_n)\end{pmatrix} = (\nabla_w y_n)\phi(x_n)^T = y_n(1-y_n)\phi(x_n)\phi(x_n)^T$$

また　ベクトルをベクトルで微分の定義　A対称なので

$$\nabla_w A\mathbf{w} = \nabla_w\begin{pmatrix}a_{11}w_1 + a_{12}w_2\\ a_{21}w_1 + a_{22}w_2\end{pmatrix} = \begin{pmatrix}\frac{\partial}{\partial w_1}(a_{11}w_1 + a_{12}w_2) & \frac{\partial}{\partial w_1}(a_{21}w_1 + a_{22}w_2)\\ \frac{\partial}{\partial w_2}(a_{11}w_1 + a_{12}w_2) & \frac{\partial}{\partial w_2}(a_{21}w_1 + a_{22}w_2)\end{pmatrix} = \begin{pmatrix}a_{11} & a_{21}\\ a_{12} & a_{22}\end{pmatrix} = A^T = A$$

これらを使って

$\nabla_w\nabla_w \ln p(\mathbf{w}|\mathbf{t},\alpha)$

$$= \sum_{n=1}^{N}(-1)y_n(1-y_n)\phi(x_n)\phi(x_n)^T - A$$

$$= -\left\{\sum_{n=1}^{N}y_n(1-y_n)\phi(x_n)\phi(x_n)^T + A\right\}$$

$$= -(\Phi^T B \Phi + A) \cdots (7.111)$$

を得る。

$B = \begin{pmatrix}y_1(1-y_1) & 0\\ 0 & y_2(1-y_2)\end{pmatrix}$

$$\Phi^T B \Phi = \Phi^T\begin{pmatrix}y_1(1-y_1) & 0\\ 0 & y_2(1-y_2)\end{pmatrix}\Phi$$

$$= \Phi^T\left\{\begin{pmatrix}y_1(1-y_1) & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & y_2(1-y_2)\end{pmatrix}\right\}\Phi$$

$$= y_1(1-y_1)\Phi^T\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Phi + y_2(1-y_2)\Phi^T\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\Phi$$

$$= \sum_{n=1}^{N}y_n(1-y_n)\phi(x_n)\phi(x_n)^T$$

$$\Phi^T\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Phi$$

$$= (\phi(x_1)\ \phi(x_2))\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\phi(x_1)^T\\ \phi(x_2)^T\end{pmatrix}$$

Block matrixの積

$$= (\phi(x_1)\ \phi(x_2))\begin{pmatrix}\phi(x_1)^T\\ 0\end{pmatrix}$$

Block matrixの積

$$= \phi(x_1)\phi(x_1)^T + \phi(x_2)\times 0$$

$$= \phi(x_1)\phi(x_1)^T$$