『入門 機械学習による異常検知 — R による実践ガイド —』（コロナ社、2015）の章末問題の解答

Tsuyoshi Idé (井手 剛) ide@ide-research.net

平成 28 年 2 月 21 日

1 異常検知の基本的な考え方

この章に章末問題はありません。

2 正規分布に従うデータからの異常検知

2.0.1

1 次元の正規分布の確率密度関数

$$\mathcal{N}(x \mid \mu, \sigma) \equiv \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

を x の関数と見て、それを x で微分することにより、極大点と変曲点を求めてください。

[解答] 確率密度関数を x で微分すると

$$\frac{\partial \mathcal{N}(x \mid \mu, \sigma)}{\partial x} = \mathcal{N}(x \mid \mu, \sigma) \times \left(-\frac{x-\mu}{\sigma^2}\right)$$

なので、これを 0 と等置して、$x = \mu$ において唯一の極値をとることがわかります。また、

$$\begin{aligned}\frac{\partial^2 \mathcal{N}(x \mid \mu, \sigma)}{\partial x^2} &= \mathcal{N}(x \mid \mu, \sigma)\left(-\frac{1}{\sigma^2}\right) + \mathcal{N}(x \mid \mu, \sigma)\left(-\frac{x-\mu}{\sigma^2}\right)^2 \\ &= \mathcal{N}(x \mid \mu, \sigma)\frac{1}{\sigma^4}\left\{-\sigma^2 + (x-\mu)^2\right\}\end{aligned}$$

したがって、変曲点は、$x = \mu \pm \sigma$ で生じます。$x = \mu$ では 2 階導関数が負になることから、極値を与える $x = \mu$ は最大値を与えます。

2.0.2

付録の定理 A.6 の式 (A.38)

$$\int_{-\infty}^{+\infty} \mathrm{d}x \ \exp(-ax^2 + bx + c) = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right)$$

を用いて、正規分布 (2.3) が規格化条件を満たすことを確かめてください †。

[解答] 正規分布の確率密度関数が規格化条件を満たすことを示すためには

$$\int_{-\infty}^{+\infty} \mathrm{d}x \, \mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{d}x \ \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

が 1 になることを示す必要があります。公式において

$$a = \frac{1}{2\sigma^2}, \ \ b = \frac{\mu}{\sigma^2}, \ \ c = -\frac{\mu^2}{2\sigma^2}$$

と置くと

$$\begin{aligned}\int_{-\infty}^{+\infty} \mathrm{d}x \, \mathcal{N}(x \mid \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{\frac{\pi}{1/(2\sigma^2)}} \exp\left\{\frac{\mu^2}{\sigma^4} \times \frac{1}{4/(2\sigma^2)} - \frac{\mu^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{\mu^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\} \\ &= 1\end{aligned}$$

のように計算できます。

2.0.3

正規分布 (2.3) に従う変数 x があった時、変数変換 $z = \dfrac{x-\mu}{\sigma}$ により定義される変数 z が**標準正規分布**に従うことを証明してください。なお、標準正規分布とは平均ゼロ、分散 1 の正規分布のことです。

[解答] 付録の定理 A.2 を使います。変換後の確率密度関数を $q(z)$ と置いたとします。z と x は

$$z = \frac{1}{\sigma}x - \frac{\mu}{\sigma}, \quad x = \sigma z + \mu$$

† なお、初版第 1 刷において、上式右辺に誤植がありました。お詫びして修正させていただきます。正誤表も参照のこと。

という関係で結ばれます。式 (A.8) において、y を z と読み替え、$T = \frac{1}{\sigma}$ および $b = -\frac{\mu}{\sigma}$ と置くと

$$\begin{aligned} q(z) &= \sigma \mathcal{N}(\sigma z + \mu \mid \mu, \sigma^2) \\ &= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\sigma z + \mu - \mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= \mathcal{N}(z \mid 0, 1) \end{aligned}$$

2.0.4

1 次元確率変数の N 個の標本を使って、次のような式を考えます。

$$\omega \equiv \frac{1}{2N^2} \sum_{n=1}^{N} \sum_{n'=1}^{N} (x^{(n)} - x^{(n')})^2$$

これは力学的には、N 個の標本が互いにばねでつながれたときのポテンシャルエネルギーの平均と解釈できます。これが標本分散の式 (2.5)

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (x^{(n)} - \hat{\mu})^2$$

と一致することを証明してください。

[解答] 標本分散の式から出発する方が簡単なのでそうします。上の式において、標本平均の式

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x^{(n)}$$

を代入して展開すると

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{N}\sum_{n=1}^{N}\left\{x^{(n)} - \frac{1}{N}\sum_{m=1}^{N}x^{(m)}\right\}^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\left\{{x^{(n)}}^2 - \frac{2}{N}x^{(n)}\sum_{m=1}^{N}x^{(m)} + \frac{1}{N^2}\sum_{m=1}^{N}\sum_{m'=1}^{N}x^{(m)}x^{(m')}\right\} \\ &= \frac{1}{N}\sum_{n=1}^{N}{x^{(n)}}^2 - \frac{1}{N^2}\sum_{m=1}^{N}\sum_{m'=1}^{N}x^{(m)}x^{(m')}\end{aligned}$$

一方 ω の式を展開すると

$$\begin{aligned}\omega &= \frac{1}{2N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left\{{x^{(n)}}^2 + {x^{(n')}}^2 - 2x^{(n)}x^{(n')}\right\} \\ &= \frac{1}{N}\sum_{n=1}^{N}{x^{(n)}}^2 - \frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}x^{(n)}x^{(n')}\end{aligned}$$

となりますので、両者の一致が分かります。

2.0.5

正方行列 A と、それと同じ次元を持つベクトル $\boldsymbol{a}, \boldsymbol{b}$ について、$\boldsymbol{a}^\top \mathsf{A}\boldsymbol{b} = \mathrm{Tr}(\mathsf{A}\boldsymbol{b}\boldsymbol{a}^\top)$ が成り立つことを証明してください。

[解答] 行列の積の定義から

$$\boldsymbol{a}^\top \mathsf{A}\boldsymbol{b} = \sum_{i,j} a_i \mathsf{A}_{i,j} b_j$$

となります。一方、行列 $\mathsf{A}\boldsymbol{b}\boldsymbol{a}^\top$ の (i,i) 成分 $[\mathsf{A}\boldsymbol{b}\boldsymbol{a}^\top]_{i,i}$ は、やはり行列の積の定義から

$$[\mathsf{A}\boldsymbol{b}\boldsymbol{a}^\top]_{i,i} = \sum_k \mathsf{A}_{i,k}[\boldsymbol{b}\boldsymbol{a}^\top]_{k,i} = \sum_k \mathsf{A}_{i,k} b_k a_i = \sum_k a_i \mathsf{A}_{i,k} b_k$$

となります。行列の跡（トレース）とは、対角成分についての和を実行したものですから、これを使うと

$$\mathrm{Tr}(\mathsf{A}\boldsymbol{b}\boldsymbol{a}^\top) = \sum_i \sum_k a_i \mathsf{A}_{i,k} b_k$$

となっていることが分かります。これは $\boldsymbol{a}^\top \mathsf{A} \boldsymbol{b}$ と一致します。

2.0.6

$\boldsymbol{z}$ と $\boldsymbol{z}'$ がそれぞれ独立に $\mathcal{N}(\boldsymbol{0}, \Sigma)$ に従うとします。$\boldsymbol{z}\boldsymbol{z}'^\top$ という量の期待値は何でしょうか。また、$\boldsymbol{z}\boldsymbol{z}^\top$ の期待値は何でしょうか。

[解答] $\boldsymbol{z}$ と $\boldsymbol{z}'$ の同時分布を $p(\boldsymbol{z}, \boldsymbol{z}')$ と表すと、仮定から

$$p(\boldsymbol{z}, \boldsymbol{z}') = \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{0}, \Sigma)\mathcal{N}(\boldsymbol{z}' \mid \boldsymbol{0}, \Sigma)$$

となります。期待値を $\langle \cdot \rangle$ と表すと、期待値の定義から、

$$\begin{aligned}
\langle \boldsymbol{z}\boldsymbol{z}'^\top \rangle &= \int \mathrm{d}\boldsymbol{z} \int \mathrm{d}\boldsymbol{z}'\, p(\boldsymbol{z}, \boldsymbol{z}') \boldsymbol{z}\boldsymbol{z}'^\top \\
&= \int \mathrm{d}\boldsymbol{z}\, \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{0}, \Sigma)\boldsymbol{z} \int \mathrm{d}\boldsymbol{z}'\, \mathcal{N}(\boldsymbol{z}' \mid \boldsymbol{0}, \Sigma)\boldsymbol{z}'^\top \\
&= \boldsymbol{0}\boldsymbol{0}^\top
\end{aligned}$$

