12.25

($p(\boldsymbol{y}|\boldsymbol{z})$ を求める)

$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ とすると

$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ なので $\boldsymbol{y}$ が値 $\boldsymbol{y}$ を取る確率 $p(\boldsymbol{y})d\boldsymbol{y}$ は
$\boldsymbol{x}$ が $\boldsymbol{x}$ の値を取る確率 $p(\boldsymbol{x})d\boldsymbol{x}$ に等しいので

$$p(\boldsymbol{y}|\boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z}) \left| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right|$$

である。ここで $\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{y}$ なので

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \frac{\partial \mathbf{A}^{-1}\boldsymbol{y}}{\partial \boldsymbol{y}} = (\mathbf{A}^{-1})^T$$

denominator layout
注 (C.18) は numerator layout である。
ここでは denominator layout に揃える

$$\left(\frac{\partial \mathbf{A}\boldsymbol{x}}{\partial \boldsymbol{x}}\right)_{ij} = \frac{\partial (\mathbf{A}\boldsymbol{x})_j}{\partial x_i}$$

$\mathbf{A}$ の j 行

$$= \frac{\partial \mathbf{A}_j \boldsymbol{x}}{\partial x_i}$$

$$= \mathbf{A}_{ji}$$

$$\therefore \frac{\partial \mathbf{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \mathbf{A}^T \text{ (denominator layout)}$$

よって

$$\left| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right| = |(\mathbf{A}^{-1})^T| = |\mathbf{A}^{-1}|$$

となる。また問題文より

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\mathbf{W}\boldsymbol{z}+\boldsymbol{\mu}, \boldsymbol{\Phi})$$

なので

$$p(\boldsymbol{y}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\mathbf{W}\boldsymbol{z}+\boldsymbol{\mu}, \boldsymbol{\Phi})\,|\mathbf{A}^{-1}|$$

$$= \mathcal{N}(\mathbf{A}^{-1}\boldsymbol{y}|\mathbf{W}\boldsymbol{z}+\boldsymbol{\mu}, \boldsymbol{\Phi})\,|\mathbf{A}^{-1}|$$

である。これより

$$|\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^T|^{\frac{1}{2}} = |\mathbf{A}|^{\frac{1}{2}}|\boldsymbol{\Phi}|^{\frac{1}{2}}|\mathbf{A}^T|^{\frac{1}{2}} = |\boldsymbol{\Phi}|^{\frac{1}{2}}|\mathbf{A}|$$

$$p(\boldsymbol{y}|\boldsymbol{z}) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Phi}|^{\frac{1}{2}}} \frac{1}{|\mathbf{A}|} \exp\left\{-\frac{1}{2}(\mathbf{A}^{-1}\boldsymbol{y}-\mathbf{W}\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Phi}^{-1} (\mathbf{A}^{-1}\boldsymbol{y}-\mathbf{W}\boldsymbol{z}-\boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^T|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{A}^{-1}(\boldsymbol{y}-\mathbf{A}\mathbf{W}\boldsymbol{z}-\mathbf{A}\boldsymbol{\mu}))^T \boldsymbol{\Phi}^{-1} (\mathbf{A}^{-1}(\boldsymbol{y}-\mathbf{A}\mathbf{W}\boldsymbol{z}-\mathbf{A}\boldsymbol{\mu}))\right\}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^T|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\mathbf{A}\mathbf{W}\boldsymbol{z}-\mathbf{A}\boldsymbol{\mu})^T (\mathbf{A}^T)^{-1} \boldsymbol{\Phi}^{-1} \mathbf{A}^{-1} (\boldsymbol{y}-\mathbf{A}\mathbf{W}\boldsymbol{z}-\mathbf{A}\boldsymbol{\mu})\right\}$$

$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$

$$= \mathcal{N}(\boldsymbol{y}|\mathbf{A}\mathbf{W}\boldsymbol{z}+\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^T)$$

