PRML 3 線形回帰モデル

大阪 PRML 読書会

2014 年 8 月 31 日

目次

導入

回帰問題に対する二つのアプローチ

- 予測関数 $y(\mathbf{x})$ を構成する。
- 条件付き確率分布関数 $p(t|\mathbf{x})$ を構成する。
  その分布の下で期待損失を最小化するように
  $\mathbf{x}$ に対する t を決定する。

線形基底関数モデル

一般形

$$
\begin{aligned}
y(\mathbf{x}, \mathbf{w}) =& w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) \\
=& \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) \\
=& \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})
\end{aligned}
$$

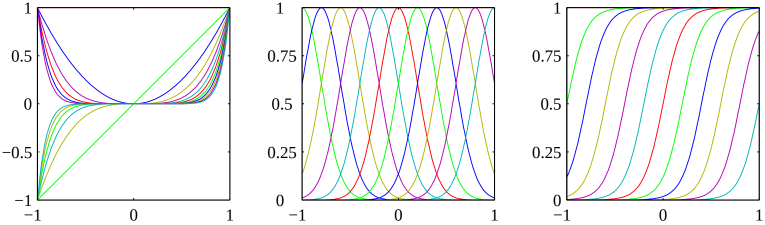
where $\phi_j(\mathbf{x})$: **基底関数**

$\phi_0(\mathbf{x}) = 1$: **ダミーの基底関数**

$\boldsymbol{\phi} = (\phi_0, ..., \phi_{M-1})^{\mathrm{T}}$

$\mathbf{w} = (w_0, ..., w_{M-1})^{\mathrm{T}}$

線形基底関数モデル



最尤推定と最小二乗法

目標変数 t が、決定論的な関数 $y(\mathbf{x}, \mathbf{w})$ とガウスノイズの和で与えられると仮定する。

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon \qquad \text{where} \qquad p(\epsilon|\beta) = \mathcal{N}(\epsilon|0, \beta^{-1})$$

これは以下と等価である。

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1})$$

最尤推定と最小二乗法

決定段階

損失関数として二乗損失関数を仮定すると、
新しい x の値に対する最適な予測値は
目標変数の条件付き期待値で与えられる。

$$t = \mathbb{E}[t|\mathbf{x}] = \int tp(t|\mathbf{x})dt$$

先に仮定した条件付き確率分布を用いると、

$$\begin{aligned} t &= \int t\mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1})dt \\ &= y(\mathbf{x}, \mathbf{w}) \end{aligned}$$

t は決定論的関数の値として与えられる。

最尤推定と最小二乗法

データ集合

$$\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)^{\mathrm{T}}$$
$$\mathsf{t} = (t_1, ..., t_N)^{\mathrm{T}}$$

尤度関数

$$p(\mathsf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})$$

最尤推定と最小二乗法

対数尤度関数

$$
\begin{aligned}
\ln p(\mathsf{t}|\mathbf{w},\beta) &= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n),\beta^{-1}) \\
&= \sum_{n=1}^{N} \ln \left[\left(\frac{\beta}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\beta}{2}(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))^2 \right\} \right] \\
&= \frac{N}{2}\ln\beta - \frac{N}{2}\ln 2\pi - \beta E_D(\mathbf{w})
\end{aligned}
$$

ただし

$$
E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n) \right\}^2
$$

E_D **は二乗和誤差関数。**

最尤推定と最小二乗法

w の最尤推定

$$
\begin{aligned}
\nabla \ln p(\mathsf{t}|\mathbf{w},\beta) &= \beta \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \\
0 &= \sum_{n=1}^{N} t_n \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} - \mathbf{w}_{\mathrm{ML}}^{\mathrm{T}} \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right) \\
0 &= \boldsymbol{\Phi}^{\mathrm{T}} \mathsf{t} - \mathbf{w}_{\mathrm{ML}}^{\mathrm{T}} \left(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \right) \\
\mathbf{w}_{\mathrm{ML}} &= \left(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \mathsf{t} = \boldsymbol{\Phi}^{\dagger} \mathsf{t} \qquad \cdots \textbf{正規方程式}
\end{aligned}
$$