$\boldsymbol{z}$ と $\boldsymbol{z}'$ が M 次元の確率変数だったとすると、これは $M \times M$ のゼロ行列です。

次に、

$$\begin{aligned}
\langle \boldsymbol{z}\boldsymbol{z}^\top \rangle &= \int \mathrm{d}\boldsymbol{z} \int \mathrm{d}\boldsymbol{z}'\, p(\boldsymbol{z}, \boldsymbol{z}') \boldsymbol{z}\boldsymbol{z}^\top \\
&= \int \mathrm{d}\boldsymbol{z}\, \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{0}, \Sigma)\boldsymbol{z}\boldsymbol{z}^\top \int \mathrm{d}\boldsymbol{z}'\, \mathcal{N}(\boldsymbol{z}' \mid \boldsymbol{0}, \Sigma) \\
&= \Sigma \times 1 \\
&= \Sigma
\end{aligned}$$

ただし、共分散行列の定義から、一般に

$$\Sigma = \int \mathrm{d}\boldsymbol{z}\, \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{\mu}, \Sigma)\, (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})^\top$$

となることを使いました。なお、これを仮に知らなかったとしても、ガウス積分の公式を使って直接積分することも可能です。そちらは読者に任せます。

2.0.7

M 変数のホテリングの統計量

$$T^2 \equiv \frac{N-M}{(N+1)M}(\boldsymbol{x}' - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}' - \hat{\boldsymbol{\mu}})$$

は、各変数が統計的に無相関であるとき、1 変数のホテリング統計量の和で表されることを証明してください。

[解答]

無相関であれば任意の 2 つの異なる変数の間の共分散が 0 になりますから、標本共分散行列が

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\sigma}_1^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_M^2 \end{pmatrix}$$

のような対角行列になります。ただし、変数の次元を M とし、変数 x_i の標本共分散を $\hat{\sigma}_i^2$ と置きました。これの逆行列は明らかに

$$\hat{\boldsymbol{\Sigma}}^{-1} = \begin{pmatrix} 1/\hat{\sigma}_1^2 & 0 & \dots & 0 \\ 0 & 1/\hat{\sigma}_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1/\sigma_M^2 \end{pmatrix}$$

ですから、

$$T^2 \equiv \frac{N-M}{(N+1)M} \sum_{l=1}^{M} \left(\frac{x_l - \hat{\mu}_l}{\sigma_l} \right)^2$$

となります。明らかにこれは、1 変数のホテリング統計量の和の定数倍になっています。

3 非正規データからの異常検知

3.0.1

ガンマ関数の定義

$$\Gamma(z) \equiv \int_0^\infty \mathrm{d}t\ t^{z-1}\mathrm{e}^{-t}$$

に部分積分を適用して直接 $\Gamma(2) = 1$ を証明してください。また、ガンマ関数の定義を用いて、ガンマ分布の規格化条件の成立を証明してください。

[解答] まず $\Gamma(2) = 1$ を示します。部分積分により容易に次の計算ができます。

$$\begin{aligned}
\Gamma(2) &= \int_0^\infty \mathrm{d}t\ t\ \mathrm{e}^{-t} = \int_0^\infty \mathrm{d}t\ t\frac{\mathrm{d}}{\mathrm{d}t}(-\mathrm{e}^{-t}) \\
&= \left[-t\ \mathrm{e}^{-t}\right]_0^\infty + \int_0^\infty \mathrm{d}t\ \mathrm{e}^{-t} \\
&= 0 - \left[\mathrm{e}^{-t}\right]_0^\infty \\
&= 1
\end{aligned}$$

ただし高校数学で習ったはずの次式を使いました。

$$\lim_{t\to\infty}\frac{t}{\mathrm{e}^t} = 0$$

ガンマ分布の規格化条件を示すためには

$$\int_0^\infty \mathrm{d}x\ \mathcal{G}(x \mid k, s) = \int_0^\infty \mathrm{d}x\ \frac{1}{s\Gamma(k)}\left(\frac{x}{s}\right)^{k-1}\exp\left(-\frac{x}{s}\right) = 1$$

を示す必要があります。$y = x/s$ と変数変換して $\Gamma(k)$ の定義式と見比べることで

$$(\text{中辺}) = \frac{1}{\Gamma(k)} \int_0^\infty \mathrm{d}y\, y^{k-1} \mathrm{e}^{-y} = \frac{1}{\Gamma(k)} \times \Gamma(k) = 1$$

が分かります。

3.0.2

確率変数 x が $\mathcal{G}(k, s)$ に従うとしたら、x はどういうカイ 2 乗分布に従うことになるでしょうか。$x \sim \chi^2(k', s')$ と書いたとき、k', s' を求めて下さい。

[解答]

$$\begin{aligned}
\mathcal{G}(x \mid k, s) &= \frac{1}{s\Gamma(k)} \left(\frac{x}{s}\right)^{k-1} \exp\left(-\frac{x}{s}\right) \\
&= \frac{1}{2 \times (s/2) \times \Gamma(2k/2)} \left(\frac{x}{2 \times (s/2)}\right)^{(2k/2)-1} \exp\left(-\frac{x}{2 \times (s/2)}\right) \\
&= \chi^2(x \mid 2k, s/2)
\end{aligned}$$

したがって $k' = 2k, s' = s/2$。

3.0.3

経験分布

$$p_{\mathrm{emp}}(\boldsymbol{x}) = \frac{1}{N} \sum_{n=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}^{(n)})$$

が規格化条件を満たすことを示してください。

[解答] M 変数のデルタ関数は 1 変数のデルタ関数を使って

$$\delta(\boldsymbol{x} - \boldsymbol{x}') \equiv \prod_{l=1}^{M} \delta(x_l - x'_l)$$

のように定義されます。ただし $\boldsymbol{x}'$ は任意の M 次元定ベクトル、x'_l はその第 l 成分です。これから、M 変数のデルタ関数が規格化条件を満たすことが分かり

ます。

$$\begin{aligned}
&\int \mathrm{d}x_1 \cdots \mathrm{d}x_M \; \delta(\boldsymbol{x} - \boldsymbol{x}^{(n)}) \\
&\qquad = \int \mathrm{d}x_1 \; \delta(x_1 - x_1^{(n)}) \times \cdots \times \int \mathrm{d}x_M \; \delta(x_M - x_M^{(n)}) \\
&\qquad = 1 \times \cdots \times 1 \\
&\qquad = 1
\end{aligned}$$

この式を項別に適用することで、

$$\int \mathrm{d}\boldsymbol{x} \; p_{\mathrm{emp}}(\boldsymbol{x}) = \frac{1}{N} \sum_{n=1}^{N} 1 = 1$$

が導かれます。

3.0.4

混合正規分布の期待値-最大化法において、共分散行列をすべての k について $\boldsymbol{\Sigma}_k = \sigma^2 \mathsf{I}_M$ と固定します。この時、混合正規分布の期待値-最大化法が、$\sigma^2 \to 0$ において k 平均クラスタリングと等価であることを次の順番で示してください。

1. 帰属度の計算式 (3.54)

$$q_k^{(n)} = \frac{\hat{\pi}_k \, \mathcal{N}(\boldsymbol{x}^{(n)} | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k)}{\sum_{l=1}^{K} \hat{\pi}_l \, \mathcal{N}(\boldsymbol{x}^{(n)} | \hat{\boldsymbol{\mu}}_l, \hat{\boldsymbol{\Sigma}}_l)}$$

が、$q_k^{(n)} = \delta(k, k^{(n)})$ に帰着されることを示して下さい。ただし

$$k^{(n)} = \arg\max_k \mathcal{N}(\boldsymbol{x}^{(n)} \mid \hat{\boldsymbol{\mu}}_k, \sigma^2 \mathsf{I}_M)$$

です（arg は「変数（argument）を取り出せ」という意味です。この場合であれば「最大値を実現する k を取り出せ」という意味でになります）。I_M は M 次元単位行列です。

2. 上記が成り立つとき、期待値-最大化法が、式 (3.47)

$$L = \sum_{c=1}^{k} \left\{ \sum_{n=1}^{N} \delta(z^{(n)}, c) \, \|\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_c\|^2 \right\}$$

