ベイズミニマムノルム推定

株式会社　シグナルアナリシス　関原謙介

1　データモデル

本当は $x_1, x_2, \ldots, x_N$ を観測したいのだがこれらは直接観測できず，代わりに一群の観測データ $y_1, y_2, \ldots, y_M$ が得られる状況を考えよう．知りたい量 $x_1, x_2, \ldots, x_N$ と観測データ $y_1, y_2, \ldots, y_M$ を，次の列ベクトル $\boldsymbol{x}$ と $\boldsymbol{y}$ で表現する．すなわち，

$$\text{未知量:}\quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \text{観測データ:}\quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \tag{1}$$

と定義する．この $\boldsymbol{x}$ は未知量ベクトルあるいは解ベクトル，$\boldsymbol{y}$ は観測ベクトルあるいはデータベクトルと呼ばれる．ベクトル $\boldsymbol{x}$ の集合は解空間 (solution space)，ベクトル $\boldsymbol{y}$ の集合は観測空間 (observation space) と呼ばれることもある．

ベクトル $\boldsymbol{x}$ とベクトル $\boldsymbol{y}$ は線形な関係,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \tag{2}$$

で結ばれているとする．ここで $\boldsymbol{H}$ は未知量 $\boldsymbol{x}$ と観測結果 $\boldsymbol{y}$ を結びつける $M \times N$ の行列である．本書では未知量 $\boldsymbol{x}$ と観測データ $\boldsymbol{y}$ は共に実数であり，$\boldsymbol{H}$ も実数行列であるとする．さらに，観測結果には，信号成分に加法的にノイズ $\boldsymbol{\varepsilon}$ が重畳すると仮定すれば，観測データのモデルは

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{3}$$

となる．ノイズベクトル $\boldsymbol{\varepsilon}$ は M 次元の列ベクトルで,

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \tag{4}$$

であり，j 番目の要素 ε_j は j 番目の観測データ y_j に重畳するノイズを表す．式 (3) で表される観測モデルを線形離散モデルと呼ぶ.

2　確率モデル

線形離散モデル

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

を仮定し，観測データ $\boldsymbol{y}$ から未知な量 $\boldsymbol{x}$ を求める問題を引き続き考えよう．事前分布として，$\boldsymbol{x}$ の各要素に対し独立で同一な正規分布，

$$x_j \sim \mathcal{N}(x_j|0, \alpha^{-1}) \tag{5}$$

を仮定して，この事前分布のもとで L_2 ノルム正則化ミニマムノルム解が導かれる．精度行列を用いて議論を進めることにして，事前確率分布とデータ尤度が

$$p(\boldsymbol{x}|\boldsymbol{\Phi}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Phi}^{-1}) \tag{6}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{H}\boldsymbol{x}, \beta^{-1}\boldsymbol{I}) \tag{7}$$

の確率モデルに従うとする．説明を簡単にするため，ノイズの精度行列は $\beta\boldsymbol{I}$ として，精度 β は既知であるとする．また，$\boldsymbol{\Phi}$ が事前分布の精度行列で，$N \times N$ の対角行列：

$$\boldsymbol{\Phi} = \begin{bmatrix} \alpha & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \alpha \end{bmatrix}$$

と仮定する．したがって，

$$p(\boldsymbol{x}|\boldsymbol{\Phi}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Phi}^{-1}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \alpha^{-1}\boldsymbol{I}) \tag{8}$$

と与えられる．

3　事後分布の導出

以上の仮定のもとで，事後分布 $p(\boldsymbol{x}|\boldsymbol{y})$ を求めてみよう．事後分布も正規分布であるので，事後分布の平均を $\bar{\boldsymbol{x}}$，精度行列を $\boldsymbol{\Gamma}$ として，

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}|\bar{\boldsymbol{x}}, \boldsymbol{\Gamma}^{-1}) \tag{9}$$

と仮定する．式 (9) 右辺の指数部分は

$$-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T\boldsymbol{\Gamma}(\boldsymbol{x} - \bar{\boldsymbol{x}}) = -\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Gamma}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{\Gamma}\bar{\boldsymbol{x}} + \mathcal{C} \tag{10}$$

となる．ここで，$\mathcal{C}$ は $\boldsymbol{x}$ を含まない項をまとめて表す．ベイズ定理から

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \tag{11}$$

が成立する．右辺に式 (8) と (7) を代入し，指数部分を書き出すと，

$$-\frac{1}{2}\left[\alpha^{-1}\boldsymbol{x}^T\boldsymbol{x} + \beta^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})\right] \tag{12}$$

となる．この式を $\boldsymbol{x}$ に関して整理すると，

$$-\frac{1}{2}\boldsymbol{x}^T(\alpha^{-1}\boldsymbol{I}+\beta^{-1}\boldsymbol{H}^T\boldsymbol{H})\boldsymbol{x}+\beta^{-1}\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{y}+\mathcal{C}' \tag{13}$$

となる．$\mathcal{C}'$ も $\boldsymbol{x}$ を含まない項をまとめて表したものである．式 (10) と (13) において，$\boldsymbol{x}$ の 2 次の項の係数行列と，$\boldsymbol{x}$ の 1 次の項の係数を比較して

$$\boldsymbol{\Gamma}=\alpha^{-1}\boldsymbol{I}+\beta^{-1}\boldsymbol{H}^T\boldsymbol{H} \tag{14}$$

$$\bar{\boldsymbol{x}}=\beta^{-1}(\alpha^{-1}\boldsymbol{I}+\beta^{-1}\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y}=(\boldsymbol{H}^T\boldsymbol{H}+\frac{\beta}{\alpha}\boldsymbol{I})^{-1}\boldsymbol{H}^T\boldsymbol{y} \tag{15}$$