ただし、$\boldsymbol{\Phi}$ は計画行列 design matrix**。**

$$
\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}} \\ \boldsymbol{\phi}(\mathbf{x}_2)^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\phi}(\mathbf{x}_N)^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}
$$

最尤推定と最小二乗法

バイアスパラメータ w_0 の役割

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \left\{ t_n - w_0 - \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}_n) \right\}^2$$

$$\frac{\partial}{\partial w_0} E_D(\mathbf{w}) = - \sum_{n=1}^{N} \left\{ t_n - w_0 - \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}_n) \right\}$$

$$w_0^{\mathrm{ML}} = \bar{t} - \sum_{j=1}^{M-1} w_j^{\mathrm{ML}} \overline{\phi_j}$$

$$\bar{t} = \frac{1}{N} \sum_{n=1}^{N} t_n$$

目標値の訓練集合に関する平均

$$\overline{\phi_j} = \frac{1}{N} \sum_{n=1}^{N} \phi_j(\mathbf{x}_n)$$

基底関数の値の訓練集合に関する平均

最尤推定と最小二乗法

β の最尤推定

$$\frac{\partial}{\partial\beta}\ln p(\mathbf{t}|\mathbf{w},\beta) = \frac{\partial}{\partial\beta}\left\{\frac{N}{2}\ln\beta - \frac{N}{2}\ln 2\pi - \beta E_D(\mathbf{w})\right\}$$
$$= \frac{N}{2}\frac{1}{\beta} - E_D(\mathbf{w})$$

$$\frac{N}{2}\frac{1}{\beta_{\mathrm{ML}}} - E_D(\mathbf{w}_{\mathrm{ML}}) = 0$$

$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N}\sum_{n=1}^{N}\left\{t_n - \mathbf{w}_{\mathrm{ML}}^{\mathrm{T}}\phi(\mathbf{x}_n)\right\}^2$$

ノイズの精度の逆数は、
回帰関数周りでの目標値の残差分散で与えられる。

最小二乗法の幾何学

各軸が t_n **で与えられる** N **次元空間。**

$\mathsf{t} = (t_1, ..., t_N)^{\mathrm{T}}$

訓練データ集合の目標値の集合。

$\boldsymbol{\varphi}_j = (\phi_j(\mathbf{x}_1), ..., \phi_j(\mathbf{x}_N))^{\mathrm{T}}$

訓練データ集合の基底関数の値の集合。

$\boldsymbol{\varphi}_j$ **は** Φ **の列ベクトル。** $\phi(\mathrm{x}_n)^{\mathrm{T}}$ **は行ベクトル。**

M **個の** $\boldsymbol{\varphi}_j$ **は** M **次元の線形部分空間** $\mathcal{S}$ **を張る。**

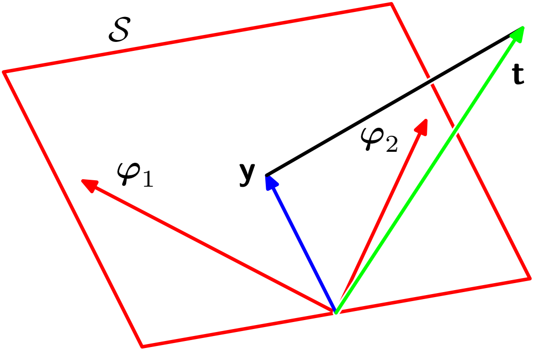
$\mathsf{y} = (y(\mathbf{x}_1, \mathbf{w}), ..., y(\mathbf{x}_N, \mathbf{w}))^{\mathrm{T}}$

目標値の予測値の集合。

y **は** $\boldsymbol{\varphi}_j$ **の線形結合なので** $\mathcal{S}$ **上にある。**

$$
\begin{aligned}
\mathsf{y} =& (\boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{w}, ..., \boldsymbol{\phi}(\mathbf{x}_N)^{\mathrm{T}}\mathbf{w})^{\mathrm{T}} \\
=& (\boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}, ..., \boldsymbol{\phi}(\mathbf{x}_N)^{\mathrm{T}})^{\mathrm{T}}\mathbf{w} \\
=& \boldsymbol{\Phi}\mathbf{w} \\
=& (\boldsymbol{\varphi}_0, ..., \boldsymbol{\varphi}_{M-1})\mathbf{w}
\end{aligned}
$$