の最小化と等価であることを示して下さい。

[解答] 1. について。記号が煩雑なので、説明の都合上、標本 $\boldsymbol{x}^{(n)}$ を $\boldsymbol{x}$ という記号で代表させ、上付きの (n) を省略します。この前提で、$\mathcal{N}(\boldsymbol{x} \mid \hat{\boldsymbol{\mu}}_k, \sigma^2 \mathsf{I}_M)$ を最大化させる k が 1 だったとします。これは要するに $\|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_k\|^2$ が $k = 1$ の時に最小になっているということです（もし 1 でなかったら番号を付け替えればいいので、一般性を失っていません)。式 (3.54) において正規分布の定義式を入れると

$$\frac{q_k}{q_1} = \frac{\hat{\pi}_k}{\hat{\pi}_1} \exp\left[-\frac{\|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_k\|^2}{2\sigma^2} \left\{ 1 - \frac{\|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1\|^2}{\|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_k\|^2} \right\} \right]$$

ですが、$k \neq 1$ なら $\{\cdot\} > 0$ ですので、$\sigma^2 \to 0$ においてこの比はゼロとなります。したがって

$$\frac{q_k}{q_1} \propto \delta(1, k)$$

です。規格化条件を考えると $q_k = \delta(1, k)$ となることがわかります。

2. について。もし上記が成り立てば、各標本はかならず 1 つのクラスターにのみ帰属します。したがって、

$$\sum_{n=1}^{N} q_k^{(n)} = (\text{クラスター } k \text{ に属する標本の数}) \equiv N_k$$

となります。式 (3.55) より、

$$\pi_k = \frac{N_k}{N}, \quad \hat{\mu}_k = \frac{1}{N_k} \sum_{n:\text{クラスタ } k \text{ に帰属}} \boldsymbol{x}^{(n)}$$

となりますが、これは式 (3.45) と同じです。したがって、式 (3.47) の最小化をしているのと同じだということが分かります。

3.0.5

上記の問題において、共分散行列を対角行列とせずにある Σ に固定したと考えます。ここでも式 (3.73) を使って $q_k^{(n)} = \delta(k, k^{(n)})$ としたとすれば、k 平均クラスタリングの手法はどのように拡張されるでしょうか。

[解答] この場合、$k^{(n)}$ は次のように選ぶことになります。

$$k^{(n)} = \arg\max_k \mathcal{N}(\boldsymbol{x}^{(n)} \mid \hat{\boldsymbol{\mu}}_k, \Sigma)$$

これは

$$k^{(n)} = \arg\min_k (\boldsymbol{x}^{(n)} - \hat{\boldsymbol{\mu}}_k)^\top \Sigma^{-1} (\boldsymbol{x}^{(n)} - \hat{\boldsymbol{\mu}}_k)$$

と同じです。普通のユークリッド距離の代わりに、マハラノビス距離が最小になるように帰属クラスターを選ぶことになります。

3.0.6

N 個の独立な M 次元標本 $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ が与えられている時に、要素数を $K = N$ と選んだ混合正規分布モデルを考え、$\pi_k = 1/N$ および $\Sigma_k = \sigma^2 \mathsf{I}_M$ と固定します。すなわち、

$$p(\boldsymbol{x} \mid \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N) = \frac{1}{N} \sum_{k=1}^{N} \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \sigma^2 \mathsf{I}_M)$$

というモデルを考えます。$\sigma^2 \to 0$ における最尤解はどのようなものになるでしょうか。

[解答] N 個の独立標本が連続分布から出てきたものだと仮定します。この時 N 個の標本のどれも一致しないと仮定できます。この時、$\sigma^2 \to 0$ において、(3.74) は N 個の重なりのない山からなります。ゆえ、対数尤度

$$L(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N \mid \mathcal{D}) = \sum_{n=1}^{N} \ln \left\{ \frac{1}{N} \sum_{k=1}^{N} \mathcal{N}(\boldsymbol{x}^{(n)} \mid \boldsymbol{\mu}_k, \sigma^2 \mathsf{I}_M) \right\}$$

は、クラスターの番号の付け替え分の重複は別にして、

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}^{(1)}, \ \boldsymbol{\mu}_2 = \boldsymbol{\mu}^{(2)}, \ \ldots, \ \boldsymbol{\mu}_N = \boldsymbol{\mu}^{(N)}$$

の時に最大になります。

3.0.7

カーネル密度推定における平均積分 2 乗誤差

$$\int \mathrm{d}\boldsymbol{x}^{(1)} \cdots \mathrm{d}\boldsymbol{x}^{(N)}\ p_{\text{真}}(\boldsymbol{x}^{(1)}) \cdots p_{\text{真}}(\boldsymbol{x}^{(N)})\ E(h \mid \mathcal{D}) \approx \frac{1}{Nh} R(K) + \frac{h^4 \sigma_K^4}{4} R(p''_{\text{真}})$$

をバンド幅 h にて微分して 0 と等値することにより式 (3.40)

$$h^* = \left[\frac{R(K)}{\sigma_K^4 R(p''_{\text{真}})} \right]^{1/5} N^{-1/5}$$

を導出してみてください。

[解答] 平均積分 2 乗誤差の式を h で微分して 0 とおくと

$$0 = -\frac{1}{Nh^2} R(K) + h^3 \sigma_K^4 R(p''_{\text{真}})$$

したがって

$$h^5 \sigma_K^4 R(p''_{\text{真}}) = \frac{1}{N} R(K)$$

これから h^* の式はただちに導かれます。

3.0.8

混合正規分布モデルに対する BIC の式

$$\mathrm{BIC}_{\text{混合正規分布}} = -2L(\hat{\boldsymbol{\Theta}}|\mathcal{D}) + \frac{K}{2}(M+1)(M+2)\ln N$$

の第 2 項は混合正規分布のパラメター数の合計を表しています。この式の形を導いてください。

[解答] $\{\pi_k\}$ に K 個、$\{\boldsymbol{\mu}_k\}$ に KM 個は自明だと思います。$\boldsymbol{\Sigma}_k$ については、$M \times M$ 対称行列は、対角成分が M 個、非対角成分の上半分が、

$$ {}_M\mathrm{C}_2 = \frac{1}{2}M(M-1) $$

個あります。したがって

$$ (\Sigma_k\text{のパラメター数}) = M + \frac{1}{2}M(M-1) = \frac{M}{2} + \frac{M^2}{2} = \frac{M}{2}(M+1) $$

です。Σ_k は全部で K 個あるので、以上全部まとめると、このモデルに含まれるパラメター数は

$$ \begin{aligned} K + KM + \frac{KM}{2}(M+1) &= K\left\{1 + M + \frac{M}{2}(M+1)\right\} \\ &= K(M+1)\left(1 + \frac{M}{2}\right) \\ &= \frac{K}{2}(M+1)(M+2) \end{aligned} $$

となります。これに BIC の定義に由来する $\ln N$ をつけたものが混合正規分布の BIC の式の第 2 項です。

なお、理論上は、混合正規分布はいわゆる特異モデルですので、BIC の導出で用いたラプラス近似は成り立ちません。しかし経験的には、BIC は、混合正規分布のクラスター数の見積もりにおいて、多くの場合に実用上妥当な目安を与えると言われています。

4 性能評価の方法

4.0.1

F 値

$$f \equiv \frac{2r_1 r_2}{r_1 + r_2} \tag{4.1}$$

の代わりに、単純平均 $(r_1 + r_2)/2$ を総合的な指標として使うのが望ましくない理由は何でしょうか。

[解答] 例えば正常標本精度が $r_1 = 0.99$、異常標本精度が $r_2 = 0.02$ だったとします。単純平均だと、$(r_1 + r_2)/2 = 0.505$ という値になり、「半分くらい合っている」という印象を与えてしまいます。しかし、実際のところ、異常標本の 98 パーセントは見逃されてしまうので、こういう印象を与えてしまうのは実用上問題があります。一方、F 値だと 0.0392... で、いかにもダメな感じがします。単純平均の最大の欠点は、r_1 と r_2 の一方が壊滅的に低くても、単純平均としては低くなるとは限らないという点です。一方、F 値のよい点は、r_1 と r_2 の一方が壊滅的に低いと、正しくそれを反映したスコアを与えてくれる点です。

4.0.2

4.3 節の例題で、実際に F 値を計算してみてください。

[解答] 下記のようになると思います。

表 4.1 4.3.1 項の例題における F 値

異常標本精度	正常標本精度	F 値
1	0	0
1	0.1428571	0.25
1	0.2857143	0.4444444
1	0.4285714	0.6
1	0.5714286	0.7272727
1	0.7142857	0.8333333
1	0.8571429	0.9230769
0.6666667	0.8571429	0.75
0.6666667	1	0.8
0.3333333	1	0.5
0	1	0

