Regresión Bayesiana - Caso Particular Probabilidad Gaussiana

- En Máxima Verosimilitud:

$$t_n = \phi(x_n)w + \eta; \quad \eta \sim \mathcal{N}(\eta|0,\sigma^2); \quad t_n \in \mathbb{R}^+, \phi(x_n), w \in \mathbb{R}^Q, \; x_n \in \mathbb{R}^P$$

$$\eta = t_n - \phi_n w \sim \mathcal{N}(t_n|\phi(x_n)w, \sigma^2)$$

- Asumiendo datos i.i.d.: $p(\mathbf{t}|\Phi, w, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(t_n|\phi(x_n)w, \sigma^2)$

$$\Phi \in \mathbb{R}^{N\times Q}; \; \mathbf{t} \in \mathbb{R}^N$$

- La solución puntual por máx verosimilitud:

$$w_{ML} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{t} \quad ; \quad \beta_{ML} = \frac{1}{\sigma^2_{ML}}$$

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}(t_n - \phi(x_n)w_{ML})^2$$

- Podemos estimar la distribución predictiva sobre t:

$$p(t|\phi(x), w_{ML}, \sigma^2_{ML}) = \mathcal{N}(t|\phi(x)w_{ML}, \sigma^2_{ML})$$

- Si incluimos un tratamiento "más" Bayesiano, podemos incluir un prior (suposición de incertidumbre sobre los parámetros):

$$p(\mathbf{w}|\sigma_w^2) = \mathcal{N}(\mathbf{w}|0, \sigma_w^2 \mathbf{I})$$

$$p(\mathbf{t}, \Phi, \mathbf{w}, \sigma^2, \sigma_w^2) = p(\mathbf{t}|\Phi\mathbf{w}, \sigma^2)\, p(\mathbf{w}|\sigma_w^2)$$

$$p(\mathbf{w}|\mathbf{t}, \Phi, \sigma^2, \sigma_w^2) = \frac{p(\mathbf{t}|\Phi\mathbf{w}, \sigma^2)\, p(\mathbf{w}|\sigma_w^2)}{p(\mathbf{t})}$$

$$p(\mathbf{w}|\mathbf{t}, \Phi, \sigma^2, \sigma_w^2) \propto p(\mathbf{t}|\Phi\mathbf{w}, \sigma^2)\, p(\mathbf{w}|\sigma_w^2)$$

MÁX A-POSTERIORI (MAP)

EJERCICIO: Demuestre que máx el log-MAP es equivalente a minimizar:

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

Distribución predictiva:

- Dado una entrada nueva $\mathbf{x}$, la incertidumbre sobre la estimación de la predicción t $p(t|\phi(\mathbf{x}), \Phi, w)$, se puede modelar como:

$$p(t|\phi(\mathbf{x}), w) = \int p(t|\phi(\mathbf{x}) w)\, p(w|\mathbf{t}, \Phi w)\, \partial w$$

EJERCICIO: Demuestre que para verosimilitud y prior Gaussianas, la predictiva en un regresor Bayesiano se obtiene como:

$$p(t|\phi(\mathbf{x}) w) = N(t|\mu(\mathbf{x}), s^2(\mathbf{x}))$$

$$\mu(\mathbf{x}) = \frac{1}{\sigma^2}\phi(\mathbf{x})^T S \sum_{n=1}^{N} \phi(\mathbf{x}_n) t_n$$

$$s^2(\mathbf{x}) = \sigma^2 + \phi^T(\mathbf{x}) S \phi(\mathbf{x})$$

$$S = \left[\frac{1}{\sigma_w^2} I + \frac{1}{\sigma^2} \sum_{n=1}^{N} \phi(\mathbf{x}_n)\phi^T(\mathbf{x}_n)\right]^{-1}$$

¿Cuál es la relación entre mín cuadrados regularizados y la predictiva?

