PRML 演習問題3.9 解答

演習問題3.9

問題は、データ $(\boldsymbol{x}_n$,$\mathrm{t}_n)$ がN個得られた時のパラメータ$\boldsymbol{w}$に関する事後分布

$$p(\boldsymbol{w}|\boldsymbol{t}) = N(\boldsymbol{w}|\boldsymbol{m}_N, \boldsymbol{S}_N) \tag{1}$$

に関して、新たに $(\boldsymbol{x}_{N+1}, t_{N+1})$ が得られた際、$\boldsymbol{m}_{N+1}, \boldsymbol{S}_{N+1}$ と導ける事を、(2.116)式を利用して示せ、というもの。ただし、N個までのデータから導けた事後分布をN+1個目のデータが得られた際の事前分布として利用する。またこの時の $\boldsymbol{m}_N$ と $\boldsymbol{S}_N$ はそれぞれ

$$\begin{aligned} \boldsymbol{m}_N &= \boldsymbol{S}_N(\boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \beta\,\boldsymbol{\Phi}^T\boldsymbol{t}) & (2)\\ \boldsymbol{S}_N^{-1} &= \boldsymbol{S}_0^{-1} + \beta\,\boldsymbol{\Phi}^T\boldsymbol{\Phi} & (3) \end{aligned}$$

となっている。
データがN+1個 ($\boldsymbol{t}_{(N+1)}$ とする) 得られた時のパラメータ$\boldsymbol{w}$の事後分布 $p(\boldsymbol{w}|\boldsymbol{t}_{(N+1)})$ は、tの尤度関数 (3.10) を考慮すると

$$\begin{aligned} p(\boldsymbol{w}|\boldsymbol{t}_{(N+1)}) &= p(\boldsymbol{t}_{(N+1)}|\boldsymbol{w})p(\boldsymbol{w})\\ &= \prod_{n=1}^{N+1} N(t_n|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_n),\,\beta^{-1})N(\boldsymbol{w}|\boldsymbol{m}_0, \boldsymbol{S}_0)\\ &= N(t_{N+1}|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_{N+1}),\,\beta^{-1})\prod_{n=1}^{N} N(t_n|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_n),\,\beta^{-1})N(\boldsymbol{w}|\boldsymbol{m}_0, \boldsymbol{S}_0)\\ &= N(t_{N+1}|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_{N+1}),\,\beta^{-1})p(\boldsymbol{w}|\boldsymbol{t})\\ &= N(t_{N+1}|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_{N+1}),\,\beta^{-1})N(\boldsymbol{w}|\boldsymbol{m}_N, \boldsymbol{S}_N) & (4) \end{aligned}$$

となる。これは、$N(\boldsymbol{w}|\boldsymbol{m}_N, \boldsymbol{S}_N)$ を事前分布と考えると、t_{N+1} の条件付き確率 $p(\boldsymbol{w}|t_{N+1})$ とみなせる。すると、tはwの条件付き確率である事から、(2.113),(2.114)におけるx,yをx→w, y→tとすることで(2.113)〜(2.116)の式を利用する事ができる。この場合、(3.8),(3.3)式と(1)から

$$\begin{aligned} p(\boldsymbol{w}) &= N(\boldsymbol{w}|\boldsymbol{m}_N, \boldsymbol{S}_N)\\ p(t_{N+1}|\boldsymbol{w}) &= N(t_{N+1}|\boldsymbol{w}^T\phi\,(\boldsymbol{x}_{N+1}),\,\beta^{-1}) \end{aligned} \tag{5}$$

となり、(2.113),(2.114)との対応を見ると、

$$
\begin{aligned}
\mu &= \boldsymbol{m}_N \\
\Lambda^{-1} &= \boldsymbol{S}_N \\
A &= \phi(\boldsymbol{x})^T \\
b &= 0 \\
L &= \beta
\end{aligned}
$$