4.0.3

正常と異常のラベルがつけられた N 個の標本を含む訓練データ $\mathcal{D}$ を使って k 近傍法による異常検知手法を構成することを考えます。一般に、新たな観測値 $\boldsymbol{x}'$ の異常を、ラベルつきデータを使って判定するためには、次の**対数尤度比**と呼ばれる量を異常度として使うのがある意味で最適であることが知られています（**ネイマン=ピアソンの補題**）。

$$a(\boldsymbol{x}') = \ln \frac{p(\boldsymbol{x}' \mid y = +1)}{p(\boldsymbol{x}' \mid y = -1)}$$

ただし $p(\boldsymbol{x} \mid y = +1)$ は、異常標本に対する $\boldsymbol{x}$ の確率密度関数で、$p(\boldsymbol{x} \mid y = -1)$ は正常標本に対する確率密度関数です。k 近傍法を使ってこれらを推定するためにはどのようにすればよいでしょうか。

[**解答**] 井手剛・杉山将『異常検知と変化検知』（講談社、2015 年 8 月）の 4.1 節参照。

4.0.4

k 近傍法で異常検知をする際、近傍数 k は、ひとつ抜き交差検証法において、F 値を最大にするように決定するのが妥当な方法のひとつです。その手順を書いてみて下さい。

[解答] 井手剛・杉山将『異常検知と変化検知』(講談社、2015 年 8 月) の 4.1 節参照。

4.0.5

2.5.1 項の定理 2.7 を用いて、独立に $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ に従う N 個の標本から作られる標本平均 $\hat{\boldsymbol{\mu}} = \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \boldsymbol{x}^{(n)}$ の従う確率分布を求めて下さい。また、その結果を用いて

$$\sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

を証明してください。同様の結果が (ゆるい条件の下) 任意の分布に従う標本集合について成り立ち、中心極限定理と呼ばれています。

[解答] 定理 2.7 において、$\mathsf{A} = \mathsf{B} = \dfrac{1}{N}\mathsf{I}_M$ としたものを繰り返し使うと

$$\left(\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}^{(n)}\text{の平均}\right) = \boldsymbol{\mu}$$
$$\left(\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}^{(n)}\text{の共分散}\right) = \frac{1}{N}\Sigma$$

のようになります。従って、

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}^{(n)} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}} \mid \boldsymbol{\mu}, \frac{1}{N}\Sigma\right)$$

が得られます。

定理 A.2 の (A.7) において、$\mathsf{T} = \sqrt{N}\mathsf{I}_M$、$\boldsymbol{b} = -\sqrt{N}\boldsymbol{\mu}$ とおくと、$\boldsymbol{y} \equiv \sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ の従う分布が

$$\boldsymbol{y} \sim \left|\sqrt{N}\mathsf{I}_M\right|^{-1} \mathcal{N}\left(\frac{1}{\sqrt{N}}\boldsymbol{y} + \boldsymbol{\mu}\right)$$

となることが分かります。ここで

$$\left|\sqrt{N}\mathsf{I}_M\right|^{-1} = N^{-M/2}, \quad \left|\frac{1}{N}\Sigma\right|^{1/2} = |\Sigma|^{1/2}\, N^{-M/2}$$

に注意して、正規分布の定義を使って明示的に $\left|\sqrt{N}\mathsf{I}_M\right|^{-1} \mathcal{N}\left(\frac{1}{\sqrt{N}}\boldsymbol{y} + \boldsymbol{\mu}\right)$ を書き下すと

$$\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

がわかります。

4.0.6

スカラー変数 θ の関数 $\ell(\theta)$ が $\hat{\theta}$ にて最大値を持ち、その近傍で 2 回微分可能だとします。今、定数 C を $C \equiv -\left.\frac{\mathrm{d}^2\ell}{\mathrm{d}\theta^2}\right|_{\theta=\hat{\theta}}$ により定義すると、これは仮定より正値となります。この時、ガウス積分に対する定理 A.6 を使い、$N \to \infty$ において次の近似式が成り立つことを証明してください。

$$\int \mathrm{d}\theta\, \mathrm{e}^{N\ell(\theta)} \approx \sqrt{\frac{2\pi}{NC}} \mathrm{e}^{N\ell(\hat{\theta})}$$

これは**ラプラス近似**と呼ばれる結果で、BIC の導出において使われます。

[解答]

$$\int_{-\infty}^{\infty} \mathrm{d}\theta\, \mathrm{e}^{N\ell(\theta)} = \mathrm{e}^{N\ell(\hat{\theta})} \int_{-\infty}^{\infty} \mathrm{d}\theta\, \mathrm{e}^{N[\ell(\theta)-\ell(\hat{\theta})]}$$

において、非積分関数は、$\theta \approx \hat{\theta}$ においては、θ に関するテイラー展開で 2 次まで考えて

$$\begin{aligned}\mathrm{e}^{N[\ell(\theta)-\ell(\hat{\theta})]} &\approx \exp\left[N\left\{\ell'(\hat{\theta})(\theta-\hat{\theta}) - \frac{NC}{2}(\theta-\hat{\theta})^2\right\}\right] \\ &= \exp\left[-\frac{NC}{2}(\theta-\hat{\theta})^2\right]\end{aligned}$$

と近似できます。ここで 1 階導関数に関する条件 $\ell'(\hat{\theta}) \equiv \left.\frac{d\ell}{d\theta}\right|_{\theta=\hat{\theta}} = 0$ を使いました。

$\theta \approx \hat{\theta}$ 以外の領域ではテイラー展開の正確性は保証されませんが、幸い、$\theta \neq \hat{\theta}$ なら左辺の指数関数の肩において $[\cdot] < 0$ ですから、$N \to \infty$ なら

$$\mathrm{e}^{N[\ell(\theta)-\ell(\hat{\theta})]} \approx 0$$

がテイラー展開せずとも分かり、なおかつ、これを $\theta \approx \hat{\theta}$ 以外の領域で積分したものは（たとえ $|\theta - \hat{\theta}| \to \infty$ の果てまで積分したとしても）0 です。

それゆえ結局、上記テイラー展開が θ の全領域で妥当だと思って差し支えありません。すなわち、$N \to \infty$ である限り

$$\int_{-\infty}^{\infty} \mathrm{d}\theta\, \mathrm{e}^{N\ell(\theta)} \approx \mathrm{e}^{N\ell(\hat{\theta})} \int_{-\infty}^{\infty} \mathrm{d}\theta\, \exp\left[-\frac{NC}{2}(\theta - \hat{\theta})^2\right]$$

が成り立ち、ガウス積分についての定理 A.6 を使うことで、

$$\int \mathrm{d}\theta\, \mathrm{e}^{N\ell(\theta)} \approx \sqrt{\frac{2\pi}{NC}} \mathrm{e}^{N\ell(\hat{\theta})}$$

が成り立つことが分かります。

5 不要な次元を含むデータからの異常検知

5.0.1

任意の対角行列の列ベクトルが直交することを証明してください。また、対角要素が a, b, c という非ゼロの実数で与えられる 3 次元の対角行列の逆行列を求めて下さい。

[解答] 一般化するまでもないので問題にある 3 次元行列で例証します。

$$\mathsf{A} \equiv \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

とすると、第 1 列と第 2 列の内積は、

$$a \times 0 + 0 \times b + 0 \times 0 = 0$$

となります。第 2 列と 3 列も同様です。また、

$$\mathsf{B} \equiv \begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{pmatrix}$$

とおくと、行列の積の定義から、$\mathsf{AB} = \mathsf{BA} = \mathsf{I}_3$ であることがわかりますので、$\mathsf{B} = \mathsf{A}^{-1}$ です。

5.0.2

M 次元正方行列 U が、$\mathsf{U}^\top\mathsf{U} = \mathsf{U}\mathsf{U}^\top = \mathsf{I}_M$ を満たすとき直交行列と呼びます。U の列ベクトルが互いに正規直交すること、また、U の行ベクトルも互いに正規直交することを証明してください。

[解答] $\mathsf{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M]$ とおきます。条件 $\mathsf{U}^\top\mathsf{U} = \mathsf{I}_M$ というのは、(i, j) 成分が

$$\boldsymbol{u}_i^\top \boldsymbol{u}_j = \delta_{i,j}$$

になるということですから、$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ が正規直交していることが分かります。

次に、$\mathsf{U}^\top = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M]$ と置きます。今度は条件 $\mathsf{U}\mathsf{U}^\top = \mathsf{I}_M$ を考えると、(i, j) 成分が

$$\boldsymbol{v}_i^\top \boldsymbol{v}_j = \delta_{i,j}$$

になるということですから、$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$ が正規直交していることが分かります。