Distribución Gaussiana

- Gaussiana univariada: $\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$

  $x, \mu \in \mathbb{R};\ \sigma^2 \in \mathbb{R}^+$

- Gaussiana multivariada:

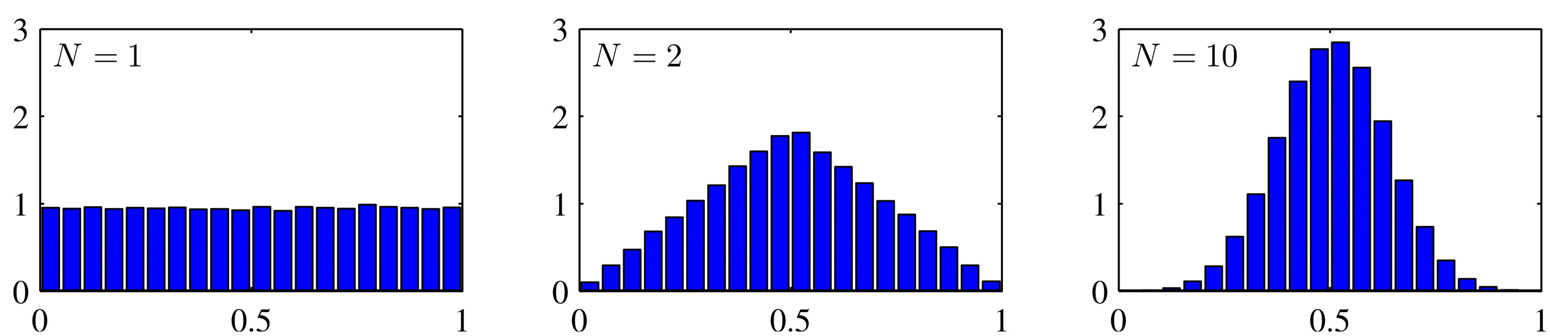
  $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^p;\ \Sigma \in \mathbb{R}^{p\times p}$

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

EJERCICIO: - Demuestre, mediante simulación, que para $x_n \sim p(x_n) = \mathcal{U}(x_n|a,b)$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x_n \sim \mathcal{N}\left(\frac{1}{N}\sum_n x_n \,\Big|\, \mu_x, \sigma_x^2\right).$$ Teorema del límite central.

- Pruebe $N \in \{1, 2, 10\}$



- La forma cuadrática en la Gaussiana Multivariada:

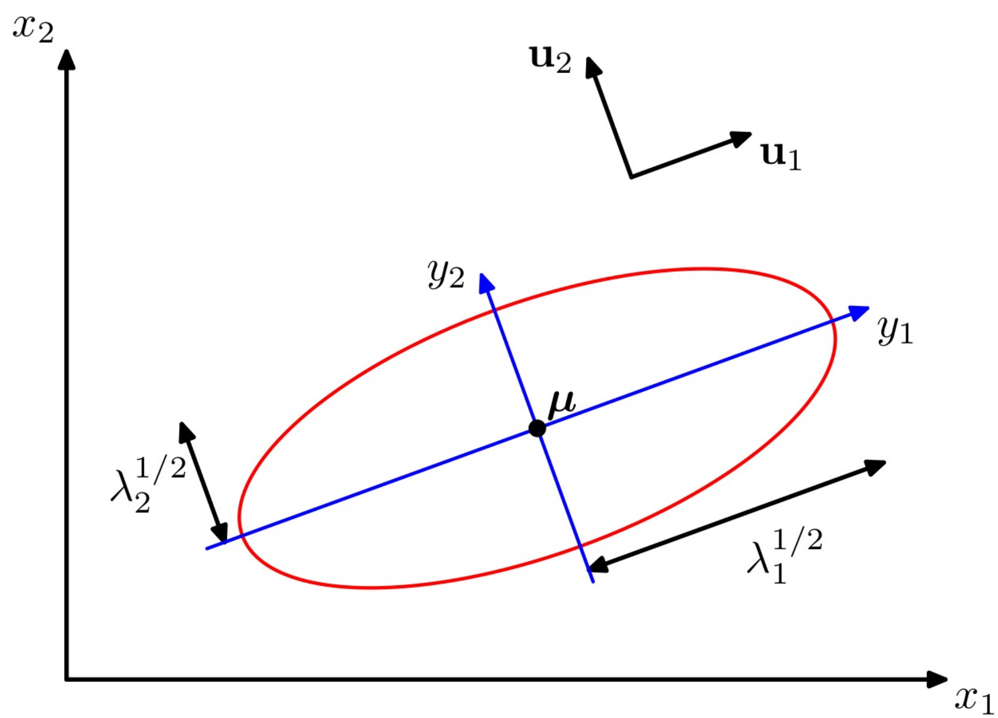
$$\Delta^2 = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$\Sigma = \sum_{j=1}^{p} \lambda_j u_j u_j^T \ , \qquad \Sigma u_j = \lambda_j u_j \rightarrow \text{eigendecomposition.}$$