最小二乗法の幾何学



$$\sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\}^2 = \| \mathsf{t} - \mathsf{y} \|^2$$

演習3.2

[定義] v' が v の $\mathcal{S}$ への正射影である。
ただし、$\mathcal{S}$ は基底ベクトル $\{\mathrm{e}_j\}$ が張る空間。
$\mathbf{A} = (\mathbf{e}_1, ..., \mathbf{e}_M)$。

1. $\mathbf{v}'$ が $\mathcal{S}$ 上にある。
   $\exists \mathbf{w}.\mathbf{v}' = \sum_j \mathbf{e}_j w_j \Leftrightarrow \exists \mathbf{w}.\mathbf{v}' = \mathbf{A}\mathbf{w}$
2. $\mathbf{v} - \mathbf{v}'$ が $\mathcal{S}$ と直交する。
   $\forall j.\mathbf{e}_j(\mathbf{v} - \mathbf{v}') = 0 \Leftrightarrow \mathbf{A}^{\mathrm{T}}(\mathbf{v} - \mathbf{v}') = 0$

$\mathbf{v}' = \mathbf{\Phi}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{v}$ と置く。

1. $\mathbf{w} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{v}$ と置けばただちに成り立つ。
2. $$\begin{aligned} \mathbf{\Phi}^{\mathrm{T}}(\mathbf{v} - \mathbf{v}') =& \mathbf{\Phi}^{\mathrm{T}}(\mathbf{v} - \mathbf{\Phi}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{v}) \\ =& (\mathbf{\Phi}^{\mathrm{T}} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}})\mathbf{v} \\ =& (\mathbf{\Phi}^{\mathrm{T}} - \mathbf{\Phi}^{\mathrm{T}})\mathbf{v} \\ =& 0 \end{aligned}$$

演習3.2

y の定義より

$$\mathrm{y} = \boldsymbol{\Phi}\mathbf{w}$$

$\mathbf{w}$ の最小二乗解は正規方程式 (3.15) により

$$\mathbf{w}_{\mathrm{ML}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathrm{t}$$

y の最小二乗解は

$$\mathrm{y}_{\mathrm{ML}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathrm{t}$$

よって、y の最小二乗解は
t の Φ の列ベクトルが張る多様体 $\mathcal{S}$ への正射影である。

逐次学習

確率的勾配降下法 stochastic gradient descent

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E_n$$

二乗和誤差関数

$$E_n(\mathbf{w}) = \frac{1}{2}(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))^2$$

を用いると

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \eta(t_n - \mathbf{w}^{(\tau)\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))\boldsymbol{\phi}(\mathbf{x}_n)$$

最小平均二乗アルゴリズム
least-mean-squares (LMS) algorithm

正則化最小二乗法

正則化された誤差関数

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

$E_D(\mathbf{w})$: **データに依存する誤差**

$E_W(\mathbf{w})$: **正則化項**

二乗和誤差関数 + 二次正則化項

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

最小化

$$\mathbf{w} = (\lambda\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathsf{t}$$

正則化最小二乗法

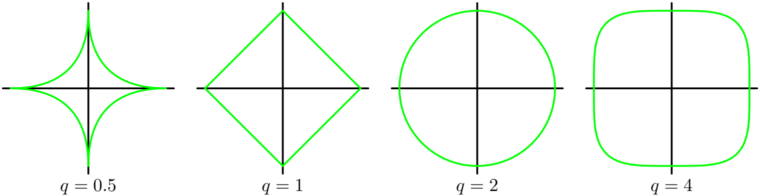
誤差関数の最小化

$$
\begin{aligned}
&\nabla \left[\frac{1}{2} \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \right] \\
&= -\boldsymbol{\Phi}^{\mathrm{T}} \mathsf{t} + \mathbf{w}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} + \lambda \mathbf{w}^{\mathrm{T}} \\
&= -\boldsymbol{\Phi}^{\mathrm{T}} \mathsf{t} + \mathbf{w}^{\mathrm{T}} \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} + \mathbf{w}^{\mathrm{T}} (\lambda \mathbf{I}) \\
&= -\boldsymbol{\Phi}^{\mathrm{T}} \mathsf{t} + \mathbf{w}^{\mathrm{T}} (\lambda \mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi})
\end{aligned}
$$

