Parcial 1 curso dirigido Posgrado

2.1. Sea la PDp conjunta $p(z) = p([x,y]) = N(z|\mu_z, \Sigma_z)$. Determine $p(y|x)$

$p(y|x) = N(y|\Lambda x + b, L^{-1})$

Supongamos que $z = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$, $\Sigma_z = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$

Si se asume que $\Delta = \Sigma^{-1}$

$\Rightarrow \Delta = \begin{bmatrix} \Delta_{aa} & \Delta_{ab} \\ \Delta_{ba} & \Delta_{bb} \end{bmatrix}$, teniendo en cuenta que z es una gaussiana multivariada

$$-\frac{1}{2}(z-\mu_z)^T \Sigma^{-1}(z-\mu_z) = -\frac{1}{2}\left([x,y]-[\mu_x,\mu_y]\right)^T$$

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}^{-1} ([x,y]-[\mu_x,\mu_y])$$

$$\Rightarrow -\frac{1}{2}(z^T - \mu_z^T)\Sigma_z^{-1}(z-\mu_z) = -\frac{1}{2}[z^T\Sigma_z^{-1}z - z^T\Sigma_z^{-1}\mu_z$$

$$-\mu_z^T\Sigma_z^{-1}z + \mu_z^T\Sigma_z^{-1}\mu_z] = -\frac{1}{2}z^T\Sigma_z^{-1}\mu_z + \frac{1}{2}z^T\Sigma_z^{-1}\mu_z$$

$$-\frac{1}{2}\mu_z^T\Sigma_z^{-1}\mu_z = -\frac{1}{2}z^T\Sigma_z^{-1}z + z^T\Sigma_z^{-1}\mu_z - \frac{1}{2}\mu_z^T\Sigma_z^{-1}\mu_z$$

$$\Rightarrow -\frac{1}{2}z^T\Sigma_z^{-1}z + z^T\Sigma_z^{-1}\mu_z - \frac{1}{2}\mu_z^T\Sigma_z^{-1}\mu_z$$

$$= -\frac{1}{2}\left[[x-\mu_x, y-\mu_y]^T \begin{bmatrix} \Delta_{xx} & \Delta_{xy} \\ \Delta_{yx} & \Delta_{yy} \end{bmatrix} [x-\mu_x, y-\mu_y]\right]$$

$$= -\frac{1}{2}\left[(x-\mu_x)^T\Delta_{xx}(x-\mu_x) + (y-\mu_y)^T\Delta_{yx}(x-\mu_x) + \right.$$

$$\left.(x-\mu_x)^T\Delta_{xy}(y-\mu_y) + (y-\mu_y)^T\Delta_{yy}(y-\mu_y)\right]$$

$= -\frac{1}{2} x^T \Delta_{xx} x + x^T \Delta_{xx} \mu_x - \frac{1}{2} x^T \Delta_{xx} \mu_x - \frac{1}{2} y^T \Delta_{yx} x$

$+ \frac{1}{2} y^T \Delta_{yx} \mu_x + \frac{1}{2} \mu_y^T \Delta_{yx} x - \frac{1}{2} \mu_y^T \Delta_{yx} \mu_x - \frac{1}{2} x^T \Delta_{xy} y$

$+ \frac{1}{2} x^T \Delta_{xy} \mu_y + \frac{1}{2} \mu_x^T \Delta_{xy} y - \frac{1}{2} \mu_x^T \Delta_{xy} \mu_y - \frac{1}{2} y^T \Delta_{yy} y$

$+ y^T \Delta_{yy} \mu_y - \frac{1}{2} \mu_y^T \Delta_{yy} \mu_y$

En este caso para poder determinar $p(y|x)$ debemos hallar la dependencia de y con x dado se asume x como constante.