$$\Sigma^{-1} = (U \Delta U^T)^{-1} = U \Delta^{-1} U^T = \sum_{j=1}^{p} \frac{1}{\lambda_j} u_j u_j^T$$

$$\Delta^2 = (x-\mu)^T \sum_{j=1}^{p} \frac{1}{\lambda_j} u_j u_j^T (x-\mu) = \sum_{j=1}^{p} \frac{1}{\lambda_j} \underbrace{(x-\mu)^T u_j}_{y_j} \underbrace{u_j^T (x-\mu)}_{y_j}$$

$$\Delta^2 = \sum_{j=1}^{p} \frac{y_j^2}{\lambda_j}$$



Distribuciones Gaussianas condicionales

- Supongamos $\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix}$; $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}$; $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}$

$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$

con $\boldsymbol{\Sigma}_{ab} = \boldsymbol{\Sigma}_{ba}^T$.

Ejercicio: Si $\mathbf{x} \in \mathbb{R}^P$, $\mathbf{x}_a \in \mathbb{R}^{\hat{P}}$

$\mathbf{x}_b \in ?$ $\boldsymbol{\mu} \in \mathbb{R}^?$ $\boldsymbol{\mu}_a \in \mathbb{R}^?$

$\boldsymbol{\mu}_b \in \mathbb{R}^?$ $\boldsymbol{\Sigma}_{aa} \in \mathbb{R}^?$

$\boldsymbol{\Sigma}_{ab} \in \mathbb{R}^?$ $\boldsymbol{\Sigma}_{bb} \in \mathbb{R}^?$

$\boldsymbol{\Sigma}_{ba} \in \mathbb{R}^?$

- A partir de la matriz de precisión:

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}; \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{bmatrix}$$

- Para $p(\mathbf{x}_a \mid \mathbf{x}_b)$; con $p(\mathbf{x}) = p(\mathbf{x}_a, \mathbf{x}_b)$:

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\left([\mathbf{x}_a, \mathbf{x}_b] - [\boldsymbol{\mu}_a, \boldsymbol{\mu}_b]\right)^T \underbrace{\begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}^{-1}}_{\boldsymbol{\Lambda}} \left([\mathbf{x}_a, \mathbf{x}_b] - [\boldsymbol{\mu}_a, \boldsymbol{\mu}_b]\right)$$

$$-\frac{1}{2}(\mathbf{x}^T - \boldsymbol{\mu}^T)\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\left[[\mathbf{x}_a - \boldsymbol{\mu}_a, \mathbf{x}_b - \boldsymbol{\mu}_b]^T \begin{bmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{bmatrix} [\mathbf{x}_a - \boldsymbol{\mu}_a, \mathbf{x}_b - \boldsymbol{\mu}_b]\right]$$

$$-\frac{1}{2}\left[x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu\right] = -\frac{1}{2}x^T\Sigma^{-1}x + \frac{2}{2}x^T\Sigma^{-1}\mu - \frac{1}{2}\mu^T\Sigma^{-1}\mu$$

$x^T\Sigma^{-1}\mu = \mu^T\Sigma^{-1}x = \langle \mu, x\rangle_{\Sigma^{-1}}$, con $\Sigma > 0 \to$ definida positiva

