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

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

- La solución <u>puntual</u> por máx verosimilitud:

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

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

- Podemos estimar la distribución predictiva sobre t:

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

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

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

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

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

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

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 $p(t|\phi(\mathbf{x}),\Phi,\mathbf{w})$, se puede modelar como:

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

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

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

$$\mu(\mathbf{x}) = \frac{1}{\sigma^2}\phi(\mathbf{x})\, S \sum_{n=1}^{N} \phi(\mathbf{x}_n) t_n = \frac{1}{\sigma^2} \underset{1\times Q}{\phi(\mathbf{x})}\, \underset{Q\times Q}{S}\, \underset{Q\times N}{\Phi^T}\, \underset{N\times 1}{\mathbf{t}}$$

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

$$S = \left[\frac{1}{\sigma_w^2} I + \frac{1}{\sigma^2}\Phi^T\Phi\right]^{-1}_{Q\times Q}$$

¿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:

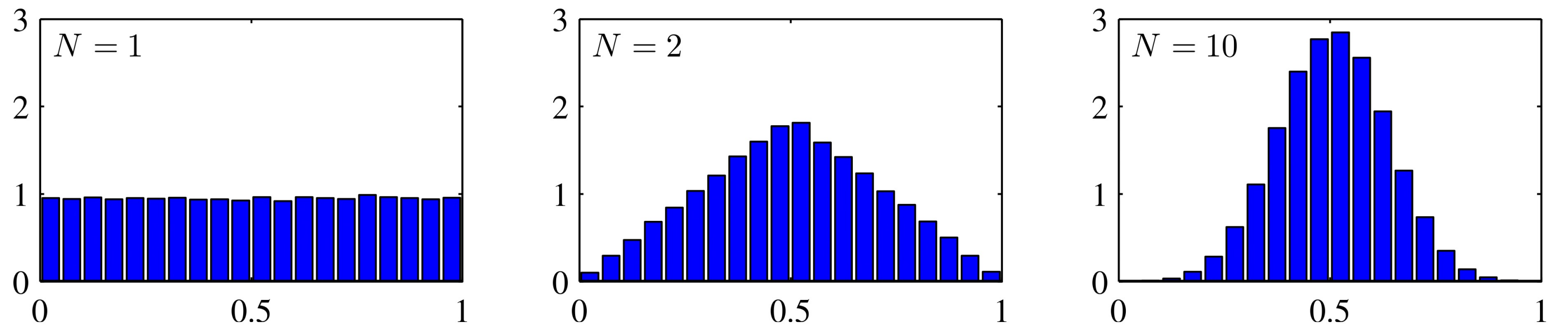
  $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^p$; $\Sigma \in \mathbb{R}^{p\times p}$

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

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:

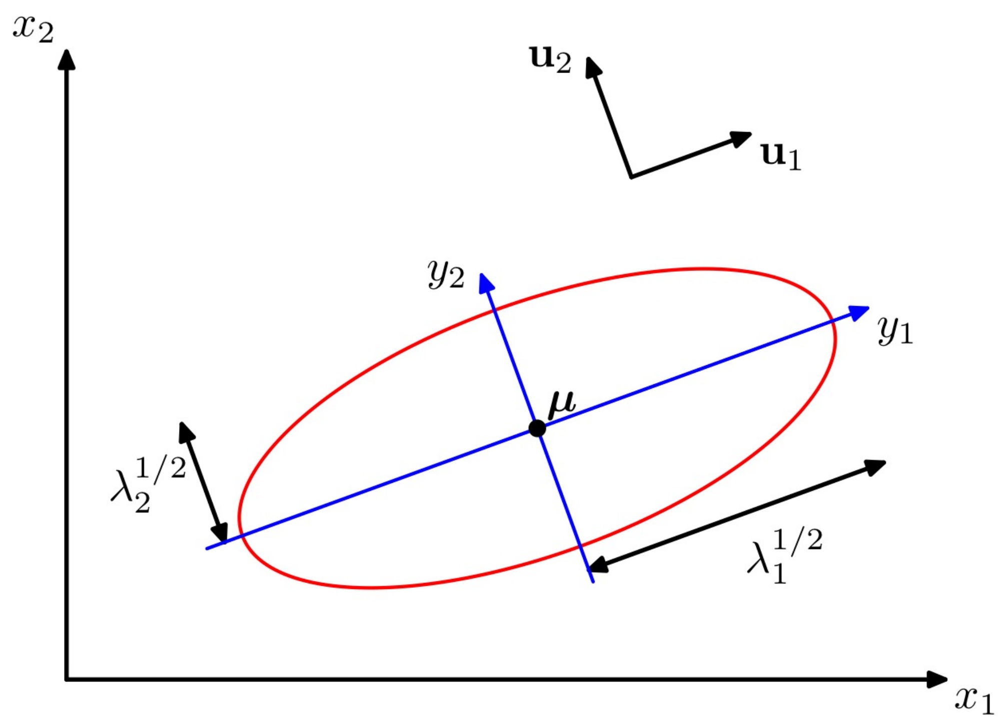
$$\Delta^2 = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