Términos lineales en y: $-\frac{1}{2} y^T \Delta_{yx} x + \frac{1}{2} y^T \Delta_{yx} \mu_x - \frac{1}{2} x^T \Delta_{xy} y$

$+ \frac{1}{2} \mu_x^T \Delta_{xy} y + y^T \Delta_{yy} \mu_y$

Por simetría tendremos: $y^T \Delta_{yy} \mu_y - y^T \Delta_{yx} x + y^T \Delta_{yx} \mu_x$

$= y^T (\Delta_{yy} \mu_y - \Delta_{yx} x + \Delta_{yx} \mu_x) = y^T (\Delta_{yy} \mu_y + \Delta_{yx} (\mu_x - x))$

Despejando el término lineal en z de $z^T \Sigma_z^{-1} \mu_z$

$y^T \Sigma_{y|x}^{-1} \mu_{y|x} = y^T (\Delta_{yy} \mu_y + \Delta_{yx} (\mu_x - x))$

$\to \Sigma_{y|x}^{-1} \mu_{y|x} = \Delta_{yy} \mu_y + \Delta_{yx} (\mu_x - x)$. Aquí asumimos que

$\Sigma_{y|x}^{-1} = \Delta_{yy}$ y aplicamos $\Sigma_{y|x}$ en cada lado

$\Rightarrow \Sigma_{y|x} \Sigma_{y|x}^{-1} \mu_{y|x} = \Sigma_{y|x} (\Delta_{yy} \mu_y + \Delta_{yx} (\mu_x - x)$

dado que $MM^{-1} = I$

$\to \mu_{y|x} = \Sigma_{y|x} \Sigma_{y|x}^{-1} \mu_y - \Sigma_{y|x} \Sigma_{y|x}^{-1} (\mu_x - x)$

$\to \mu_{y|x} = \mu_y - \Delta_{yy}^{-1} \Delta_{yx} (\mu_x - x)$

Teniendo que $\Sigma^{-1} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} = \Delta \begin{bmatrix} \Delta_{xx} & \Delta_{xy} \\ \Delta_{yx} & \Delta_{yy} \end{bmatrix}$ (Aplicando la identidad de matriz inversa por partes)

$\Sigma_{y|x} = \Delta_{yy}^{-1} \to \Sigma_{y|x}^{-1} = \Delta_{yy} = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1} \Sigma_{yx} (\Sigma_{xx} - \ldots$

$\ldots \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx})^{-1} \Sigma_{xy} \Sigma_{yy}^{-1}$

Así $p(y|x) = N(y | \mu_{y|x}, \Sigma_{y|x})$

2.2 Problema de optimización

$t_n = \phi(x_n) w^T + R_n$ Mínimos Cuadrados

$R_n = t_n - \phi(x_n) w^T$

$$w^* = \underset{w}{\arg\min} \ \frac{1}{N} \sum_{R_n}^{n=1} \| R_n \|_2^2$$

$$\Rightarrow w^* = \underset{w}{\arg\min} \ \frac{1}{N} \sum_{R_n}^{n=1} \| t_n - \phi(x_n) w^T \|_2^2$$

$$w^* = \underset{w}{\arg\min} \ \| t_n - \phi(x_n) w^T \|_2^2$$

$$w^* = \underset{w}{\arg\min} \ \langle t_n - \phi(x_n) w^T, t_n - \phi(x_n) w^T \rangle$$

$$w^* = \underset{w}{\arg\min} \ (t_n^T t_n - t_n^T \phi(x_n) w^T - (\phi(x_n) w^T)^T$$

$$t_n (\phi(x_n) w^T)^T (\phi(x_n) w^T) \phi$$

$$w^* = \underset{w}{\arg\min} \ (t_n^2 - t_n^T \phi w^T - \phi^T w t_n + \phi^T w \phi w^T$$

$$\frac{d}{dw} = \phi t_n - \phi^T t_n + 2 \phi^T \phi w = -2 \phi^T t_n + 2 \phi^T \phi w = 0$$

$$0 = -2 \phi^T t_n + 2 \phi^T \phi w \rightarrow \phi^T t_n = \phi^T \phi w$$

$$\Rightarrow w = (\phi^T \phi)^{-1} \phi^T t_n$$

- $t_n = \phi(x_n) w^T + R_n + \lambda$ Mínimos cuadrados regularizados

$R_n = t_n - \phi(x_n) w^T + \lambda, \quad \lambda \in \mathbb{R}^+$

$$w^* = \underset{w}{\arg\min} \ \frac{1}{N} \sum_{n}^{n=1} \| t_n - \phi(x_n) w^T \|_2^2 + \lambda \| w \|_2^2$$