5.0.3

任意の $M \times N$ 行列 X の第 n 列ベクトルを $\boldsymbol{x}_n$ とする時、$\mathsf{X}\mathsf{X}^\top = \sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top$ が成り立つことを証明してください。

[解答] $\mathsf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ ですので、$\boldsymbol{x}_i$ の第 k 成分 $[\boldsymbol{x}_i]_k$ は $\mathsf{X}_{k,i}$ のことです。行列の積の定義から $\mathsf{X}\mathsf{X}^\top$ の (k, l) 成分は

$$[\mathsf{X}\mathsf{X}^\top]_{k,l} = \sum_{i=1}^{M} \mathsf{X}_{k,i}\mathsf{X}_{l,i} = \sum_{i=1}^{M} [\boldsymbol{x}_i]_k \times [\boldsymbol{x}_i]_l$$

となります。一方 $\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top$ の (k, l) 成分は

$$\left[\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^\top\right]_{k,l} = \sum_{n=1}^{N} [\boldsymbol{x}_n \boldsymbol{x}_n^\top]_{k,l} = \sum_{n=1}^{N} [\boldsymbol{x}_n]_k \times [\boldsymbol{x}_n]_l$$

ですので一致が分かります。

5.0.4

行列 A の 2 ノルムの定義 (5.72) が、$\mathsf{A}^\top\mathsf{A}$ の固有値方程式と等価であることを証明してください。

[解答] 最適化問題

$$\max_{\boldsymbol{\varphi}} \frac{\|\mathsf{A}\boldsymbol{\varphi}\|_p}{\|\boldsymbol{\varphi}\|_p}$$

は

$$\max_{\boldsymbol{\varphi}} \|\mathsf{A}\boldsymbol{\varphi}\|_2 \quad \text{subject to} \quad \|\boldsymbol{\varphi}\|_2 = 1$$

と同じで、これは

$$\max_{\boldsymbol{\varphi}} \|\mathsf{A}\boldsymbol{\varphi}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\varphi}\|_2^2 = 1$$

と同じです。さらにこれは

$$\max_{\boldsymbol{\varphi}} \boldsymbol{\varphi}^\top\mathsf{A}^\top\mathsf{A}\boldsymbol{\varphi} \quad \text{subject to} \quad \boldsymbol{\varphi}^\top\boldsymbol{\varphi} = 1$$

とも同じです。ラグランジュ乗数 λ を使うと、最適性の条件は

$$\mathbf{0} = \frac{1}{2}\frac{\partial}{\partial\boldsymbol{\varphi}}\left\{\boldsymbol{\varphi}^\top\mathsf{A}^\top\mathsf{A}\boldsymbol{\varphi} - \lambda\boldsymbol{\varphi}^\top\boldsymbol{\varphi}\right\} = \mathsf{A}^\top\mathsf{A}\boldsymbol{\varphi} - \lambda\boldsymbol{\varphi}$$

これより

$$\mathsf{A}^\top\mathsf{A}\boldsymbol{\varphi} = \lambda\boldsymbol{\varphi}$$

が得られます。これを解いて固有値が大きい順に $\lambda_1, \lambda_2, \ldots$ と求まったとします。λ_i に対応する正規化された固有ベクトルを $\boldsymbol{\varphi}_i$ とすれば、

$$\frac{\|\mathsf{A}\boldsymbol{\varphi}_i\|_2}{\|\boldsymbol{\varphi}_i\|_2} = \sqrt{\lambda_i}$$

ですので、行列 2 ノルムとしては $\mathsf{A}^\top\mathsf{A}$ の最大固有値の平方根を選べばいいということになります。これは A の最大特異値と同じです。

5.0.5

実対称行列 A が重複のない最大固有値を持つとし、対応する固有ベクトルを $\boldsymbol{u}_1$ とします。$\boldsymbol{u}_1^\top \boldsymbol{z} \neq 0$ を満たす任意の M 次元単位ベクトル $\boldsymbol{z}$ を考え

$$\boldsymbol{z} \leftarrow \mathsf{A}\boldsymbol{z}$$

$$\boldsymbol{z} \leftarrow \boldsymbol{z}/\|\boldsymbol{z}\|_2$$

という計算を K 回繰り返します。$K \to \infty$ の時、$\boldsymbol{z}$ が A の最大固有値に属する規格化された固有ベクトルになっていることを証明してください。ちなみにこれは数値計算において**べき乗法**として知られる計算手法になっています。

[**解答**] 実対称行列の固有値は実数です。大きい順からその固有値を $\lambda_1, \lambda_2, \ldots$ とし、規格化された固有ベクトルを $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ とします。固有値分解の定義から

$$\mathsf{A} = \sum_i \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$$

とできるので、固有ベクトルの正規直交性から

$$\mathsf{A}^K \boldsymbol{z} = \sum_i \lambda_i^K \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{z} = \lambda_1^K \sum_i \left(\frac{\lambda_i}{\lambda_1}\right)^K \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{z}$$

となります。したがって、$K \to \infty$ において

$$\mathsf{A}^K \boldsymbol{z} \to \boldsymbol{u}_1 \boldsymbol{u}_1^\top \boldsymbol{z}$$

となります。右辺を規格化すると $\boldsymbol{u}_1$ そのものです。これで問題が証明できました。

5.0.6

付録の定理 A.9 を使って式 (5.45) を導出してみて下さい。

[**解答**] $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{0}, \mathsf{I}_m)$ および $p(\boldsymbol{x} \mid \boldsymbol{z}) = \mathcal{N}(\boldsymbol{x} \mid \mathsf{W}\boldsymbol{z} + \bar{\boldsymbol{x}}, \sigma^2 \mathsf{I}_M)$ を使って $p(\boldsymbol{z} \mid \boldsymbol{x})$ を求めるのが問題です。定理 A.9 の (A.54) において

$$\boldsymbol{\mu} \leftarrow \mathbf{0}, \Sigma \leftarrow \mathsf{I}_m, \mathsf{A} \leftarrow \mathsf{W}, \boldsymbol{b} \leftarrow \bar{\boldsymbol{x}}, \mathsf{D} \leftarrow \sigma^2 \mathsf{I}_m$$

とおくと

$$p(\boldsymbol{z} \mid \boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{z} \mid \mathsf{R}\mathsf{W}^\top \frac{1}{\sigma^2}(\boldsymbol{x} - \bar{\boldsymbol{x}}), \mathsf{R}\right)$$

ただし

$$\mathsf{R} = \left[\mathsf{W}^\top \frac{1}{\sigma^2}\mathsf{W} + \mathsf{I}_m\right]^{-1} = \sigma^2[\mathsf{W}^\top\mathsf{W} + \sigma^2\mathsf{I}_m]^{-1}$$

したがって

$$p(\boldsymbol{z} \mid \boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{z} \mid [\mathsf{W}^\top\mathsf{W} + \sigma^2\mathsf{I}_m]^{-1}\mathsf{W}^\top(\boldsymbol{x} - \bar{\boldsymbol{x}}), \sigma^2[\mathsf{W}^\top\mathsf{W} + \sigma^2\mathsf{I}_m]^{-1}\right)$$

5.0.7

任意の $M \times N$ 行列 F について、$M \times r$ 行列 C と、$N \times r$ 行列 H を用いて、$\mathsf{F} \approx \mathsf{C}\mathsf{H}^\top$ のような近似式を作ります。最適化問題

$$(\mathsf{H}^*, \mathsf{C}^*) = \arg\min_{\mathsf{C},\mathsf{H}} \|\mathsf{F} - \mathsf{C}\mathsf{H}^\top\|_{\mathrm{F}}^2 \ \text{ subject to } \ \mathsf{H}^\top\mathsf{H} = \mathsf{I}_r \qquad (5.1)$$

を解くことで、特異値分解が、階数 r の条件の下での最善の近似となっていることを証明して下さい。この結果はしばしば**エッカート＝ヤングの定理**と呼ばれます。

[**解答**] 行列 H を $[\boldsymbol{h}_1, \ldots, \boldsymbol{h}_r]$ とおくと、これら列ベクトルは正規直交します。この問題の制約条件をラグランジュ乗数で取り込むとすると、関係

$$\sum_{i,j=1}^{r} \mathsf{L}_{i,j}(\boldsymbol{h}_i^\top\boldsymbol{h}_j - \delta_{i,j}) = \mathrm{Tr}(\mathsf{L}(\mathsf{H}^\top\mathsf{H} - \mathsf{I}_r))$$