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

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

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

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



Distribuciones Gaussianas condicionales

- Supongamos $x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}$; $\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}$; $\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$

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

con $\Sigma_{ab} = \Sigma_{ba}^{T}$.

Ejercicio: Si $x \in \mathbb{R}^{P}$, $x_a \in \mathbb{R}^{\bar{P}}$

$x_b \in ?$ $\mu \in \mathbb{R}^{?}$ $\mu_a \in \mathbb{R}^{?}$

$\mu_b \in \mathbb{R}^{?}$ $\Sigma_{aa} \in \mathbb{R}^{?}$

$\Sigma_{ab} \in \mathbb{R}^{?}$ $\Sigma_{bb} \in \mathbb{R}^{?}$

$\Sigma_{ba} \in \mathbb{R}^{?}$

- A partir de la matriz de precisión:

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

- Para $p(x_a \mid x_b)$; con $p(x) = p(x_a, x_b)$:

$$-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) = -\frac{1}{2}\left([x_a, x_b] - [\mu_a, \mu_b] \right)^T \underbrace{\begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1}}_{\Lambda} \left([x_a, x_b] - [\mu_a, \mu_b] \right)$$

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

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

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

Por lo tanto:

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

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

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

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

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

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

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:

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

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

$$\mu_{a|b} = \mu_a + \Lambda_{aa}^{-1}\Lambda_{ab}(\mu_b - x_b)$$; dado que:

$$(\Sigma_{a|b}^{-1})^{-1} = (\Lambda_{aa})^{-1} \rightarrow \Sigma_{a|b} = \Lambda_{aa}^{-1} \quad y \quad \Lambda_{aa} = M = (A - BD^{-1}C)^{-1} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$

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

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

Además:

$$\mu_{a|b} = \mu_a + \Sigma_{a|b}\Lambda_{ab}(\mu_b - x_b);$$ con $$\Lambda_{ab} = -MBD^{-1} = -(A - BD^{-1}C)^{-1}BD^{-1}$$

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

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

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

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

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

Marginalización de Gaussianas

- Bajo el mismo contexto podemos encontrar:

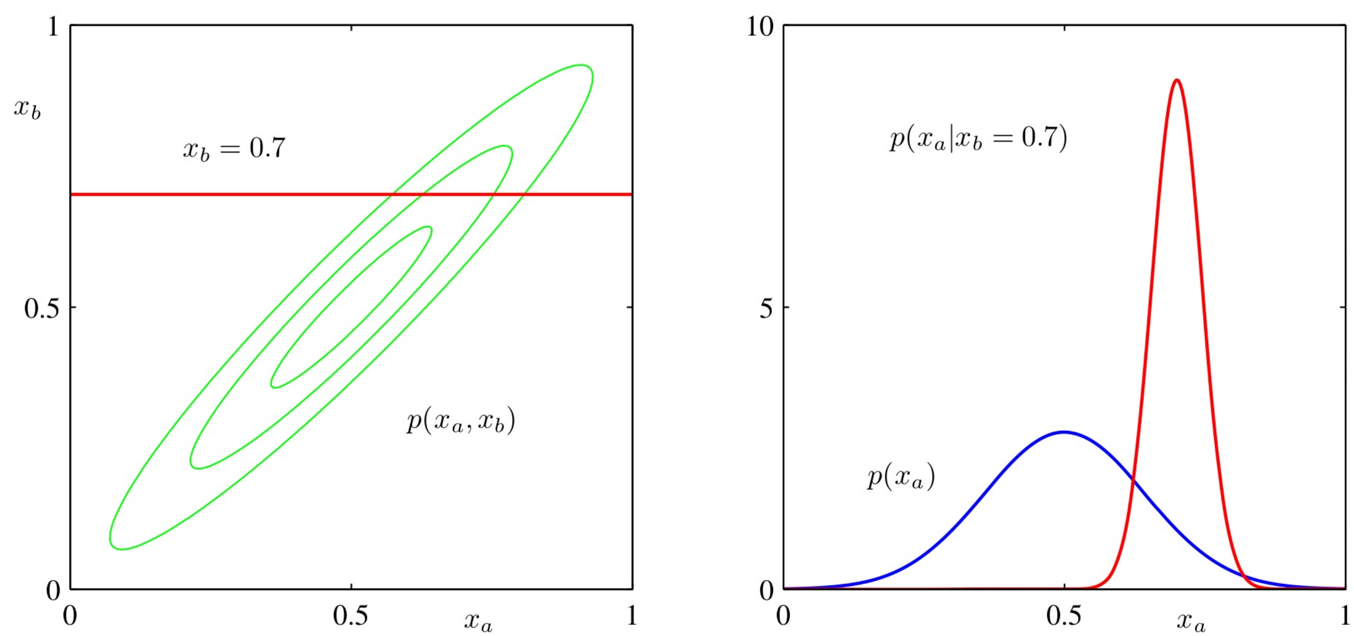
$$\boxed{p(x_a) = \int p(x_a, x_b)\, dx_b}$$

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

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

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

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

Λ, L matrices de precisión.

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

- 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.

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

$$= -\frac{1}{2}\left[Y^T L Y - Y^T L A x - Y^T L b - x^T A^T L Y + x^T A^T L A x + x^T A^T L b - b^T L Y + b^T L A x + b^T L b\right]$$
$$-\frac{1}{2}\left[x^T \Lambda x - x^T \Lambda \mu - \mu^T \Lambda x + \mu^T \Lambda \mu\right] + cte.$$

- Factorizamos para encontrar términos cuadráticos de la forma:

$$x^T A_{xx} x, \quad Y^T A_{YY} Y, \quad Y^T A_{Yx} x, \quad x^T A_{xY} Y.$$

$$\log(p(Y,x)) = -\frac{1}{2} Y^T L Y - \frac{1}{2} x^T(\Lambda + A^T L A)x + \frac{1}{2} Y^T L A x + \frac{1}{2} x^T A^T L Y$$
$$+ \text{lineales en } x \text{ y } Y + cte$$

Así:

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

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: $(R = \Sigma_z^{-1};\ R^{-1} = \Sigma_z)$

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

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

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

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

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

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

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

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 $p(x_a, x_b)$:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$S_N^{-1} = S_0^{-1} + \beta \phi^T \phi \in \mathbb{R}^{Q \times Q}$$

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

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

Demuestre que la predictiva toma la forma:

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

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

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

Si: $p(x) = \mathcal{N}(x \mid \mu, \Delta^{-1})$; $p(y \mid x) = \mathcal{N}(y \mid Ax+b, L^{-1})$

con $y = f(x \mid A, b) = Ax+b$.