Asumiendo

$\|t_n - \phi w^T\|_2^2 = \langle t_n - \phi w^T, t_n - \phi w^T \rangle$

$\Rightarrow t_n^T t_n - t_n^T \phi w^T - (\phi w^T)^T + (t_n + \phi w^T)^T (\phi w^T)$

$\Rightarrow \|t_n - \phi w^T\| = t_n^2 - t_n^T \phi w^T - \phi^T w t_n + \phi^T w \phi w$

$\frac{d}{dw} = -2\phi^T t_n + 2\phi^T \phi w = 0$

$\lambda \|w\|_2^2 = \lambda \langle w, w \rangle = \lambda w^T w$ dado que $w \in \mathbb{R}^q$

$\frac{d}{dw} \Rightarrow -2\lambda I w = 0$

$-2\phi^T t_n + 2\phi^T \phi w - 2\lambda I w = 0 = -\phi^T t_n + \phi^T \phi w - 2\lambda I w$

$-\phi^T t_n + (\phi^T \phi - \lambda I) w = 0$

$\phi^T = t_n = (\phi^T \phi - \lambda I) w$

$w = (\phi^T \phi - \lambda I) \phi^T t_n$

- Maxima verosimilitud

$t_n = \phi(x_n) w^T + \eta_n$

$\eta \sim p(n) = N(\eta_n \mid 0, \sigma^2_n)$

$\eta_n = t_n - \phi(x_n) w^T \sim p(t_n - \phi(x_n) w^T)$

$= P(t_n \mid \phi(x_n) w^T, 0, \sigma^2_n)$

$\Rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|t_n - \phi(x_n) w^T\|_2^2}{2\sigma^2_n}\right)$

Verosimilitud con media cero

$p(t_n \mid \phi(x_n) w^T, \sigma^2_n) = \frac{1}{\sqrt{2\pi\sigma^2_n}} \exp\left(-\frac{\|t_n - \phi(x_n) w^T\|_2^2}{2\sigma^2_n}\right)$

i.i.d

$$p(t_n \mid \phi(x_n)w^T, \sigma^2_n) = p(t_n \mid \phi(x_n)w^T)$$

$$\Rightarrow \prod_{n=1}^{N} p(t_n \mid \phi(x_n)w^T)$$

Ahora aplicando log

$$\log(p(t_n \mid \phi(x_n)w^T)) = \log\left(\prod_{n=1}^{N} p(t_n) \phi(x_n)w^T\right)$$

$$= \log\left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_n}} \exp\left(-\frac{\|t_n - \phi(x_n)w^T\|_2^2}{2\sigma^2_n}\right)\right)$$

$$= \log\left(\frac{1}{(2\pi\sigma^2_n)^{N/2}}\right) + \log\left(\exp\left(-\sum_{n=1}^{N}\frac{\|t_n - (x_n)w^T\|_2^2}{2\sigma^2_2}\right)\right)$$

$$-\frac{N}{2}\log(2\pi\sigma^2_n) - \frac{1}{2\sigma_n^2}\sum_{n=1}^{N}\|t_n - \phi(x_n)w^T\|_2^2$$

$$w^* = \arg\min_w \log(p(t_n \mid \phi(x_n)w^T) = -\frac{N}{2}\log(2\pi\sigma^2_n)$$

$$-\frac{1}{2\sigma^2_n}\sum_{n=1}^{N}\|t_n - \phi(x_n)w^t\|_2^2$$

$$w^* = \arg\min_w \frac{1}{2\sigma^2_n}\|t_n - \phi(x_n)w^T\|_2^2 + \frac{N}{2}\log(2\pi\sigma^2_n) \text{ tct}$$

$$\Rightarrow \eta \sim N(\eta \mid 0, \sigma^2_n) \qquad w^* = (\phi^T\phi)^{-1}\phi^T t_n$$