から、対称行列 L をラグランジュ乗数と考えればよいことが分かります。I_r は r 次元の単位行列。したがって、最小化すべき目的関数は次のようになります。

$$\|\mathsf{F} - \mathsf{C}\mathsf{H}^\top\|_{\mathrm{F}}^2 + \mathrm{Tr}(\mathsf{L}(\mathsf{H}^\top\mathsf{H} - \mathsf{I}_r))$$

この目的関数は、C, H に無関係な定数を除くと、条件 $\mathsf{H}^\top\mathsf{H} = \mathsf{I}_r$ を使うことにより

$$J_1(\mathsf{C},\mathsf{H}) \equiv \mathrm{Tr}(\mathsf{C}^\top\mathsf{C} - 2\mathsf{C}^\top\mathsf{F}\mathsf{H}) + \mathrm{Tr}(\mathsf{L}(\mathsf{H}^\top\mathsf{H} - \mathsf{I}_r))$$

に帰着できます。あとの都合上、

$$J(\mathsf{C},\mathsf{H}) \equiv \mathrm{Tr}(\mathsf{C}^\top\mathsf{C} - 2\mathsf{C}^\top\mathsf{F}\mathsf{H})$$

と置いておきます。

最適解の条件は次のようになります。

$$\mathsf{0} = \frac{\partial J_1}{\partial \mathsf{C}} = 2\mathsf{C} - 2\mathsf{F}\mathsf{H}$$
$$\mathsf{0} = \frac{\partial J_1}{\partial \mathsf{H}} = -2\mathsf{F}^\top\mathsf{C} + 2\mathsf{H}\mathsf{L}$$

これから容易に次式が導けます。これが未知行列 H,C の満たすべき方程式です。

$$\mathsf{F}^\top\mathsf{F}\mathsf{H} = \mathsf{H}\mathsf{L}$$
$$\mathsf{F}\mathsf{F}^\top\mathsf{C} = \mathsf{C}\mathsf{L}$$

ついでに、元の最適解の条件の最初の式から $\mathsf{C} = \mathsf{F}\mathsf{H}$ が得られ、これと、元の最適解の条件の第 2 の式に左から $\mathsf{H}^\top$ をかけた式を連立させると、$\mathsf{L} = \mathsf{C}^\top\mathsf{C}$ が得られることが分かります。したがって

$$J(\mathsf{C},\mathsf{H}) = -\mathrm{Tr}(\mathsf{L})$$

です。

上に導いた最適解の条件式は、$\mathsf{F}^\top\mathsf{F}$ と $\mathsf{F}\mathsf{F}^\top$ についての固有値方程式に似た形をしていますが、L は一般には対角行列ではありませんので、このままでは固有値方程式とは言えません。しかし次のようにして L を対角行列として仮定しても一般性を失わないことが分かります。まず、J は、直交行列 Q により

$$\mathsf{H} \leftarrow \mathsf{H}\mathsf{Q}^\top, \quad \mathsf{C} \leftarrow \mathsf{C}\mathsf{Q}^\top,$$

と変換しても影響を受けません。したがって未知行列 H,C には上記の変換の任意性があります。それをどう選ぶかは自由です。ここで J_1 がこの変換により

$$J_1(\mathsf{C},\mathsf{H}) \leftarrow \mathrm{Tr}(\mathsf{C}^\top\mathsf{C} - 2\mathsf{C}^\top\mathsf{F}\mathsf{H}) + \mathrm{Tr}(\tilde{\mathsf{L}}(\mathsf{H}^\top\mathsf{H} - \mathsf{I}_r))$$

となることに注目します。ただし

$$\tilde{\mathsf{L}} \equiv \mathsf{Q}^\top\mathsf{L}\mathsf{Q}$$

と置きました。ここで Q を、L を対角化するように選んだとします。L は定数ではなくて最適化の結果として出てくるものですが、どういう解が得られるにせよ、L を対角化するように Q を選ぶことは可能です。したがって、最初から L が対角行列と思っても同じことです。

ということで、未知行列 H,C は、それぞれ $\mathsf{F}^\top\mathsf{F}$ と $\mathsf{F}\mathsf{F}^\top$ の固有ベクトルを列ベクトルとして並べたものになることが分かりました。L は対角行列で、その対角要素には固有値が入ります。これは、未知行列 H が F の右特異ベクトル、C が同じく左特異ベクトルであることを意味します。

さらに $J(\mathsf{C},\mathsf{H}) = -\mathrm{Tr}(\mathsf{L})$ であることから、目的関数の最小化のためには大きい順から r 個の特異ベクトルを選べばよいことがわかります。

6 入力と出力があるデータからの異常検知

6.0.1

普通の最小 2 乗法による線形回帰で

$$\boldsymbol{\phi} = \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}$$

により新しい $M+1$ 次元の入力変数 $\boldsymbol{\phi}$ を定義し、それに対応して係数ベクトルを

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha_0 \\ \boldsymbol{\alpha} \end{pmatrix}$$

と置きます。この時、対数尤度の式 (6.6)

$$L(\alpha_0, \boldsymbol{\alpha}, \sigma^2 \mid \mathcal{D}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left[y^{(n)} - \alpha_0 - \boldsymbol{\alpha}^\top \boldsymbol{x}^{(n)}\right]^2$$

はどのように変わるでしょうか。

[解答] $\boldsymbol{\beta}^\top \boldsymbol{\phi} = \alpha_0 + \boldsymbol{\alpha}^\top \boldsymbol{x}$ となるので、

$$L(\alpha_0, \boldsymbol{\alpha}, \sigma^2 \mid \mathcal{D}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left[y^{(n)} - \boldsymbol{\beta}^\top \boldsymbol{\phi}^{(n)}\right]^2$$

となります。$\boldsymbol{\beta}$ で L を微分して 0 とおくことで、α_0 と $\boldsymbol{\alpha}$ の解 (6.7) と (6.13) が得られます。

6.0.2

任意の M 次元ベクトル N 本、$\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(N)}$ と、N 個のスカラー定数 $w_1, \ldots, w_N$ に対して、次の式が成り立つことを示して下さい。

$$\sum_{n=1}^{N} w_n {\boldsymbol{z}^{(n)}}^{\top} \boldsymbol{z}^{(n)} = \mathrm{Tr}\left(\mathsf{Z}\mathsf{W}\mathsf{Z}^{\top}\right)$$
$$\sum_{n=1}^{N} w_n \boldsymbol{z}^{(n)} {\boldsymbol{z}^{(n)}}^{\top} = \mathsf{Z}\mathsf{W}\mathsf{Z}^{\top}$$

ただし、$\mathsf{Z} \equiv [\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(N)}]$、$\mathsf{W} = \mathrm{diag}(w_1, \ldots, w_N)$ です。

[解答] 行列の積とトレースの定義に従い、素朴に要素を比較します。最初の式に関しては、

$$\begin{aligned}
(\text{左辺}) &= \sum_{n=1}^{N} w_n {\boldsymbol{z}^{(n)}}^{\top} \boldsymbol{z}^{(n)} \\
&= \sum_{n=1}^{N} w_n \sum_{i=1}^{M} \mathsf{Z}_{i,n}^2 \\
(\text{右辺}) &= \sum_{i=1}^{M} \left(\sum_{n=1}^{N} \sum_{n'=1}^{N} \mathsf{Z}_{i,n} \mathsf{W}_{n,n'} \mathsf{Z}_{i,n'} \right) \\
&= \sum_{i=1}^{M} \left(\sum_{n=1}^{N} w_n \mathsf{Z}_{i,n} \mathsf{Z}_{i,n} \right) \\
&= \sum_{i=1}^{M} \left(\sum_{n=1}^{N} w_n \mathsf{Z}_{i,n}^2 \right)
\end{aligned}$$

なので両辺は一致します。ただし $\mathsf{W}_{n,j} = \delta_{n,j} w_n$ を使いました。

次の式については

$$
\begin{aligned}
(\text{左辺の } (i,j) \text{ 成分}) &= \sum_{n=1}^{N} w_n z_i^{(n)} z_j^{(n)} \\
&= \sum_{n=1}^{N} w_n \mathsf{Z}_{i,n} \mathsf{Z}_{j,n} \\
(\text{右辺の } (i,j) \text{ 成分}) &= \sum_{n=1}^{N} \sum_{n'=1}^{N} \mathsf{Z}_{i,n} \mathsf{W}_{n,n'} \mathsf{Z}_{j,n'} \\
&= \sum_{n=1}^{N} w_n \mathsf{Z}_{i,n} \mathsf{Z}_{j,n}
\end{aligned}
$$

