W8PRML 演習問題 2.17, 2.18, 2.20, 2.21, 2.22

大上 雅史　(@tonets)

2012/9/28

2.17（基本）

(2.43) の多変量ガウス分布を考える．精度行列（逆共分散行列）$\mathbf{\Sigma}^{-1}$ を対称行列と反対称行列（歪対称行列）の和の形で書くと，反対称行列の項がガウス分布の指数分布には現れなくなるため，一般性を失うことなく精度行列は対象であるとしてよいことを示せ．この結果から，対称行列の逆行列も対象 (▷ **演習 2.22**) なので，一般性を失うことなく，共分散行列にも対象なものを選んでよいことになる．

$$\text{(2.43)} — \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\mathbf{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

まず，反対称行列 $\mathbf{T}$ ($\mathbf{T}^\top = -\mathbf{T}$) が任意のベクトル $\mathbf{a} \neq \mathbf{0}$ に対し，$\mathbf{a}^\top\mathbf{T}\mathbf{a} = 0$ となることを示す．
$\mathbf{a}^\top\mathbf{T}\mathbf{a}$ はスカラー値なので，$(\mathbf{a}^\top\mathbf{T}\mathbf{a})^\top = \mathbf{a}^\top\mathbf{T}\mathbf{a}$ である．これより

$$\mathbf{a}^\top\mathbf{T}\mathbf{a} = (\mathbf{a}^\top\mathbf{T}\mathbf{a})^\top = \mathbf{a}^\top\mathbf{T}^\top\mathbf{a} = -\mathbf{a}^\top\mathbf{T}\mathbf{a}$$

よって，$\mathbf{a}^\top\mathbf{T}\mathbf{a} = 0$.
精度行列 $\mathbf{\Sigma}^{-1}$ は以下のように行列 $\mathbf{S}$ と $\mathbf{T}$ に分解すると，$\mathbf{S}$ は対称行列，$\mathbf{T}$ は反対称行列となる．

$$\begin{aligned}\mathbf{\Sigma}^{-1} &= \mathbf{S} + \mathbf{T}\\ \mathbf{S} &= \frac{1}{2}\left(\mathbf{\Sigma}^{-1} + (\mathbf{\Sigma}^{-1})^\top\right)\\ \mathbf{T} &= \frac{1}{2}\left(\mathbf{\Sigma}^{-1} - (\mathbf{\Sigma}^{-1})^\top\right)\end{aligned}$$

これより，(2.43) の指数部を考えると，

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= (\mathbf{x}-\boldsymbol{\mu})^\top(\mathbf{S}+\mathbf{T})(\mathbf{x}-\boldsymbol{\mu})\\ &= (\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{S}(\mathbf{x}-\boldsymbol{\mu}) + (\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{T}(\mathbf{x}-\boldsymbol{\mu})\end{aligned}$$

となるが，先に証明した反対称行列の性質から $(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{T}(\mathbf{x}-\boldsymbol{\mu}) = 0$ となるので

$$(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{S}(\mathbf{x}-\boldsymbol{\mu})$$

となり，対称行列のみ残る．

2.18（難問）

実対称行列 $\mathbf{\Sigma}$ を考える．この行列について (2.45) の固有値の方程式が成立する．この式の複素共役から，もとの式を引いた後，固有ベクトル $\mathbf{u}_i$ との内積をとることで，固有値 λ_i が実数となることを示せ．同様に，$\mathbf{\Sigma}$ の対称性を用いて，2 つの固有ベクトル $\mathbf{u}_i$ と $\mathbf{u}_j$ が $\lambda_i \neq \lambda_j$ であれば，直交することを示せ．最後に，たとえいくつかの固有値が 0 であっても，一般性を失うことなく，(2.46) を満たす，正規直交となるように固有ベクトル集合を選ぶことが可能であることを示せ．

$$(2.45) — \mathbf{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i \qquad\qquad (2.46) — \mathbf{u}_i^\top \mathbf{u}_j = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases}$$