を得る。

ここで $\mathbf{M} = \mathbf{A}\mathbf{W}$, $\boldsymbol{\nu} = \mathbf{A}\boldsymbol{\mu}$, $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^T$ とおくと

$$p(\boldsymbol{y}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{y}|\mathbf{M}\boldsymbol{z}+\boldsymbol{\nu}, \boldsymbol{\Psi})$$

となる。

($p(\mathbf{y})$ を求める)

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$$

$$p(\mathbf{y} \mid \mathbf{z}) = \mathcal{N}(\mathbf{y} \mid \mathbf{M}\mathbf{z} + \boldsymbol{\nu}, \boldsymbol{\Psi})$$

なので、(2.115) より

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \boldsymbol{\nu}, \boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T)$$

を得る。

($\mathbf{Y} = \{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$ の対数尤度を求める)

$\leftarrow$ $\mathbf{y}_n$は独立なので

$$\ln p(\mathbf{Y}) = \ln \prod_{n=1}^{N} p(\mathbf{y}_n) = \sum_{n=1}^{N} \ln \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y}_n - \boldsymbol{\nu})^T (\boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T)^{-1} (\mathbf{y}_n - \boldsymbol{\nu})\right\}$$

$$= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{y}_n - \boldsymbol{\nu})^T(\boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T)^{-1}(\mathbf{y}_n - \boldsymbol{\nu})$$

($p(\mathbf{x})$ を求める)

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x} \mid \mathbf{z}) = \mathcal{N}(\mathbf{x} \mid \mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \boldsymbol{\Phi})$$

なので、(2.115) より

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T)$$

($\mathbf{X} = \{\mathbf{x}_1 \cdots \mathbf{x}_N\}$ の対数尤度)

$$\ln p(\mathbf{X}) = \ln \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}, \boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T) = \sum_{n=1}^{N} \ln \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu})^T(\boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})\right\}$$

$$= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^T(\boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

（$\ln p(\mathbf{X})$ の最尤解 $\boldsymbol{\mu}_{ML}$ と $\ln p(\mathbf{Y})$ の最尤解 $\boldsymbol{\nu}_{ML}$ をくらべる）

$\boldsymbol{\mu}_{ML}$ を求める

$$0 = \frac{\partial}{\partial \boldsymbol{\mu}_{ML}} \ln p(\mathbf{X}) = -\frac{1}{2}\sum_{n=1}^{N}(-1)\,2\,(\boldsymbol{\Phi} + \mathbf{W}\mathbf{W}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

$$\therefore\ \boldsymbol{\mu}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

- ベクトル微分の chain rule
  $g = g(\mathbf{u}),\ \mathbf{u} = \mathbf{u}(\mathbf{x})$ のとき
  $\frac{\partial g}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial g}{\partial \mathbf{u}}$ (denominator layoutのとき)
- $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A}\mathbf{x} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{x}} \mathbf{A}\mathbf{x} + \mathbf{x}^T \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}}$
  $= \mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}^T \leftarrow \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^T$
  $= (\mathbf{A} + \mathbf{A}^T)\mathbf{x} \leftarrow$ denominator layout
  $= 2\mathbf{A}\mathbf{x}$
  $\uparrow$ $\mathbf{A}$ 対称のとき

同様に $\boldsymbol{\nu}_{ML}$ を求める。

$$0 = \frac{\partial}{\partial \boldsymbol{\nu}_{ML}} \ln p(\mathbf{Y}) = -\frac{1}{2}\sum_{n=1}^{N}(-1)\,2\,(\boldsymbol{\Psi} + \mathbf{M}\mathbf{M}^T)^{-1}(\mathbf{y}_n - \boldsymbol{\nu})$$

$$\therefore\ \boldsymbol{\nu}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n$$

ここで $\mathbf{y}_n = \mathbf{A}\mathbf{x}_n$ なので

$$\boldsymbol{\nu}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{A}\mathbf{x}_n = \mathbf{A}\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n = \mathbf{A}\boldsymbol{\mu}_{ML}$$

となっていることが確認できる。

($\ln p(X)$ の最尤解 W_{ML} と $\ln p(Y)$ の最尤解 M_{ML} を比べる)