とやはり一致します。

6.0.3

普通の最小 2 乗法による線形回帰で、N 個の標本それぞれに異なる定数の重み w_n を付与するモデルを考えます。この時、回帰係数 $\boldsymbol{\alpha}$ を求める問題は

$$
\|\tilde{\boldsymbol{y}}_N - \tilde{\mathsf{X}}^\top \boldsymbol{\alpha}\|^2 = \sum_{n=1}^{N} \left[y^{(n)} - \bar{y} - \boldsymbol{\alpha}^\top (\boldsymbol{x}^{(n)} - \bar{\boldsymbol{x}}) \right]^2 \ \rightarrow \ \text{最小化}
$$

の代わりに

$$
\sum_{n=1}^{N} w_n \left[y^{(n)} - \bar{y} - \boldsymbol{\alpha}^\top (\boldsymbol{x}^{(n)} - \bar{\boldsymbol{x}}) \right]^2 \ \rightarrow \ \text{最小化}
$$

のようになります。$\boldsymbol{\alpha}$ で微分して 0 と等置することで、最尤解 $\hat{\boldsymbol{\alpha}}$ を、式 (6.13) と同様の行列表現にて求めて下さい。ヒント: 前の問題の W を使います。

[解答]

$$
\sum_{n=1}^{N} w_n \left[y^{(n)} - \bar{y} - \boldsymbol{\alpha}^\top (\boldsymbol{x}^{(n)} - \bar{\boldsymbol{x}}) \right]^2
$$

は

$$
(\tilde{\boldsymbol{y}}_N - \tilde{\mathsf{X}}^\top \boldsymbol{\alpha})^\top \mathsf{W} (\tilde{\boldsymbol{y}}_N - \tilde{\mathsf{X}}^\top \boldsymbol{\alpha})
$$

すなわち

$$\tilde{\boldsymbol{y}}_N^\top \mathsf{W} \tilde{\boldsymbol{y}}_N - 2\boldsymbol{\alpha}^\top \tilde{\mathsf{X}} \mathsf{W} \tilde{\boldsymbol{y}}_N + \boldsymbol{\alpha}^\top \tilde{\mathsf{X}} \mathsf{W} \tilde{\mathsf{X}}^\top \boldsymbol{\alpha}$$

となります。$\boldsymbol{\alpha}$ で微分してゼロベクトルと等置すると

$$\mathbf{0} = -2\tilde{\mathsf{X}} \mathsf{W} \tilde{\boldsymbol{y}}_N + 2\tilde{\mathsf{X}} \mathsf{W} \tilde{\mathsf{X}}^\top \boldsymbol{\alpha}$$

が得られるので、結局

$$\hat{\boldsymbol{\alpha}} = [\tilde{\mathsf{X}} \mathsf{W} \tilde{\mathsf{X}}^\top]^{-1} \tilde{\mathsf{X}} \mathsf{W} \tilde{\boldsymbol{y}}_N$$

6.0.4

リッジ回帰の最適化問題

$$\|\tilde{\boldsymbol{y}}_N - \tilde{\mathsf{X}}^\top \boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^\top \boldsymbol{\alpha} \quad \rightarrow \quad \text{最小}$$

において、左辺の第 2 項を、$\lambda \boldsymbol{\alpha}^\top \boldsymbol{\alpha}$ の代わりに、ある M 次元正方行列 L を使って $\lambda \boldsymbol{\alpha}^\top \mathsf{L} \boldsymbol{\alpha}$ のように置き換えたとします。この時、リッジ解はどのように変わるでしょうか。

[解答] $\|\tilde{\boldsymbol{y}}_N - \tilde{\mathsf{X}}^\top \boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^\top \mathsf{L} \boldsymbol{\alpha}$ を $\boldsymbol{\alpha}$ で微分して 0 と置いた結果として

$$\hat{\boldsymbol{\alpha}} = [\tilde{\mathsf{X}} \tilde{\mathsf{X}}^\top + \lambda \mathsf{L}]^{-1} \tilde{\mathsf{X}} \tilde{\boldsymbol{y}}_N$$

6.0.5

式 (6.44) と (6.45) を導出してみて下さい。

[解答] 式 (6.42) に左から $\Sigma_{yy}^{-1/2}$ をかけると

$$\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \boldsymbol{\beta} = \lambda^2 \Sigma_{yy}^{1/2} \boldsymbol{\beta}$$

となります。ここで

$$\mathsf{W} \equiv \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2}, \quad \tilde{\boldsymbol{\beta}} \equiv \Sigma_{yy}^{1/2} \boldsymbol{\beta}$$

とおくと、右辺は $\lambda^2\tilde{\boldsymbol{\beta}}$、左辺は $\mathsf{W}^\top\mathsf{W}\tilde{\boldsymbol{\beta}}$ となります。

また、(6.43) に左から $\boldsymbol{\Sigma}_{xx}^{-1/2}$ をかけると、$\tilde{\boldsymbol{\alpha}} \equiv \boldsymbol{\Sigma}_{xx}^{1/2}\boldsymbol{\alpha}$ に対して、右辺は $\lambda^2\tilde{\boldsymbol{\alpha}}$ となり、左辺は、

$$\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\alpha} = \mathsf{W}\mathsf{W}^\top\tilde{\boldsymbol{\alpha}}$$

となりますので、結局次が導かれます。

$$\mathsf{W}\mathsf{W}^\top\tilde{\boldsymbol{\alpha}} = \lambda^2\tilde{\boldsymbol{\alpha}} \quad \text{subject to} \quad \tilde{\boldsymbol{\alpha}}^\top\tilde{\boldsymbol{\alpha}} = 1$$

$$\mathsf{W}^\top\mathsf{W}\tilde{\boldsymbol{\beta}} = \lambda^2\tilde{\boldsymbol{\beta}} \quad \text{subject to} \quad \tilde{\boldsymbol{\beta}}^\top\tilde{\boldsymbol{\beta}} = 1$$

7 時系列データの異常検知

7.0.1

M 次元の観測値が T 個 $\{\boldsymbol{\xi}^{(1)},\ldots,\boldsymbol{\xi}^{(T)}\}$ のように得られている時、過去に行くほど寄与が小さくなるように平均値の計算法を工夫することを考えます。$0<\gamma\leqq 1$ に対して標本平均を $\bar{\boldsymbol{\xi}}_T \equiv \dfrac{1}{Z}\displaystyle\sum_{t=1}^{T}\gamma^{T-t}\boldsymbol{\xi}^{(t)}$ と書いたとき、次の条件を満たすように Z の式を求めて下さい。（１）$\gamma=1$ の時に通常の標本平均に一致すること。（２）$\boldsymbol{\xi}^{(t)}$ が t によらず一定値 $\boldsymbol{\xi}$ をとるとき、平均がその値に一致すること。

[解答]

条件（２）を考えると

$$\boldsymbol{\xi}=\frac{1}{Z}\sum_{t=1}^{T}\gamma^{T-t}\boldsymbol{\xi}$$

なので、

$$1=\frac{1}{Z}\sum_{t=1}^{T}\gamma^{T-t}$$

が成り立つ必要があります。これより

$$Z=\sum_{t=1}^{T}\gamma^{T-t}$$

となり、これは条件（１）も満たします。

7.0.2

上の結果を利用して、時間ごとに減衰する重みを入れた共分散行列を求めることを考えます。$w_t \equiv \gamma^{T-t}/Z$ とおき、時刻 T における共分散行列を

$$\Sigma_T \equiv \sum_{t=1}^{T} w_t(\boldsymbol{\xi}^{(t)} - \bar{\boldsymbol{\xi}}_T)(\boldsymbol{\xi}^{(t)} - \bar{\boldsymbol{\xi}}_T)^\top$$

のように定義します。この時、式 (6.74) を用いて、上式の行列表現が

$$\Sigma_T = \mathsf{X}_T(\mathsf{W}_T - \boldsymbol{w}_T\boldsymbol{w}_T^\top)\mathsf{X}_T^\top$$

となることを示して下さい。ただし、$\boldsymbol{w}_T \equiv (w_1, \ldots, w_T)^\top$、$\mathsf{W}_T \equiv \mathrm{diag}(\boldsymbol{w}_T)$、$\mathsf{X}_T \equiv [\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(T)}]$ とおきました。