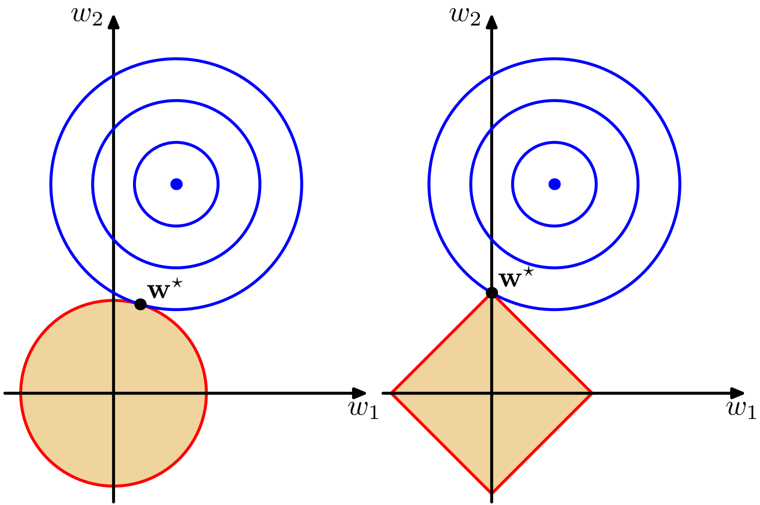
正則化最小二乗法

二乗和誤差関数 + q 次正則化項

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$$



正则化最小二乘法



演習3.5

制約条件 (3.30) **は以下のように書き換えられる。**

$$\sum_{j=1}^{M} |w_j|^q \leq \eta \Leftrightarrow -\frac{1}{2}\left(\sum_{j=1}^{M} |w_j|^q - \eta\right) \geq 0$$

ラグランジュ関数

$$L(\mathbf{w}, \lambda) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\left(\sum_{j=1}^{M} |w_j|^q - \eta\right)$$

where $\quad g(\mathbf{w}) \geq 0 \qquad \lambda \geq 0 \qquad \lambda g(\mathbf{w}) = 0$

停留条件

$\nabla L(\mathbf{w}, \lambda) = 0 \rightarrow$ **正則化された誤差関数の最小化と同じ**

$$\frac{\partial}{\partial \lambda} L(\mathbf{w}, \lambda) = 0 \rightarrow \qquad \eta = \sum_{j=1}^{M} |w_j|^q = \sum_{j=1}^{M} |w_j(\lambda)|^q$$

出力変数が多次元の場合

t のすべての要素に同じ基底関数を用いたモデル

$$\mathbf{y}(\mathbf{x}, \mathbf{w}) = \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

t の条件付き分布を等方性ガウス分布と仮定する。

$$p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \beta) = \mathcal{N}(\mathbf{t}|\mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}), \beta^{-1}\mathbf{I})$$

出力変数が多次元の場合

データ集合　$\mathbf{T} = (\mathbf{t}_1, ..., \mathbf{t}_N)^{\mathrm{T}}$、$\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)^{\mathrm{T}}$

対数尤度関数

$$
\begin{aligned}
&\ln p(\mathbf{T}|\mathbf{X}, \mathbf{W}, \beta) \\
=&\sum_{n=1}^{N} \ln \mathcal{N}(\mathbf{t}_n|\mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1}\mathbf{I}) \\
=&\sum_{n=1}^{N} \ln \left[\left(\frac{\beta}{2\pi} \right)^{K/2} \exp\left\{ -\frac{\beta}{2} \|\mathbf{t}_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\|^2 \right\} \right] \\
=&\frac{NK}{2} \ln \left(\frac{\beta}{2\pi} \right) - \frac{\beta}{2} \sum_{n=1}^{N} \|\mathbf{t}_n - \mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\|^2
\end{aligned}
$$

最尤解

$$
\begin{aligned}
\mathbf{W}_{\mathrm{ML}} =& (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{T} = \boldsymbol{\Phi}^{\dagger}\mathbf{T} \\
\mathbf{w}_k =& (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathsf{t}_k = \boldsymbol{\Phi}^{\dagger}\mathsf{t}_k
\end{aligned}
$$