Por lo tanto:

$$-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu \underbrace{-\frac{1}{2}\mu^T\Sigma^{-1}\mu}_{\text{cte en } x} = -\frac{1}{2}\left[[x_a-\mu_a, x_b-\mu_b]^T \begin{bmatrix}\Lambda_{aa} & \Lambda_{ab}\\ \Lambda_{ba} & \Lambda_{bb}\end{bmatrix}[x_a-\mu_a, x_b-\mu_b]\right]$$

$$-\frac{1}{2}\underbrace{x^T\Sigma^{-1}x}_{\substack{\text{cuadrático}\\ \text{en } x}} + \underbrace{x^T\Sigma^{-1}\mu}_{\substack{\text{lineal}\\ \text{en } x}} + cte = -\frac{1}{2}\Big[(x_a-\mu_a)^T\Lambda_{aa} + (x_b-\mu_b)^T\Lambda_{ba}, (x_a-\mu_a)^T\Lambda_{ab} + (x_b-\mu_b)^T\Lambda_{bb}]\ldots$$

$$\ldots[x_a-\mu_a, x_b-\mu_b]\Big]$$

$$= -\frac{1}{2}\Big[(x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) + (x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a)$$
$$+ (x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b) + (x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)\Big]$$

$p(x_a|x_b) = ?$ ENCONTREMOS $\mu_{a|b}$ y $\Sigma_{a|b}$ COMPLETANDO CUADRADOS

$$p(x) = p([x_a, x_b]) \quad ; \quad p(x_a|x_b) = \mathcal{N}(x_a | \mu_{a|b}, \Sigma_{a|b})$$

Reescribiendo:

$$-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = -\frac{1}{2} x_a^T \Lambda_{aa} x_a + x_a^T \Lambda_{aa} \mu_a - \frac{1}{2} \mu_a^T \Lambda_{aa} \mu_a$$

$$- \frac{1}{2} x_b^T \Lambda_{ba} x_a + \frac{1}{2} x_b^T \Lambda_{ba} \mu_a + \frac{1}{2} \mu_b^T \Lambda_{ba} x_a - \frac{1}{2} \mu_b^T \Lambda_{ba} \mu_a$$

$$- \frac{1}{2} x_a^T \Lambda_{ab} x_b + \frac{1}{2} x_a^T \Lambda_{ab} \mu_b + \frac{1}{2} \mu_a^T \Lambda_{ab} x_b - \frac{1}{2} \mu_a^T \Lambda_{ab} \mu_b$$

$$- \frac{1}{2} x_b^T \Lambda_{bb} x_b + x_b^T \Lambda_{bb} \mu_b - \frac{1}{2} \mu_b^T \Lambda_{bb} \mu_b$$

$$\boxed{\Lambda_{ab}^T = \Lambda_{ba}}$$

- Para determinar $p(x_a|x_b)$ encontramos la dependencia de x_a con x_b asumiendo x_b constante.

- Buscamos el término cuadrático en x_a: $-\frac{1}{2} x_a^T \Lambda_{aa} x_a$

- Del término cuadrático $\Sigma_{a|b} = \Lambda_{aa}^{-1}$

- Ahora buscamos los términos lineales en x_a:

$$x_a^T \Lambda_{aa} \mu_a \underbrace{- \frac{1}{2} x_b^T \Lambda_{ba} x_a} \overbrace{+ \frac{1}{2} \mu_b^T \Lambda_{ba} x_a} \underbrace{- \frac{1}{2} x_a^T \Lambda_{ab} x_b} \overbrace{+ \frac{1}{2} x_a^T \Lambda_{ab} \mu_b}$$