W_{ML} は

$$0 = \frac{\partial}{\partial W} \ln p(X) = \frac{\partial}{\partial W}\left\{ -\frac{N}{2} \ln|\Phi + WW^T| - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^T(\Phi + WW^T)^{-1}(x_n - \mu)\right\}$$

$$\therefore \quad 0 = \frac{\partial}{\partial W}\left\{ N \ln|\Phi + WW^T| + \sum_{n=1}^{N}(x_n - \mu)^T(\Phi + WW^T)^{-1}(x_n - \mu)\right\} \cdots ①$$

の解で与えられる。

M_{ML} は、

$$0 = \frac{\partial}{\partial M} \ln p(Y) = \frac{\partial W}{\partial M}\frac{\partial}{\partial W} \ln p(Y)$$

$$\therefore \quad 0 = \frac{\partial}{\partial W} \ln p(Y) \quad \cdots ②$$

② の解 W を使って $M_{ML} = AW$ で与えられる。

ここで $y_n = Ax_n$, $\Psi = A\Phi A^T$, $M = AW$ を使いまた ν, μ として最尤解 $\nu_{ML} = A\mu_{ML}$ を使うと ② は

$$0 = \frac{\partial}{\partial W} \ln p(Y)$$

$$= \frac{\partial}{\partial W}\left\{ -\frac{N}{2} \ln|\Psi + MM^T| - \frac{1}{2}\sum_{n=1}^{N}(y_n - \nu)^T(\Psi + MM^T)^{-1}(y_n - \nu)\right\}$$

$$= -\frac{1}{2}\frac{\partial}{\partial W}\left\{ N \ln|A\Phi A^T + AW(AW)^T| + \sum_{n=1}^{N}(Ax_n - A\mu_{ML})^T(A\Phi A^T + AW(AW)^T)^{-1}(Ax_n - A\mu_{ML})\right\}$$

$$= -\frac{1}{2}\frac{\partial}{\partial W}\left\{ N \ln|A| + \ln|\Phi + WW^T| + \ln|A^T| + \sum_{n=1}^{N}(x_n - \mu_{ML})A^T(A^T)^{-1}(\Phi + WW^T)^{-1}A^{-1}A(x_n - \mu_{ML})\right\}$$

$$\therefore \quad 0 = \frac{\partial}{\partial W}\left\{ N \ln|\Phi + WW^T| + \sum_{n=1}^{N}(x_n - \mu_{ML})^T(\Phi + WW^T)^{-1}(x_n - \mu_{ML})\right\}$$

となり、② は ① と同じ条件式になっている。

なので ② の解 W は ① の解 W_{ML} と同じになる。

よって $\ln p(Y)$ の最尤解 M_{ML} は $\ln p(X)$ の最尤解 W_{ML} を使って

$M_{ML} = A W_{ML}$ で与えられる。

($\ln p(\mathbb{X})$ の最尤解 Ψ_{ML} と $\ln p(\mathbb{Y})$ の最尤解 Φ_{ML} をくらべる)

Φ_{ML} は

$$0 = \frac{\partial}{\partial \Phi} \ln p(\mathbb{X}) = \frac{\partial}{\partial \Phi}\left\{ -\frac{N}{2} \ln |\Phi + WW^T| - \frac{1}{2}\sum_{n=1}^{N} (x_n - \mu)^T (\Phi + WW^T)^{-1} (x_n - \mu) \right\}$$

$$\therefore \quad 0 = \frac{\partial}{\partial \Phi}\left\{ N \ln |\Phi + WW^T| + \sum_{n=1}^{N} (x_n - \mu)^T (\Phi + WW^T)^{-1} (x_n - \mu) \right\} \cdots ③$$

の解で与えられる。

Ψ_{ML} は

$$0 = \frac{\partial}{\partial \Psi} \ln p(\mathbb{Y}) = \frac{\partial \Phi}{\partial \Psi} \frac{\partial}{\partial \Phi} \ln p(\mathbb{Y})$$

$$\therefore \quad 0 = \frac{\partial}{\partial \Phi} \ln p(\mathbb{Y}) \quad \cdots ④$$