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

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

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

Para: $t_n = w^T \phi(x_n) + \eta$; $\phi(x_n), w \in \mathbb{R}^{Q \times 1}$; $\Phi \in \mathbb{R}^{N \times Q}$; $\mathbf{t} \in \mathbb{R}^{N \times 1}$; $\beta \in \mathbb{R}^+$

$\eta \sim \mathcal{N}(\eta \mid 0, \beta^{-1}) \rightarrow \eta = t_n - w^T \phi(x_n) \sim \mathcal{N}(t_n - w^T \phi(x_n) \mid 0, \beta^{-1})$

Entonces: $p(t_n \mid w^T \phi(x_n), \beta^{-1}) = \mathcal{N}(t_n \mid w^T \phi(x_n), \beta^{-1}) \rightarrow p(\mathbf{t} \mid w) = \mathcal{N}(\mathbf{t} \mid \Phi w, \beta^{-1} I)$

Ahora: $w \sim p(w) = \mathcal{N}(w \mid m_0, S_0) \longrightarrow \mu = m_0$; $\Delta^{-1} = S_0$; $S_0^{-1} = \Delta$

$A = \Phi$, $x = w$, $L^{-1} = \beta^{-1} I$, $b = 0$

$L^{-1} = \beta^{-1} I$

$L = \beta I$

POSTERIOR: $p(w \mid \mathbf{t}) = \mathcal{N}(w \mid \mu_{w|t}, \Sigma_{w|t})$

$$m_N = \mu_{w|t} = (S_0^{-1} + \Phi^T \beta I \Phi)^{-1} [\Phi^T \beta I (\mathbf{t} - 0) + S_0^{-1} m_0]$$

$$m_N = \mu_{w|t} = (S_0^{-1} + \beta \Phi^T \Phi)^{-1} [\beta \Phi^T \mathbf{t} + S_0^{-1} m_0]$$

$$\boxed{m_N = \mu_{w|t} = S_N (S_0^{-1} m_0 + \beta \Phi^T \mathbf{t})}$$

$$S_N = (S_0^{-1} + \beta \Phi^T \Phi)^{-1}; \quad \boxed{S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi}$$

Para la predictiva: $t_* = w^T \phi(x_*) + \eta$; x_*: entrada nueva

t_*: salida nueva → predicción

↓ predictiva

$$p(t_* | t, \phi, w) = \mathcal{E}_{w|t}\{p(t_*|w)\} = \int p(t_*|w)\, p(w|t)\, \partial w = \mathcal{E}_{w|t}\{N(t^*| w^T\phi(x_*), \beta^{-1})\}$$

Dado que: $\eta = t_* - w^T\phi(x_*) \sim N(t_* - w^T\phi(x_*) | 0, \beta^{-1})$

$p(t^*|w) = N(t_* | w^T\phi(x_*), \beta^{-1})$ → verosimilitud para t_*

$p(w|t) = N(w | \mu_{w|t}, \Sigma_{w|t}) = N(w | m_N, S_N)$ → posterior

$$p(t_* | t, \phi, w) = \int N(t_* | w^T\phi(x_n), \beta^{-1})\, N(w | m_N, S_N)\, \partial w.$$

$$p(t_* | t, \phi, w) = \mathcal{E}_{w|t}\{p(t^*|w)\}$$

Ahora, si se analiza desde la marginalización de la predictiva:

$$p(t_*|w) = \mathcal{N}(t_*|w^T\phi(x_*), \beta^{-1})$$

$$p(w|t) = \mathcal{N}(w|m_N, S_N)$$

$$p(t_*|w,t) = \int p(t_*, w|t)\, \partial w$$

$$p(t_*|t,w) = \int p(t_*|w)\, p(w|t)\, \partial w$$

$$w^T\phi(x_*) = \underbrace{\phi^T(x_*)}_{A}\underbrace{w}_{x} \; ; \; b =$$

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

$$p(x|y) = \mathcal{N}(x|\Sigma(A^TL(y-b)+\Lambda\mu, \Sigma)$$

$$\Sigma = (\Lambda + A^TLA)^{-1}$$

$$p(x) = \mathcal{N}(x|\mu, \Lambda^{-1})$$

$$p(y) = \int p(x,y)\,\partial x = \int p(y|x)\,p(x)\,\partial x$$

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

$$\boxed{\begin{aligned} p(t_*|t,w) &= \mathcal{N}(t_*|\phi^T(x_*)m_N, \beta^{-1}+\phi^T(x_*)S_N\phi(x_*)) \\ p(t_*|t,w) &= \mathcal{N}(t_*|m_N^T\phi(x_*), \beta^{-1}+\phi^T(x_*)S_N\phi(x_*)) \end{aligned}}$$