- Tenemos que:

$$x_a^T \Lambda_{aa} \mu_a - x_a^T \Lambda_{ab} x_b + x_a^T \Lambda_{ab} \mu_b = x_a^T \left(\Lambda_{aa} \mu_a - \Lambda_{ab} x_b + \Lambda_{ab} \mu_b \right)$$

$$= x_a^T \left(\Lambda_{aa} \mu_a + \Lambda_{ab} (\mu_b - x_b) \right)$$

- Buscamos despejar el término lineal en x desde $x^T \Sigma^{-1} \mu$:

$$x_a^T \Sigma_{a|b}^{-1} \mu_{a|b} = x_a^T \left(\Lambda_{aa} \mu_a - \Lambda_{ab} (\mu_b - x_b) \right)$$

- Sabemos que $\Sigma_{a|b}^{-1} = \Lambda_{aa}$ y:

$$\Sigma_{a|b} \Sigma_{a|b}^{-1} \mu_{a|b} = \Sigma_{a|b} \left(\Lambda_{aa} \mu_a - \Lambda_{ab} (\mu_b - x_b) \right)$$

$$\mu_{a|b} = \Sigma_{a|b} \Sigma_{a|b}^{-1} \mu_a - \Sigma_{a|b} \Sigma_{ab}^{-1} (\mu_b - x_b)$$

$$\boxed{\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (\mu_b - x_b)}$$

NOTA: Dado que:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = \Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

Usando la identidad de la matriz inversa por partes:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}+D^{-1}CMBD^{-1} \end{bmatrix}$$

Siendo $M = (A-BD^{-1}C)^{-1}$; con M^{-1} el complemento de Schur.

EJERCICIO: Demuestre que:

$$\boxed{\begin{aligned} \mu_{a|b} &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b) \\ \Sigma_{a|b} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{aligned}}$$

Nota: $MM^{-1} = I = M^{-1}M$

$$\Sigma_{a|b} = \Lambda_{aa}^{-1}$$

$$\Sigma_{a|b}^{-1} = \Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$

$$(\Sigma_{a|b}^{-1})^{-1} = \Lambda_{aa}^{-1} = \Sigma_{a|b}$$

$$= ((\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1})^{-1}$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}.$$

Marginalización de Gaussianas

- Bajo el mismo contexto podemos encontrar:

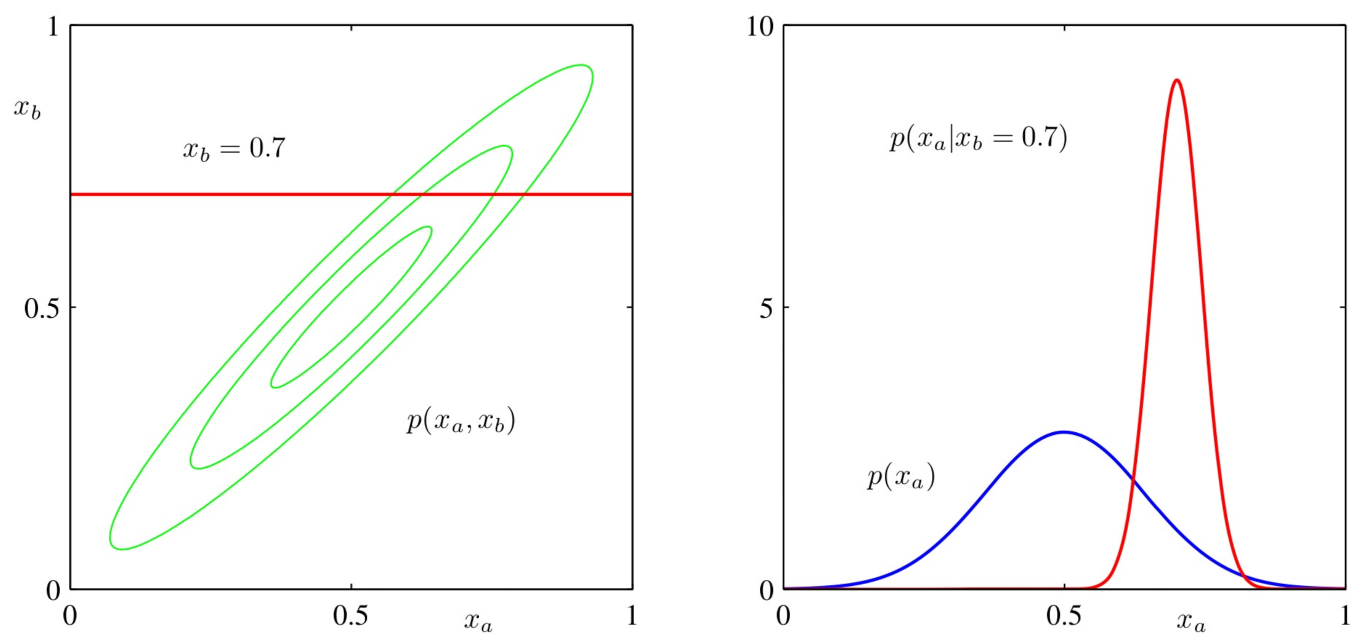
$$\boxed{p(x_a) = \int p(x_a, x_b)\, dx_b}$$