の解 Φ より $\Psi_{ML} = A \Phi A^T$ で与えられる。

ここで $y_n = A x_n$, $\Psi = A \Phi A^T$, $M = AW$ を使いまた ν, μ として最尤解 $\nu_{ML} = A \mu_{ML}$ を使うと ④ は

$$0 = \frac{\partial}{\partial \Phi} \ln p(\mathbb{Y})$$

$$= \frac{\partial}{\partial \Phi}\left\{ -\frac{N}{2} \ln |\Psi + MM^T| - \frac{1}{2}\sum_{n=1}^{N} (y_n - \nu)^T (\Psi + MM^T)^{-1} (y_n - \nu) \right\}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \Phi}\left\{ N \ln |A\Phi A^T + AW(AW)^T| + \sum_{n=1}^{N} (Ax_n - A\mu_{ML})^T (A\Phi A^T + AW(AW)^T)^{-1} (Ax_n - A\mu_{ML}) \right\}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \Phi}\left\{ N \ln |A| + \ln |\Phi + WW^T| + \ln |A^T| + \sum_{n=1}^{N} (x_n - \mu_{ML}) A^T (A^T)^{-1} (\Phi + WW^T)^{-1} A^{-1} A (x_n - \mu_{ML}) \right\}$$

$$\therefore \quad 0 = \frac{\partial}{\partial \Phi}\left\{ N \ln |\Phi + WW^T| + \sum_{n=1}^{N} (x_n - \mu_{ML})^T (\Phi + WW^T)^{-1} (x_n - \mu_{ML}) \right\}$$

となり、④ は ③ と同じ条件式になっている。

なので ④ の解 Φ は ③ の解 Φ_{ML} と同じになる。

よって $\ln p(\mathbb{Y})$ の最尤解 Ψ_{ML} は $\ln p(\mathbb{X})$ の最尤解 Φ_{ML} を使って

$\Psi_{ML} = A \Phi_{ML} A^T$ で与えられる。

（(i)(ii)の問題文の「…モデルが共変的…」のイミについて）

良くわからないが、Web解答を見ると多分次のような仕と思われる。

$$p(\mathbf{x}) = N(\mathbf{x} \mid \boldsymbol{\mu}, \Phi + WW^T)$$

に対してデータを $\mathbf{y} = A\mathbf{x}$ で座標変換して、

$$p(\mathbf{y}) = N(\mathbf{y} \mid \boldsymbol{\nu}, \Psi + MM^T)$$

$p(\mathbf{x}) \to p(\mathbf{y})$ の変換において $p(\mathbf{x})$ と $p(\mathbf{y})$ のガウス分布の構造が同じことを共変と言っている

を得たとき、Φ の構造と Ψ の構造が同じになることを"共変的"と言っているみたい。

- Φ が対角で Ψ も対角とか
- $\Phi = \sigma^2 I$ で $\Psi = \sigma^2 I$ とか

(i)

Φ が対角で A も対角のとき

$$\Psi_{ML} = A \Phi_{ML} A^T = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_D \end{pmatrix} \begin{pmatrix} \phi_1 & & 0 \\ & \ddots & \\ 0 & & \phi_D \end{pmatrix}_{ML} \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_D \end{pmatrix} = \begin{pmatrix} a_1^2\phi_1 & & 0 \\ & \ddots & \\ 0 & & a_D^2\phi_D \end{pmatrix}$$

となり Ψ_{ML} も対角である。なので $p(\mathbf{x}) \to p(\mathbf{y})$ の変換は共変的である。

(ii)

$\Phi = \sigma^2 I$ で、A が直交行列のとき

A は直交なので $AA^T = I$

$$\Psi_{ML} = A \Phi_{ML} A^T = A \sigma^2 I A^T = \sigma^2 AA^T = \sigma^2 I$$

となり Ψ_{ML} についても $\Psi_{ML} = \sigma^2 I$ である。なので $p(\mathbf{x}) \to p(\mathbf{y})$ の変換は共変的である。