PRML 解答集

2022 年 4 月 17 日

問題 3.3

それぞれのデータ点 t_n に重み要素 $r_n > 0$ が割り当てられており，二乗和誤差関数が

$$E_D(w) = \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \boldsymbol{w}^T\phi(\boldsymbol{x}_n)\}^2 \tag{3.104}$$

となるデータ集合を考える．このとき，この誤差関数を最小にする解 $\boldsymbol{w}^*$ についての式を求めよ．また，(i) ノイズの分散がデータに依存する場合，(ii) データ点に重複がある場合に照らして，それぞれ重み付き二乗和誤差関数の解釈を与えよ。

方針 3.3

解答 3.3

$\boldsymbol{\phi} = (\phi(\boldsymbol{x_1}), \phi(\boldsymbol{x_2}), \dots, \phi(\boldsymbol{x_n}))^T$, $\boldsymbol{D} = diag(r_1, \dots, r_n)$ とすると

$$E_D(w) = \frac{1}{2}(\boldsymbol{t} - \boldsymbol{\phi w})^T\boldsymbol{D}(\boldsymbol{t} - \boldsymbol{\phi w})$$

と表せる．上式を $\boldsymbol{w}$ で微分すると

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}}E_D(\boldsymbol{w}) &= \frac{\partial}{\partial \boldsymbol{w}}\left\{\frac{1}{2}(\boldsymbol{t}^T - \boldsymbol{w}^T\boldsymbol{\phi}^T)\boldsymbol{D}(\boldsymbol{t} - \boldsymbol{\phi w})\right\} \\
&= \frac{\partial}{\partial \boldsymbol{w}}\left\{\frac{1}{2}(\boldsymbol{t}^T\boldsymbol{Dt} - \boldsymbol{t}^T\boldsymbol{D\phi w} - \boldsymbol{w}^T\boldsymbol{\phi}^T\boldsymbol{Dt} + \boldsymbol{w}^T\boldsymbol{\phi}^T\boldsymbol{D\phi w})\right\} \\
&= \boldsymbol{\phi}^T\boldsymbol{D\phi w} - \boldsymbol{\phi}^T\boldsymbol{Dt}
\end{aligned}$$

なお，ここでは以下の微分公式を利用.

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{a}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}^T\boldsymbol{x}) = \boldsymbol{a} \tag{C.19}$$

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{Ax}) = 2\boldsymbol{Ax}(A\text{ は対角行列})$$

$\dfrac{\partial}{\partial \boldsymbol{w}}E_D(\boldsymbol{w}) = 0$ を満たす $\boldsymbol{w}$ が誤差関数を最小にする $\boldsymbol{w}^*$ であり，計算すると

$$\boldsymbol{w}^* = (\boldsymbol{\phi}^T \boldsymbol{D} \boldsymbol{\phi})^{-1} \boldsymbol{\phi}^T \boldsymbol{D} \boldsymbol{t}$$

となる.

誤差関数の解釈

(i) ノイズの分散がデータに依存する場合

(3.10) にて、右辺の β が n に依存しているということなので、

$$P(\boldsymbol{t}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\beta}) = \prod_{n=1}^{N} N(t_n|\boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x_n}), \beta_n^{-1})$$

$$\begin{aligned}\log P(\boldsymbol{t}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\beta}) &= \sum_{n=1}^{N} \log \exp \left\{ -\frac{1}{2} \beta_n (t_n - \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x_n}))^2 \right\} + (w \text{ に依らない項}) \\ &= \frac{1}{2} \sum_{n=1}^{N} \beta_n (t_n - \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x_n}))^2 + const\end{aligned}$$

となり，(3.104) の誤差関数が出現する.

(ii) データ点の重複がある場合

(3.104) の式の形から，ある n についてのデータ $(t_n, \boldsymbol{x_n})$ が r_n 個ずつある時の二乗和誤差を表していることが分かる.

問題 3.6

ガウス分布に従う複数の目標変数 $\boldsymbol{t}$ を持つ次の形の線形基底関数モデルを考える．

$$p(\boldsymbol{t}|\boldsymbol{W},\boldsymbol{\Sigma}) = N(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{W},\boldsymbol{\Sigma}) \tag{3.107}$$

ただし，

$$y(\boldsymbol{x},\boldsymbol{W}) = \boldsymbol{W}^T\phi(\boldsymbol{x}) \tag{3.108}$$