- Expandiendo y buscando términos relacionados con x_b:

$$\boxed{p(x_a) = \mathcal{N}(x_a \mid \mu_a, \Sigma_{aa})}$$

Ejercicio: Demostrar.

Ejercicio: Simular



Teorema de Bayes para variables Gaussianas.

- Sea $p(\mathbf{x})$ la pdf marginal sobre $\mathbf{x} \in \mathbb{R}^M$
- Sea $p(\mathbf{y}|\mathbf{x})$ una distribución Gaussiana condicional con $\mu_{y|x}$ dada como una función lineal en $\mathbf{x}$. y $\Sigma_{y|x}$ independiente de $\mathbf{x}$. $(\mathbf{y} \in \mathbb{R}^D)$
- Las anteriores condiciones permiten generar un modelo linear Gaussiano.

- Nos interesa encontrar la marginal $p(\mathbf{y})$ y la condicional $p(\mathbf{x}|\mathbf{y})$.
- Asumiendo:

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

donde μ, A, b son parámetros que modelan las medias

$\Lambda, \mathbb{L}$ matrices de precisión.

$\mu \in \mathbb{R}^M$, $b \in \mathbb{R}^D$, $A \in \mathbb{R}^{D\times M}$, $\mathbb{L} \in \mathbb{R}^{D\times D}$, $\Lambda \in \mathbb{R}^{M\times M}$

- Sea $z = [x, y]$, la pdf conjunta en log se puede representar como:

$$p(z) = p(x,y) = p(y|x)\,p(x)$$

$$\log(p(z)) = \log(p(y|x)p(x)) = \log(p(y|x)) + \log(p(x))$$

$$= \log\left[\frac{1}{(2\pi)^{D/2}|\mathbb{L}^{-1}|^{1/2}}\exp\left[-\frac{1}{2}(y-Ax-b)^T\mathbb{L}(y-Ax-b)\right]\right]$$

$$+\log\left[\frac{1}{(2\pi)^{M/2}\Lambda}\exp\left[-\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)\right]\right]$$

$$= -\frac{1}{2}(y-Ax-b)^T\mathbb{L}(y-Ax-b) - \frac{1}{2}(x-\mu)^T\Lambda(x-\mu) + cte.$$

EJERCICIO: Determine el termino cte.

$$\log(p(z)) = \log(p(y,x)) = -\frac{1}{2}\left(y^T-(Ax)^T-b^T\right)L\left(y-Ax-b\right) - \frac{1}{2}\left(x^T-\mu^T\right)\Lambda\left(x-\mu\right) + cte.$$

$$= -\frac{1}{2}\Big[y^TLy - y^TLAx - y^TLb - x^TA^TLy + x^TA^TLAx + x^TA^TLb - b^TLy + b^TLAx + b^TLb\Big]$$

$$-\frac{1}{2}\Big[x^T\Lambda x - x^T\Lambda\mu - \mu^T\Lambda x + \mu^T\Lambda\mu\Big] + cte.$$