---

(1) 実対称行列 $\mathbf{\Sigma}$ の固有値 λ_i がすべて実数であることを示す．固有値 $\lambda_i = \sigma + j\tau$，固有ベクトル $\mathbf{u}_i = \mathbf{v} + j\mathbf{w}$ とおく．ただし，σ, τ は実数，$\mathbf{v}, \mathbf{w}$ は実ベクトル，j は虚数単位である．$\mathbf{\Sigma}$ が実対称行列なので，$\overline{\mathbf{\Sigma}\mathbf{u}_i} = \mathbf{\Sigma}\overline{\mathbf{u}}_i$ である（上線は複素共役を表す）ことを利用して，固有方程式と固有方程式の複素共役について，和と差を計算する．

$$\begin{aligned}
\mathbf{\Sigma}\mathbf{u}_i + \mathbf{\Sigma}\overline{\mathbf{u}}_i &= \lambda_i \mathbf{u}_i + \overline{\lambda_i \mathbf{u}_i} \\
2\mathbf{\Sigma}\Re[\mathbf{u}_i] &= 2\Re[\lambda_i \mathbf{u}_i] \\
2\mathbf{\Sigma}\mathbf{v} &= 2(\sigma\mathbf{v} - \tau\mathbf{w}) \\
\mathbf{\Sigma}\mathbf{v} &= \sigma\mathbf{v} - \tau\mathbf{w} \text{ — (和)} \\
\mathbf{\Sigma}\mathbf{u}_i - \mathbf{\Sigma}\overline{\mathbf{u}}_i &= \lambda_i \mathbf{u}_i - \overline{\lambda_i}\overline{\mathbf{u}}_i \\
2j\mathbf{\Sigma}\Im[\mathbf{w}] &= 2j\Im[\lambda_i \mathbf{u}_i] \\
2j\mathbf{\Sigma}\mathbf{w} &= 2j(\tau\mathbf{v} + \sigma\mathbf{w}) \\
\mathbf{\Sigma}\mathbf{w} &= \tau\mathbf{v} + \sigma\mathbf{w} \text{ — (差)}
\end{aligned}$$

ここで，$\mathbf{w}^\top\mathbf{\Sigma}\mathbf{v} = \mathbf{w}^\top\mathbf{\Sigma}^\top\mathbf{v} = \mathbf{w}^\top(\mathbf{v}^\top\mathbf{\Sigma})^\top = ((\mathbf{v}^\top\mathbf{\Sigma})\mathbf{w})^\top = \mathbf{v}^\top\mathbf{\Sigma}\mathbf{w}$ が成り立つので，(和) に左から $\mathbf{w}^\top$ をかけたものと，(差) に左から $\mathbf{v}^\top$ をかけたものは等しい．よって，

$$\begin{aligned}
&\sigma\mathbf{w}^\top\mathbf{v} - \tau\mathbf{w}^\top\mathbf{w} = \tau\mathbf{v}^\top\mathbf{v} + \sigma\mathbf{v}^\top\mathbf{w} \\
&\tau(\mathbf{v}^\top\mathbf{v} + \mathbf{w}^\top\mathbf{w}) = 0 \\
&\tau(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2) = 0 \\
&\qquad\qquad \therefore \tau = 0
\end{aligned}$$

よって，$\lambda_i = \sigma$ であり，固有値は実数である．

(2) 2 つの固有ベクトル $\mathbf{u}_i$ と $\mathbf{u}_j$ が $\lambda_i \neq \lambda_j$ であれば直交することを示す．2 つの固有方程式 $\mathbf{\Sigma}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ (a) と $\mathbf{\Sigma}\mathbf{u}_j = \lambda_j\mathbf{u}_j$ (a) を考える．(1) の証明で行ったのと同様にして，$\mathbf{u}_j^\top\mathbf{u}_i = \mathbf{u}_i^\top\mathbf{\Sigma}\mathbf{u}_j$ となるので，(a) に $\mathbf{u}_i^\top$，(b) に $\mathbf{u}_i^\top$ をかけて，

$$\begin{aligned}
\lambda_i\mathbf{u}_j^\top\mathbf{u}_i &= \lambda_j\mathbf{u}_i^\top\mathbf{u}_j \\
\lambda_i\mathbf{u}_i^\top\mathbf{u}_j &= \lambda_j\mathbf{u}_i^\top\mathbf{u}_j \quad (\mathbf{a}^\top\mathbf{b} = \mathbf{b}^\top\mathbf{a} \text{ を使った}) \\
(\lambda_i - \lambda_j)\mathbf{u}_i^\top\mathbf{u}_j &= 0
\end{aligned}$$

$\lambda_i \neq \lambda_j$ ならば，$\mathbf{u}_i^\top \mathbf{u}_j = 0$ なので，$\mathbf{u}_i \top \mathbf{u}_j$.
（問題文の「この式の複素共役から，もとの式を引いた後，固有ベクトル $\mathbf{u}_i$ との内積をとることで」で計算できなかったので無視しました.）