[解答] 標本平均について $\bar{\boldsymbol{\xi}} = \sum_{t=1}^{T} w_t\boldsymbol{\xi}^{(t)} = \mathsf{X}_T\boldsymbol{w}_T$ が成り立つので

$$\begin{aligned}\Sigma_T &= \sum_{t=1}^{T} w_t\boldsymbol{\xi}^{(t)}{\boldsymbol{\xi}^{(t)}}^\top - \bar{\boldsymbol{\xi}}_T\bar{\boldsymbol{\xi}}_T^\top \\ &= \mathsf{X}_T\mathsf{W}\mathsf{X}_T^\top - (\mathsf{X}_T\boldsymbol{w}_T)(\mathsf{X}_T\boldsymbol{w}_T)^\top \\ &= \mathsf{X}_T(\mathsf{W} - \boldsymbol{w}_T\boldsymbol{w}_T^\top)\mathsf{X}_T^\top\end{aligned}$$

が言えます。

7.0.3

次数 r を 0 と置いたベクトル自己回帰モデルによる異常検知が、ホテリング理論と等価であることを示して下さい。

[解答] この場合係数行列が $\mathsf{A} = 0$ を満たすはずですので、式 (7.24) においてそのようにおくと

$$a_M(\boldsymbol{\xi}^{(t)}) = {\boldsymbol{\xi}^{(t)}}^\top\hat{\Sigma}^{-1}\boldsymbol{\xi}^{(t)}$$

となります。これは平均 $\mathbf{0}$ におけるマハラノビス距離です。したがって、ホテリング理論と等価です。

なお、本文で議論したベクトル自己回帰モデルでは定数項を含めないモデルを使いましたが、含めることは可能です。そのような場合、元のベクトル自己回帰モデルは定数 $\boldsymbol{\mu}$ により

$$\boldsymbol{\xi}^{(t)} \approx \boldsymbol{\mu}$$

と近似するモデルになります。より正確に書くと

$$\boldsymbol{\xi}^{(t)} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

ということです。これは 2.4 節のモデルと同等です。

7.0.4

時刻 t と $t+1$ における正則な行列 Q_t と Q_{t+1} が、ベクトル $\boldsymbol{x}$ に対して

$$\mathsf{Q}_{t+1} = (1-\beta)\mathsf{Q}_t + \beta \boldsymbol{x}\boldsymbol{x}^\top$$

という関係を満たすとします（$0 < \beta < 1$ はある定数)。この時、ウッドベリー行列恒等式 (A.22) を使うことで、それぞれの逆行列が

$$\mathsf{Q}_{t+1}^{-1} = \frac{1}{1-\beta}\mathsf{Q}_t^{-1} - \left(\frac{\beta}{1-\beta}\right)\frac{\mathsf{Q}_t^{-1}\boldsymbol{x}\boldsymbol{x}^\top\mathsf{Q}_t^{-1}}{1-\beta+\beta\boldsymbol{x}^\top\mathsf{Q}_t^{-1}\boldsymbol{x}}$$

という関係を満たすことを示して下さい。この式は、前の時刻の逆行列がわかっていれば、逆行列の再計算をすることなしに現在の逆行列を計算できることを意味しています。一般にこのような式を**階数 1 更新式**と呼びます。

[解答] ウッドベリー行列恒等式 (A.22) において

$$\mathsf{A} \leftarrow (1-\beta)\mathsf{Q},\ \mathsf{B} \leftarrow \boldsymbol{x},\ \ \mathsf{C} \leftarrow \boldsymbol{x}^\top,\ \ \mathsf{D} \leftarrow -\frac{1}{\beta}$$

と置くと

$$\begin{aligned}\mathsf{Q}_{t+1}^{-1} &= \frac{1}{1-\beta}\mathsf{Q}_t^{-1} + \frac{1}{1-\beta}\mathsf{Q}_t^{-1}\boldsymbol{x}\left(-\frac{1}{\beta} - \frac{\boldsymbol{x}^\top\mathsf{Q}_t^{-1}\boldsymbol{x}}{1-\beta}\right)^{-1}\boldsymbol{x}^\top\mathsf{Q}_t^{-1}\frac{1}{1-\beta} \\ &= \frac{1}{1-\beta}\mathsf{Q}_t^{-1} - \frac{\beta}{1-\beta}\times\frac{\mathsf{Q}_t^{-1}\boldsymbol{x}\boldsymbol{x}^\top\mathsf{Q}_t^{-1}}{1-\beta+\beta\boldsymbol{x}^\top\mathsf{Q}_t^{-1}\boldsymbol{x}}\end{aligned}$$

7.0.5

カルマンフィルタを使って、道路を走る車を追跡することを 考えます。$\boldsymbol{z}^{(t)}$ を時刻 t での車の真の位置、$\boldsymbol{x}^{(t)}$ をある観測機器が報告した（誤差を含むかもしれない）位置だとします。$\mathsf{C} = \mathsf{I}_M$ かつ $\mathsf{A} = \mathsf{I}_M$ とします。さらに、$\mathsf{R} = \epsilon\mathsf{I}_M$ として、$\epsilon \to 0$ が成り立つとき、どのような更新式が得られるでしょうか。状態空間モデル自体から想像される状況と、カルマンフィルタの結果が直感的に首尾一貫していることを説明してください。

[解答] A と C が単位行列ということは、状態変数を直接観測できる状況を意味しています。状態変数は時々刻々変わりますが、前の時刻での状態変数の周りのランダムな位置に遷移するだけです。モデル (7.27) と (7.28) において

$$\begin{aligned}p(\boldsymbol{x}^{(t)} \mid \boldsymbol{z}^{(t)}) &= \mathcal{N}(\boldsymbol{x}^{(t)} \mid \boldsymbol{z}^{(t)}, \epsilon\mathsf{I}_M) \\ p(\boldsymbol{z}^{(t)} \mid \boldsymbol{z}^{(t-1)}) &= \mathcal{N}(\boldsymbol{z}^{(t)} \mid \boldsymbol{z}^{(t-1)}, \mathsf{Q})\end{aligned}$$

となりますが、第 1 式から分かるとおり、潜在変数は無限の精度で測定できます。

カルマンフィルタの式では、$m = M$ かつ $\mathsf{K}_t = \mathsf{Q}_{t-1}(\epsilon\mathsf{I}_M + \mathsf{Q}_{t-1})^{-1} \approx \mathsf{I}_M$ となるので

$$\boldsymbol{\mu}_t = \boldsymbol{x}^{(t)}, \quad \mathsf{Q}_t = \mathsf{Q}$$

が更新式となり

$$p(\boldsymbol{z}^{(t)} \mid \mathsf{X}_t) = \mathcal{N}(\boldsymbol{z}^{(t)} \mid \boldsymbol{x}^{(t)}, 0)$$

が成り立ちます。やはり、状態すなわち真の位置 $\boldsymbol{z}^{(t)}$ は常に観測量 $\boldsymbol{x}^{(t)}$ と一致していることがわかります。

メ　　モ

- [ガウス積分 $(a > 0)$]

$$\int_{-\infty}^{+\infty} \mathrm{d}x \ \exp(-ax^2 + bx + c) = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right)$$
$$\int_{-\infty}^{+\infty} \mathrm{d}x \ x^2 \exp(-ax^2) = \frac{1}{2a}\sqrt{\frac{\pi}{a}}$$

- [多次元正規変数の 1 次結合] $\boldsymbol{x}$ と $\boldsymbol{x}'$ が独立に多次元正規分布 $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ に従うとき、行列 A と B により作られる確率変数 $\mathsf{A}\boldsymbol{x} + \mathsf{B}\boldsymbol{x}'$ は、多次元正規分布 $\mathcal{N}\left((\mathsf{A}+\mathsf{B})\boldsymbol{\mu}\ ,\ \mathsf{A}\Sigma\mathsf{A}^\top + \mathsf{B}\Sigma\mathsf{B}^\top\right)$ に従う。
- [多変数正規分布のベイズ公式] $p(\boldsymbol{y} \mid \boldsymbol{x})$ および $p(\boldsymbol{x})$ が

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{y} \mid \mathsf{A}\boldsymbol{x} + \boldsymbol{b},\ \mathsf{D})$$
$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma)$$

で与えられる時、$p(\boldsymbol{x} \mid \boldsymbol{x})$ および $p(\boldsymbol{y})$ は次で与えられる。

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{x} \ \middle|\ \mathsf{M}\left\{\mathsf{A}^\top\mathsf{D}^{-1}(\boldsymbol{y} - \boldsymbol{b}) + \Sigma^{-1}\boldsymbol{\mu}\right\}, \mathsf{M}\right)$$
$$p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y} \mid \mathsf{A}\boldsymbol{\mu} + \boldsymbol{b},\ \mathsf{D} + \mathsf{A}\Sigma\mathsf{A}^\top)$$

ただし、M は次式で定義される。

$$\mathsf{M} \equiv (\mathsf{A}^\top\mathsf{D}^{-1}\mathsf{A} + \Sigma^{-1})^{-1}$$