- Factorizamos para encontrar terminos cuadráticos de la forma:

$$x^TA_{xx}x, \quad y^TA_{yy}y, \quad y^TA_{yx}x, \quad x^TA_{xy}y.$$

$$\log(p(y,x)) = -\frac{1}{2}y^TLy - \frac{1}{2}x^T(\Lambda + A^TLA)x + \frac{1}{2}y^TLAx + \frac{1}{2}x^TA^TLy$$

$+$ lineales en x y y $+$ cte

Así:

$$-\frac{1}{2}\begin{bmatrix} x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} \Lambda + A^TLA & -A^TL \\ -LA & L \end{bmatrix}}_{\text{MATRIZ DE PRECISIÓN DISTRIBUCIÓN CONJUNTA}} \begin{bmatrix} x \\ y \end{bmatrix} = -\frac{1}{2}z^TRz;$$

EJERCICIO: Utilizando el teorema de la matriz inversa, demostrar que:

$$R^{-1} = \begin{bmatrix} \Lambda + A^T L A & -A^T L \\ -LA & L \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1} A^T \\ A \Lambda^{-1} & L^{-1} + A \Lambda^{-1} A^T \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

- Ahora, agrupando los términos lineales en $\log(p(z))$:

$$\log(p(z)) = \frac{2}{2} y^T L b - \frac{2}{2} x^T A^T L b + \frac{2}{2} x^T \Lambda \mu + \text{términos cuadráticos} + \text{cte.}$$

$$= x^T(\Lambda \mu - A^T L b) + y^T L b + \text{términos cuadráticos} + \text{cte.}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} \Lambda \mu - A^T L b \\ L b \end{bmatrix}$$

NOTA: Recuerda que $-\frac{1}{2}(z-\mu_z)^T \Sigma_z^{-1} (z-\mu_z) = -\frac{1}{2} z^T \Sigma_z^{-1} z + z^T \Sigma_z^{-1} \mu_z + \text{cte.}$

Entonces:

$$z^T \Sigma_z^{-1} \mu_z = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} \Lambda \mu - A^T L b \\ L b \end{bmatrix}$$

$$\Sigma_z^{-1} \mu_z = \begin{bmatrix} \Lambda \mu - A^T L b \\ L b \end{bmatrix}$$

- Reescribiendo desde la matriz de precisión: $(R = \Sigma_z^{-1};\ R^{-1} = \Sigma_z)$

$$-\frac{1}{2}(z-\mu_z)^T R (z-\mu_z) = -\frac{1}{2} z^T R z + z^T R \mu_z + \text{cte.}$$

$$z^T R \mu_z = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} \Lambda\mu - A^T L b \\ L b \end{bmatrix}; \quad \Sigma_z = R^{-1} = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

$$R\mu_z = \begin{bmatrix} \Lambda\mu - A^T L b \\ L b \end{bmatrix} \rightarrow \quad \mu_z = R^{-1} \begin{bmatrix} \Lambda\mu - A^T L b \\ L b \end{bmatrix}$$

$$\mu_z = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{bmatrix} \begin{bmatrix} \Lambda\mu - A^T L b \\ L b \end{bmatrix} = \begin{bmatrix} \Lambda^{-1}(\Lambda\mu - A^T L b) + \Lambda^{-1}A^T L b \\ A\Lambda^{-1}(\Lambda\mu - A^T L b) + (L^{-1} + A\Lambda^{-1}A^T) L b \end{bmatrix}$$

$$\mu_z = \begin{bmatrix} \mu - \underline{\Lambda^{-1}A^T L b} + \underline{\Lambda^{-1}A^T L b} \\ A\underline{\Lambda^{-1}\Lambda}\mu - \underline{A\Lambda^{-1}A^T L b} + \underline{L^{-1}L}b + \underline{A\Lambda^{-1}A^T L b} \end{bmatrix} = \begin{bmatrix} \mu \\ b + A\mu \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

$$\boxed{\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu \\ b + A\mu \end{bmatrix}}\ ;$$

$$\boxed{\begin{aligned} p(x) &= \mathcal{N}(x \mid \mu, \Lambda^{-1}) \\ p(y|x) &= \mathcal{N}(y \mid Ax + b, L^{-1}) \\ p(y) &= \mathcal{N}(y \mid \mu_y, \Sigma_{yy}) = \mathcal{N}(y \mid b + A\mu, L^{-1} + A\Lambda^{-1}A^T) \end{aligned}}$$