(3) たとえいくつかの固有値が 0 であっても，一般性を失うことなく，(2.46) を満たす，正規直交となるように固有ベクトル集合を選ぶことが可能であることを示す．まず，球面 $\|\mathbf{x}\| = 1$ を考え，この球面上で $|\mathbf{x}^\top \mathbf{\Sigma} \mathbf{x}|$ を最大にする単位ベクトル $\mathbf{u}_1$ をとる．また，その最大値を λ とおく．このとき，以下の証明により，$\mathbf{\Sigma}\mathbf{u}_1 = \lambda \mathbf{u}_1$ である．
(証明)　$\mathbf{u}_1^\top \mathbf{x} = 0, \|\mathbf{x}\| = 1$ を満たす任意の $\mathbf{x}$ と，$\cos\theta \sin\theta \neq 0$ となる任意の θ に対して，$\mathbf{u}_1^\top \mathbf{\Sigma} \mathbf{u}_1 = \lambda$ より，

$$\begin{aligned}
\lambda &\geq (\cos\theta \cdot \mathbf{u}_1 \pm \sin\theta \cdot \mathbf{x})^\top \mathbf{\Sigma} (\cos\theta \cdot \mathbf{u}_1 \pm \sin\theta \cdot \mathbf{x}) \\
&= \cos^2\theta \cdot \lambda \pm 2\cos\theta \sin\theta \cdot \mathbf{u}_1^\top \mathbf{\Sigma}\mathbf{x} + \sin^2\theta \cdot \mathbf{x}^\top \mathbf{\Sigma}\mathbf{x} \\
&= \pm 2\cos\theta \sin\theta \cdot \mathbf{u}_1^\top \mathbf{\Sigma}\mathbf{x} + (1 - \sin^2\theta)\lambda + \sin^2\theta \cdot \mathbf{x}^\top \mathbf{\Sigma}\mathbf{x} \\
\pm 2\cos\theta \sin\theta \cdot \mathbf{u}_1^\top \mathbf{\Sigma}\mathbf{x} &\leq 2\sin^2\theta(\lambda - \mathbf{x}^\top \mathbf{\Sigma}\mathbf{x}) \\
|\mathbf{u}_1^\top \mathbf{\Sigma}\mathbf{x}| &\leq |\tan\theta|(\lambda - \mathbf{x}^\top \mathbf{\Sigma}\mathbf{x})
\end{aligned}$$

$\theta \to 0$ の極限をとると，$\mathbf{u}_1^\top \mathbf{\Sigma}\mathbf{x} = \mathbf{x}^\top \mathbf{\Sigma}\mathbf{u}_1 = 0$．よって $\mathbf{u}_1$ と直交する任意の $\mathbf{x}$ は，$\mathbf{\Sigma}\mathbf{u}_1$ とも直交する．このことは $\mathbf{\Sigma}\mathbf{u}_1$ と $\mathbf{u}_1$ の方向が同じ，即ち $\mathbf{\Sigma}\mathbf{u}_1 = \lambda\mathbf{u}_1$ を意味する．(証明終)

n を $\mathbf{\Sigma}$ の次数（$\mathbf{\Sigma}$ は $n \times n$ 行列）とおき，適当に $\mathbf{v}_2, ..., \mathbf{v}_n$ をとって正規直交基底 $\{\mathbf{u}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ を作る．これを横に並べた行列 $\mathbf{U} = (\mathbf{u}_1 \mathbf{v}_2 ... \mathbf{v}_n)$ は直交行列となる．$\mathbf{u}_1$ については固有ベクトルであることが分かっているので，

$$\mathbf{U}^\top \mathbf{\Sigma} \mathbf{U} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{pmatrix}$$

となる．即ち，これにより第 1 行第 1 列の掃き出しが可能であり，かつ残る $(n-1) \times (n-1)$ 行列は実対称行列であることが分かる．今，$(n-1)$ 次実対称行列の固有ベクトルが正規直交となるように $n-1$ 個取れると仮定すると，その $n-1$ 個の固有ベクトルを並べた行列（直交行列）によって先ほどの残った $(n-1) \times (n-1)$ 行列の対角化ができ，$\mathbf{\Sigma}$ 全体を対角化することができる．したがって，数学的帰納法により n 次実対称行列は直交行列を用いて対角化できることが証明され，これは固有値の値に関わらず固有ベクトルが n 個取れることを意味する．