これを、(2.116) にあてはめると

$$
\begin{aligned}
p(\boldsymbol{w}|\boldsymbol{t}_{(N+1)}) = p(\boldsymbol{w}|t_{N+1}) &= N(\boldsymbol{w}|\ \Sigma\ \{\ \phi(\boldsymbol{x}_{N+1})\ \beta\ t_{N+1} + \boldsymbol{S}_N^{-1}\boldsymbol{m}_N\}, \Sigma\ ) \\
&= N(\boldsymbol{w}|\ \Sigma\ \{\ \beta\ \phi(\boldsymbol{x}_{N+1}) t_{N+1} + \boldsymbol{S}_N^{-1}\boldsymbol{m}_N\}, \Sigma\ ) \quad (6) \\
\Sigma &= \left(\boldsymbol{S}_N^{-1} + \phi(\boldsymbol{x}_{N+1})\ \beta\ \phi(\boldsymbol{x}_{N+1})^T\right)^{-1} \\
&= \left(\boldsymbol{S}_N^{-1} + \beta\ \phi(\boldsymbol{x}_{N+1})\ \phi(\boldsymbol{x}_{N+1})^T\right)^{-1} \quad (7)
\end{aligned}
$$

まず Σ から見ると、(7) 式は (3) 式を代入すると

$$
\Sigma = \left(\boldsymbol{S}_0^{-1} + \beta\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \phi(\boldsymbol{x}_{N+1})\ \phi(\boldsymbol{x}_{N+1})^T\right)\right)^{-1} \quad (8)
$$

となる。ここで $\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \phi(\boldsymbol{x}_{N+1})\ \phi(\boldsymbol{x}_{N+1})^T$ は (3.16) 式から $\boldsymbol{\Phi}^T_{(N+1)}\boldsymbol{\Phi}_{(N+1)}$ と表せる事がわかる。ただし、$\boldsymbol{\Phi}_{(N+1)}$ は $\boldsymbol{\Phi}$ を N+1 行 N+1 列に拡張したものである。すると、(7) 式は

$$
\begin{aligned}
\Sigma &= \left(\boldsymbol{S}_0^{-1} + \beta\ \boldsymbol{\Phi}^T_{(N+1)}\boldsymbol{\Phi}_{(N+1)}\right)^{-1} \\
&= \boldsymbol{S}_{N+1} \quad (9)
\end{aligned}
$$

となる。
続いて (6) の平均について見てみると、(2) と (9) 式を代入すると、

$$
\begin{aligned}
\Sigma\ \{\ \beta\ \phi(\boldsymbol{x}_{N+1}) t_{N+1} + \boldsymbol{S}_N^{-1}\boldsymbol{m}_N\} &= \boldsymbol{S}_{N+1}\{\ \beta\ \phi(\boldsymbol{x}_{N+1}) t_{N+1} + \boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \beta\ \boldsymbol{\Phi}^T\boldsymbol{t}\} \\
&= \boldsymbol{S}_{N+1}\{\boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \beta\left(\boldsymbol{\Phi}^T\boldsymbol{t} + \phi(\boldsymbol{x}_{N+1}) t_{N+1}\right)\} \quad (10)
\end{aligned}
$$

ここで、$\boldsymbol{\Phi}^T\boldsymbol{t} + \phi(\boldsymbol{x}_{N+1}) t_{N+1}$ は (3.16) 式から $\boldsymbol{\Phi}^T_{(N+1)}\boldsymbol{t}_{(N+1)}$ と表せる事がわかる。すると (10) 式は、

$$
\begin{aligned}
\Sigma\ \{\ \beta\ \phi(\boldsymbol{x}_{N+1}) t_{N+1} + \boldsymbol{S}_N^{-1}\boldsymbol{m}_N\} &= \boldsymbol{S}_{N+1}\{\boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \beta\ \boldsymbol{\Phi}^T_{(N+1)}\boldsymbol{t}_{(N+1)}\} \\
&= \boldsymbol{m}_{N+1} \quad (11)
\end{aligned}
$$

となる。
以上より、(6),(9),(11) 式から、

$$
p(\boldsymbol{w}|\boldsymbol{t}_{(N+1)}) = N(\boldsymbol{w}|\boldsymbol{m}_{N+1}, \boldsymbol{S}_{N+1})
$$

が導ける。