を得る．式 (15) は逆行列の公式を用いて

$$\bar{\boldsymbol{x}}=(\boldsymbol{H}^T\boldsymbol{H}+\frac{\beta}{\alpha}\boldsymbol{I})^{-1}\boldsymbol{H}^T\boldsymbol{y}=\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{H}^T+\frac{\beta}{\alpha}\boldsymbol{I}\right)^{-1}\boldsymbol{y} \tag{16}$$

とも変形できる．上式 (16) は L_2 ノルム正則化を用いたミニマムノルム解に等しい．すなわち，線形正規モデルのベイズ推定解から，L_2 ノルム正則化が組み込まれた形のミニマムノルム解を導くことができる．

4 EM アルゴリズムの導出

以下にハイパーパラメータ α と β の更新式を導いてみよう．平均データ尤度は，定数項を無視して，

$$\Theta(\alpha,\beta)=\frac{N}{2}\log\alpha-\frac{\alpha}{2}E\left[\boldsymbol{x}^T\boldsymbol{x}\right]+\frac{M}{2}\log\beta-\frac{\beta}{2}E\left[(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\right] \tag{17}$$

である．従って，

$$\frac{\partial}{\partial\alpha}\Theta(\alpha,\beta)=\frac{N}{2\alpha}-\frac{1}{2}E\left[\boldsymbol{x}^T\boldsymbol{x}\right]=0$$

および

$$E\left[\boldsymbol{x}^T\boldsymbol{x}\right]=\mathrm{tr}\left(E[\boldsymbol{x}\boldsymbol{x}^T]\right)=\mathrm{tr}\left(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T+\boldsymbol{\Gamma}^{-1}\right)=\bar{\boldsymbol{x}}^T\bar{\boldsymbol{x}}+\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\right)$$

から，

$$\widehat{\alpha}^{-1}=\frac{1}{N}E\left[\boldsymbol{x}^T\boldsymbol{x}\right]=\frac{1}{N}\left[\|\bar{\boldsymbol{x}}\|^2+\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\right)\right] \tag{18}$$

を得る．

β については，同様に

$$\frac{\partial}{\partial\beta}Q(\alpha,\beta)=\frac{M}{2\beta}-\frac{1}{2}E\left[(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\right]$$

から，

$$\widehat{\beta}^{-1}=\frac{1}{M}E\left[(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\right] \tag{19}$$

である．ここで，

$$E\left[(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\right]==\boldsymbol{y}^T\boldsymbol{y}-\bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{y}-\boldsymbol{y}^T\boldsymbol{H}\bar{\boldsymbol{x}}+E[\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}] \tag{20}$$

であり，さらに，

$$E[\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}] = E[\mathrm{tr}(\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}\boldsymbol{x}^T)] = \mathrm{tr}(\boldsymbol{H}^T\boldsymbol{H}E\left(\boldsymbol{x}\boldsymbol{x}^T\right))$$
$$= \mathrm{tr}\left[\boldsymbol{H}^T\boldsymbol{H}\left(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T + \boldsymbol{\Gamma}^{-1}\right)\right] = \bar{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{H}\bar{\boldsymbol{x}} + \mathrm{tr}\left[\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Gamma}^{-1}\right] \quad (21)$$

であるので，

$$\widehat{\beta}^{-1} = \frac{1}{M}\left[\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \mathrm{tr}\left[\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Gamma}^{-1}\right]\right] \quad (22)$$

を得る．

5　ベイズミニマムノルムアルゴリズム

以上，ベイズ L_2 正則化ミニマムノルム解をまとめると，この方法では EM アルゴリズムを用いて，未知量 $\boldsymbol{x}$ とそれを求めるための正則化定数を再帰的に求める．実際の計算では，まず，α と β に適当な初期値を与え，E ステップにおいて，

$$\boldsymbol{\Gamma} = \alpha\boldsymbol{I} + \beta\boldsymbol{H}^T\boldsymbol{H} \quad (23)$$
$$\bar{\boldsymbol{x}} = \beta\boldsymbol{\Gamma}^{-1}\boldsymbol{H}^T\boldsymbol{y} = \boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{H}^T + \frac{\alpha}{\beta}\boldsymbol{I}\right)^{-1}\boldsymbol{y} \quad (24)$$

を用いて事後分布のパラメター$\boldsymbol{\Gamma}$ と $\bar{\boldsymbol{x}}$ を計算する．次に M ステップでは，

$$\widehat{\alpha}^{-1} = \frac{1}{N}\left[\|\bar{\boldsymbol{x}}\|^2 + \mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\right)\right]$$
$$\widehat{\beta}^{-1} = \frac{1}{M}\left[\|\boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{x}}\|^2 + \mathrm{tr}\left[\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Gamma}^{-1}\right]\right]$$

から，パラメータ α と β の値を更新する．上記 E ステップと M ステップを収束条件が満たされるまで繰り返す．収束の判定は，周辺尤度を以下の式

$$\log p(\boldsymbol{y}|\alpha,\beta) = -\frac{1}{2}|\beta^{-1}\boldsymbol{I} + \alpha^{-1}\boldsymbol{H}\boldsymbol{H}^T| - \frac{1}{2}\boldsymbol{y}^T(\beta^{-1}\boldsymbol{I} + \alpha^{-1}\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{y} \quad (25)$$

から計算し，周辺尤度 $\log p(\boldsymbol{y}|\alpha,\beta)$ の増加が，更新を繰り返してもほとんど増加しなくなったときに計算を打ち切る．