2.20（標準）

正定値行列 $\boldsymbol{\Sigma}$ は，次の二次形式が，任意の実ベクトル $\mathbf{a}$ について正になるということで定義できる．

$$\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$$

$\boldsymbol{\Sigma}$ が正定値になる必要十分条件は，(2.45) で定義される $\boldsymbol{\Sigma}$ のすべての固有値 λ_i が正となることであることを示せ．

（$\mathbf{a} \neq \mathbf{0}$ と仮定する.）任意の実ベクトル $\mathbf{a}$ は $\boldsymbol{\Sigma}$ の固有ベクトル $\mathbf{u}_i$ の線形結合

$$\mathbf{a} = \sum_{i=1}^{D} c_i \mathbf{u}_i$$

で表せる．ただし，D は各ベクトルの次元数，$c_i \in \mathbb{R}$ は定数である．$\boldsymbol{\Sigma}$ の固有値を λ_i とすると，固有方程式 $\boldsymbol{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ を用いて，

$$\begin{aligned}
\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} &= \left(\sum_{i=1}^{N} c_i \mathbf{u}_i\right)^\top \boldsymbol{\Sigma} \left(\sum_{j=1}^{N} c_j \mathbf{u}_j\right) \\
&= \left(\sum_{i=1}^{N} c_i \mathbf{u}_i\right)^\top \left(\sum_{j=1}^{N} c_j \boldsymbol{\Sigma} \mathbf{u}_j\right) \\
&= \left(\sum_{i=1}^{N} c_i \mathbf{u}_i\right)^\top \left(\sum_{j=1}^{N} c_j \lambda_j \mathbf{u}_j\right) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \lambda_j \mathbf{u}_i^\top \mathbf{u}_j
\end{aligned}$$

となるが，

$$\mathbf{u}_i^\top \mathbf{u}_j = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

なので，$\sum_j$ については $j = i$ の部分のみ残る．よって，

$$\begin{aligned}
\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} &= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \lambda_j \mathbf{u}_i^\top \mathbf{u}_j \\
&= \sum_{i=1}^{N} c_i^2 \lambda_i
\end{aligned}$$

となる．これより，$\boldsymbol{\Sigma}$ が正定値であれば，

$$\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} = \sum_{i=1}^{N} c_i^2 \lambda_i > 0$$

なので，$\lambda_i \geq 0$．ここである固有値 λ_k が 0 であると仮定すると，固有方程式より $\boldsymbol{\Sigma}\mathbf{u}_k = \lambda_k \mathbf{u}_k = \mathbf{0}$ となり，$\boldsymbol{\Sigma}^{-1}$ を左からかけて $\mathbf{u}_k = \mathbf{0}$．しかし，演習 2.18 より，$\mathbf{u}_k \neq \mathbf{0}$ なので，背理法より，$\lambda_k \neq 0$．したがって，全ての固有値に対し，

$$\lambda_i > 0$$

が成り立つ．逆に $\lambda_i > 0$ であれば，$c_i^2 \geq 0$（ただし全ての c_i が同時に 0 となることはない）．よって，

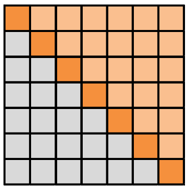
$$\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} = \sum_{i=1}^{N} c_i^2 \lambda_i > 0$$

となる．

2.21（基本）

大きさが $D \times D$ の実対称行列の独立なパラメータは $D(D+1)/2$ 個であることを示せ.

対称行列の独立なパラメータは，以下の図に示すように非対角領域の片方と対角成分である.



$$7+6+5+4+3+2+1=\sum_{k=1}^{7} k$$

よって，

$$\sum_{k=1}^{D} k = \frac{D(D+1)}{2}$$

2.22（基本）

対称行列の逆行列も対称であることを示せ.

$\mathbf{A}$ を対称行列，$\mathbf{I}$ を単位行列とする．$\mathbf{A}$ に逆行列が存在するならば，

$$\begin{aligned}
&\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{I}^{\top} = (\mathbf{A}^{-1}\mathbf{A})^{\top} = \mathbf{A}^{\top}(\mathbf{A}^{-1})^{\top} = \mathbf{A}(\mathbf{A}^{-1})^{\top} \\
&\mathbf{A}(\mathbf{A}^{-1})^{\top} = \mathbf{I} \\
&\mathbf{A}^{-1}\mathbf{A}(\mathbf{A}^{-1})^{\top} = \mathbf{A}^{-1}\mathbf{I} \\
&\therefore (\mathbf{A}^{-1})^{\top} = \mathbf{A}^{-1}
\end{aligned}$$