NOTA: Si $A = I \rightarrow$ la marginal $p(y)$ se puede escribir como la convolución de dos Gaussianas:

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

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

Ahora, para $p(x|y)$, recordemos que para $p(x_a, x_b)$:

$$p(x_a|x_b) = \mathcal{N}(x_a|\mu_{a|b}, \Sigma_{a|b})$$

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \; ; \quad \mu_{a|b} = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \; ; \quad \Sigma_{a|b}^{-1} = \Lambda_{aa} \; ; \quad \Sigma_{a|b} = \Lambda_{aa}^{-1}$$

Entonces, a partir de la matriz de precisión y μ_z:

$$R = \begin{bmatrix} \Lambda + A^T L A & -A^T L \\ -LA & L \end{bmatrix} = \begin{bmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{bmatrix} ; \quad \mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu \\ b + A\mu \end{bmatrix}$$

$$\Sigma_{x|y} = \Lambda_{xx}^{-1} = (\Lambda + A^T L A)^{-1}$$

$$\mu_{x|y} = \mu_x - \Lambda_{xx}^{-1}\Lambda_{xy}(y - \mu_y) = \mu + (\Lambda + A^T L A)^{-1} A^T L (y - A\mu - b)$$

$$\Sigma_{x|y} = (\Lambda + A^T L A)^{-1}$$

$$\mu_{x|y} = \mu + (\Lambda + A^T L A)^{-1}(A^T L y - A^T L b - A^T L A \mu)$$

$$= (\Lambda + A^T L A)^{-1}(\Lambda + A^T L A)\mu + (\Lambda + A^T L A)^{-1}(A^T L (y-b) - A^T L A \mu)$$

$$= (\Lambda + A^T L A)^{-1}\left[A^T L (y-b) - \underbrace{A^T L A \mu} + \Lambda \mu + \underbrace{A^T L A \mu}\right]$$

$$\mu_{x|y} = (\Lambda + A^T L A)^{-1}\left[A^T L (y-b) + \Lambda \mu\right]$$

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

EJERCICIO: Sea $t_n = \mathbb{w}^T \phi(x_n) + \eta$;

con $\mathbb{w} \sim p(\mathbb{w}) = \mathcal{N}(\mathbb{w} \mid m_0, S_0)$ y $\eta \sim p(\eta) = \mathcal{N}(\eta \mid 0, \beta^{-1})$

Demuestre que $p(\mathbb{w} \mid \mathbb{t}) = \mathcal{N}(\mathbb{w} \mid m_N, S_N)$

$$m_N = S_N (S_0^{-1} m_0 + \beta \phi^T \mathbb{t})$$

$$S_N^{-1} = S_0^{-1} + \beta \phi^T \phi$$

$\phi = [\phi(x_1), \phi(x_2), \ldots, \phi(x_N)]^T \in \mathbb{R}^{Q \times N}$; $\mathbb{t} = [t_1, t_2, \ldots, t_N]^T \in \mathbb{R}^N$

$\mathbb{w} \in \mathbb{R}^Q$ $\qquad$ $\beta \in \mathbb{R}^+$

Demuestre que la predictiva toma la forma:

$$p(t^* \mid \mathbb{t}, \phi, \mathbb{w}) = \int p(t^* \mid \mathbb{w})\, p(\mathbb{w} \mid \mathbb{t})\, \partial \mathbb{w}$$

$$p(t^* \mid \mathbb{t}, \phi, \mathbb{w}) = \mathcal{N}(t^* \mid m_N^T \phi(x^*), \sigma_N^2(x^*))$$

$$\sigma_N^2(x^*) = \frac{1}{\beta} + \phi(x^*)^T S_N \phi(x^*)$$