

Regresión Bayesiana - Caso Particular Probabilidad Gaussiana

- En Máxima Verosimilitud:

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

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

- Asumiendo datos i.i.d.: $p(t | \phi, w, \sigma^2) = \prod_{n=1}^N N(t_n | \phi(x_n)w, \sigma^2)$
 $\phi \in \mathbb{R}^{N \times Q}; t \in \mathbb{R}^N$

- La solución puntual por máx verosimilitud:

$$w_{ML} = (\phi^T \phi)^{-1} \phi^T t; \quad B_{ML} = \frac{1}{\sigma_{ML}^2}$$

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

- Podemos estimar la distribución predictiva sobre t :

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

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

$$p(w|\sigma_w^2) = N(w|0, \sigma_w^2 I)$$

$$p(t, \phi, w, \sigma^2, \sigma_w^2) = p(t|\phi w, \sigma^2) p(w|\sigma_w^2)$$

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

$$p(w|t, \phi, \sigma^2, \sigma_w^2) \propto p(t|\phi w, \sigma^2) p(w|\sigma_w^2)$$

MÁX A-POSTERIORI (MAP)

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

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

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:

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

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

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

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

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

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

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

Distribución Gaussiana.

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

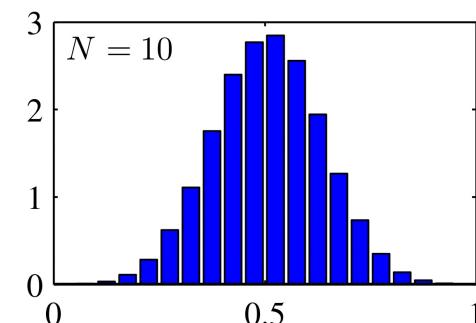
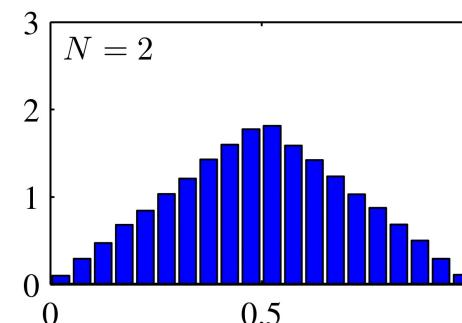
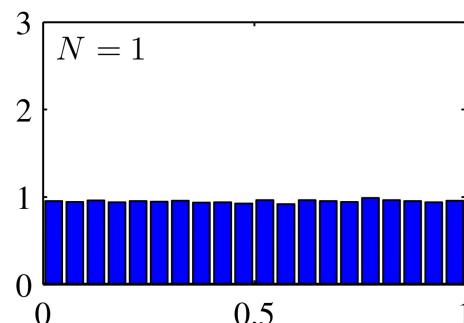
- Gaussiana multivariada:

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

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

EJERCICIO: - Demuestre, mediante simulación, que para $X_n \sim p(x_n) = U(x_n|a, b)$
 $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N X_n \sim N\left(\frac{1}{N} \sum_{n=1}^N x_n | \mu_x, \sigma_x^2\right)$. Teorema del límite central.

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



- La forma cuadrática en la Gaussiana Multivariada:

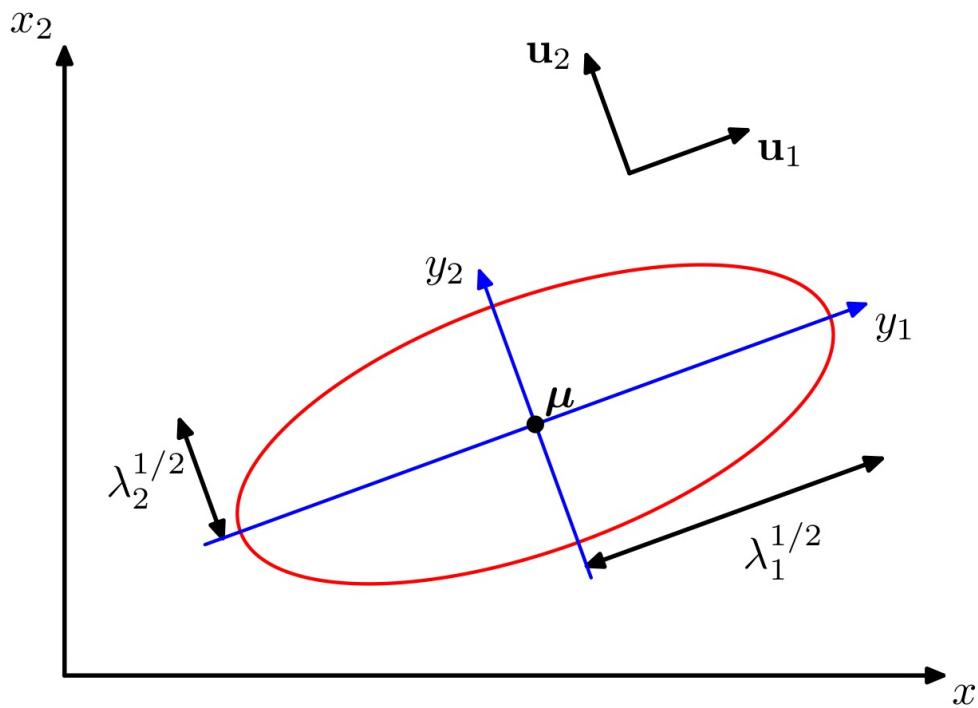
$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