である．入力基底ベクトル $\phi(\boldsymbol{x}_n)(n=1,\ldots,N)$ とそれに対応する目標ベクトル $\boldsymbol{t}_n$ が訓練データ集合として与えられるとき，パラメータ行列 $\boldsymbol{W}$ の最尤推定解 $\boldsymbol{W}_{ML}$ のそれぞれの列が，等方性のノイズ分布に対する解の (3.15) の形の式で与えられることを示せ．これは共分散行列 $\boldsymbol{\Sigma}$ にはよらないことに注意せよ．さらに，$\boldsymbol{\Sigma}$ の最尤指定解が

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(t_n - \boldsymbol{W}_{ML}\phi(\boldsymbol{x}_n))(t_n - \boldsymbol{W}_{ML}\phi(\boldsymbol{x}_n))^T \tag{3.109}$$

で与えられることを示せ．

方針 3.6

解答 3.6

X: 入力ベクトル $x_1,\ldots x_n$ を並べた行列

とする．対数尤度関数は

$$\ln P(T|X,W,\Sigma) = \sum_{n=1}^{N}\ln N(t_n|W^T\phi(x_n),\Sigma)$$

$$= -\frac{NK}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi)^T\Sigma^{-1}(t_n - w^T\phi_n)$$

となる．

$$\frac{\partial}{\partial X}(Xb+c)^TD(Xb+c) = (D+D^T)(Xb+c)b^T$$

ここで,T を n 番目の行が t_n^T となる $N\times K$ 行列とし，上記の公式を用いて W^T に関する微分計算をすると

$$\frac{\partial}{\partial W^T}\ln P(T|X,W,\Sigma) = \sum_{n=1}^{N}\Sigma^{-1}(t_n - w^T\phi_n)\phi_n^T$$

これを $\mathbf{0}$ とおくと,

$$\begin{aligned}\sum_{n=1}^{N}\Sigma^{-1}(t_n - w^T\phi_n)\phi_n^T = \mathbf{0}\\ \Leftrightarrow W^T\sum_{n=1}^{N}\phi_n\phi_n^T = \sum_{n=1}^{N}t_n\phi_n^T\\ \Leftrightarrow W^T\Phi^T\Phi = T^T\Phi\end{aligned}$$

よって,

$$\begin{aligned}W_{ML}^T = T^T\Phi(\Phi^T\Phi)^{-1}\\ W_{ML} = (\Phi^T\Phi)^{-1}\Phi^T T\end{aligned}$$

となり, T は n 番目の行が t_n^T となるようにした行列であるから, W_{ML} のそれぞれの列が, 等方性のノイズ分布に対する解の (3.15) の形の式で与えられることが分かる.
また Σ の最尤解は, 対数尤度関数を Σ で微分して $\mathbf{0}$ と置き, Σ について解くことで求まる.
これは演習 2.34(以前解いたもの) と全く同手順であるため省略する.

問題 3.7

$\boldsymbol{m}_N$ と $\boldsymbol{S}_N$ がそれぞれ (3.50) と (3.51) で定義される線形基底関数モデルを考える．平方完成を用いて，このモデルパラメータ $\boldsymbol{w}$ の事後分布が (3.49) で与えられることを確かめよ．

方針 3.7

解答 3.7

ベイズの公式から

$$p(w|t) \propto p(t|w)p(w)$$

(3.10),(3.48) より，

$$\begin{aligned}
&\propto \left(\prod_{n=1}^{N} exp\left[-\frac{1}{2}\left\{t_n - w^T\phi(x_n)\right\}^2\beta\right]\right) exp\left[-\frac{1}{2}(w-m_0)^T S_0^{-1}(w-m_0)\right] \\
&= exp\left[-\frac{1}{2}\left\{\beta\sum_{n=1}^{N}\left(t_n - w^T\phi(x_n)\right)^2 + (w-m_0)^T S_0^{-1}(w-m_0)\right\}\right]
\end{aligned}$$

以下では exp の中の $-\frac{1}{2}$ の中身を見ていく．

$$= \beta \begin{pmatrix} t_1 - w^T\phi(x_1) \\ \vdots \\ t_N - w^T\phi(x_N) \end{pmatrix}^T \begin{pmatrix} t_1 - w^T\phi(x_1) \\ \vdots \\ t_N - w^T\phi(x_N) \end{pmatrix} + (w-m_0)^T S_0^{-1}(w-m_0)$$

$w^T\phi(x_1) = \phi^T(x_1)w$ なので

$$= \beta \begin{pmatrix} t_1 - \phi^T(x_1)w \\ \vdots \\ t_N - \phi^T(x_N)w \end{pmatrix}^T \begin{pmatrix} t_1 - \phi^T(x_1)w \\ \vdots \\ t_N - \phi^T(x_N)w \end{pmatrix} + (w-m_0)^T S_0^{-1}(w-m_0)$$

ここで,

$$\Phi = \begin{pmatrix} \phi^T(x_1) \\ \vdots \\ \phi^T(x_N) \end{pmatrix}$$

であるので, 定数項を C とおくとして

$$\begin{aligned}
&= \beta(t-\Phi w)^T(t-\Phi w) + (w-m_0)^T S_0^{-1}(w-m_0) \\
&= \beta(w^T\Phi^T\Phi w - w^T\Phi^T t - t^T\Phi w) + w^T S_0^{-1}w - w^T S_0^{-1}m_0 - m_0^T S_0^{-1}w + C
\end{aligned}$$

S_0 は共分散なので対称行列となる．よってその逆行列も対称行列なので $(S_0^{-1})^T = S_0^{-1}$.

$$= w^T(S_0^{-1} + \beta\Phi^T\Phi)w - w^T(S_0^{-1}m_0 + \beta\Phi^T t) - (S_0^{-1}m_0 + \beta\Phi^T t)^T w + C$$

ここで, m_N, S_N はそれぞれ以下のように与えられている.

$$\begin{aligned} m_N = S_N(S_0^{-1}m_0 + \beta\Phi^T t) = S_N R \\ S_N^{-1} = S_0^{-1} + \beta\Phi^T\Phi \end{aligned}$$

これらを用いて平方完成すると

$$\begin{aligned} = w^T S_N^{-1} w - w^T S_N^{-1} m_N - S_N^{-1} m_N w + C \\ = (w - m_N)^T S^{-1} (w - m_N) + C' \end{aligned}$$

なおパラメータ w に依存しない部分を新たに定数 C' と置いた.
以上より, C を定数として

$$p(w|t) \propto C \exp\{(w - m_N)^T S^{-1}(w - m_N)\}$$

と表せるので, モデルパラメータ w の事後分布が (3.49) で与えられることが分かる.
問題 3.10

(2.115) の結果を用いて (3.57) の積分を評価し，ベイズ線形回帰モデルの予測分布が (3.58) で与えられることを確かめよ．ただし，入力に依存する分散は (3.59) で与えられる．
方針 3.10

解答 3.10
(3.57) の条件付き分布と事後分布は, それぞれ以下の式で表される.

$$p(t \mid \mathbf{w}, \beta) = \mathcal{N}\left(t \mid \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}), \beta^{-1}\right) \tag{3.3, 3.8}$$

$$p(\mathbf{w} \mid \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \mathbf{S}_N) \tag{3.49}$$

ここで, (2.115) の式は,

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right) \tag{2.113}$$

と

$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}\right) \tag{2.114}$$

が与えられた際の周辺分布だったことに注意して, (2.113) から (2.115) について,

$$\mathbf{y} \to t, \quad \mathbf{x} \to \mathbf{w}, \quad \boldsymbol{\mu} \to \mathbf{m}_N, \quad \boldsymbol{\Lambda}^{-1} \to \mathbf{S}_N, \quad \mathbf{A} \to \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}, \quad \mathbf{L}^{-1} \to \beta^{-1}$$

と置き換えると, (3.57) を評価することができる.
したがって, (2.115) にそれぞれを代入すると,

$$p(t \mid \mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}\left(t \mid \mathbf{m}_N^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}), \sigma_N^2(\mathbf{x})\right) \tag{3.58}$$

と求まる. ここで, 入力に依存する分散は,

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}) \tag{3.59}$$

である.
問題 3.15
線形基底関数からなる回帰モデルの超パラメータ α,β をエビデンス枠組みを用いて決定する場合を考える.(3.82) で定義される関数 $E(m_N)$ が関係式 $2E(m_N) = N$ を満たすことを示せ

解答 3.15
エビデンス関数を最大にする α と β は (3.92) 式と (3.95) 式より

$$\begin{cases} \alpha = \dfrac{\gamma}{m_N^T m_N} \\ \beta = (N-\gamma)\{\sum_{n=1}^{N}\{t_N - m_N^T \phi(x_N)\}^2\}^{-1} \end{cases}$$

これらを (3.82) 式に代入すると

$$\begin{aligned} E(m_N) &= \frac{\beta}{2}||t - \phi m_N||^2 + \frac{\alpha}{2} m_N^T m_N \\ &= \frac{N-\gamma}{2} + \frac{\gamma}{2} \\ &= \frac{N}{2} \end{aligned}$$

なお

$$||t - \phi m_N||^2 = \sum_{n=1} (t_n - m_N^T \phi(x_n))^2$$

である。