$$\Sigma = \sum_{j=1}^p \lambda_j \mathbf{u}_j \mathbf{u}_j^\top, \quad \Sigma \mathbf{u}_j = \lambda_j \mathbf{u}_j \rightarrow \text{eigen decomposition.}$$

$$\Sigma^{-1} = (\mathbf{U} \Delta \mathbf{U}^\top)^{-1} = \mathbf{U} \Delta^{-1} \mathbf{U}^\top = \sum_{j=1}^p \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^\top$$

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \sum_{j=1}^p \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^\top (\mathbf{x} - \boldsymbol{\mu}) = \sum_{j=1}^p \frac{1}{\lambda_j} \underbrace{(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{u}_j}_{y_j} \underbrace{\mathbf{u}_j^\top (\mathbf{x} - \boldsymbol{\mu})}_{y_j}$$

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



Distribuciones Gaussianas condicionales

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

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

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

Ejercicio: Si $\mathbf{x} \in \mathbb{R}^P$, $x_a \in \mathbb{R}^P$
 $x_b \in ?$ $\boldsymbol{\mu} \in ?$ $\boldsymbol{\mu}_a \in \mathbb{R}^?$

$\boldsymbol{\mu}_b \in \mathbb{R}^?$ $\Sigma_{aa} \in \mathbb{R}^?$
 $\Sigma_{ab} \in \mathbb{R}^?$ $\Sigma_{bb} \in \mathbb{R}^?$
 $\Sigma_{ba} \in \mathbb{R}^?$

- A partir de la matriz de precisión:

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

- Para $p(x_a | x_b)$; con $p(\mathbf{x}) = p(x_a, x_b)$:

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

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

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

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

Por lo tanto:

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

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

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

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

$p(x_a | x_b) = ?$ ENCONTRAREMOS Λ_{ab} y Σ_{ab} 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:

$$\begin{aligned}
 -\frac{1}{2} (x - \mu)^T \bar{\Sigma}^{-1} (x - \mu) &= -\frac{1}{2} \cancel{x_a^T \Delta_{aa} x_a} + \cancel{x_a^T \Delta_{aa} \mu_a} - \frac{1}{2} \mu_a^T \Delta_{aa} \mu_a \\
 &\quad - \cancel{\frac{1}{2} x_b^T \Delta_{bb} x_b} + \frac{1}{2} x_b^T \Delta_{bb} \mu_a + \cancel{\frac{1}{2} \mu_b^T \Delta_{bb} x_a} - \frac{1}{2} \mu_b^T \Delta_{bb} \mu_a \\
 &\quad - \cancel{\frac{1}{2} x_a^T \Delta_{ab} x_b} + \cancel{\frac{1}{2} x_a^T \Delta_{ab} \mu_b} + \frac{1}{2} \mu_a^T \Delta_{ab} x_b - \frac{1}{2} \mu_a^T \Delta_{ab} \mu_b \\
 &\quad - \frac{1}{2} x_b^T \Delta_{bb} x_b + x_b^T \Delta_{bb} \mu_b - \frac{1}{2} \mu_b^T \Delta_{bb} \mu_b
 \end{aligned}$$

- 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 : $-\frac{1}{2} x_a^T \Delta_{aa} x_a$
- Del término cuadrático $\Sigma_{ab} = \Delta_{aa}$
- Ahora buscamos los términos lineales en x_a :

$$\underbrace{x_a^T \Delta_{aa} \mu_a}_{-\frac{1}{2} x_a^T \Delta_{aa} x_a} + \underbrace{\frac{1}{2} \mu_b^T \Delta_{bb} x_a}_{\cancel{x_b^T \Delta_{bb} x_a}} - \underbrace{\frac{1}{2} x_a^T \Delta_{ab} \mu_b}_{\cancel{\frac{1}{2} x_a^T \Delta_{ab} x_b}} + \underbrace{\frac{1}{2} x_a^T \Delta_{ab} \mu_b}_{\cancel{\frac{1}{2} \mu_a^T \Delta_{ab} x_b}}$$

- Tenemos que:

$$\mathbf{x}_a^T \Delta_{aa} \mathbf{u}_{aa} - \mathbf{x}_a^T \Delta_{ab} \mathbf{u}_b + \mathbf{x}_a^T \Delta_{ab} \mathbf{u}_b = \mathbf{x}_a^T (\Delta_{aa} \mathbf{u}_{aa} - \Delta_{ab} \mathbf{u}_b + \Delta_{ab} \mathbf{u}_b)$$
$$= \mathbf{x}_a^T (\Delta_{aa} \mathbf{u}_{aa} + \Delta_{ab} (\mathbf{u}_b - \mathbf{x}_b))$$

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

$$\mathbf{x}_a^T \bar{\Sigma}_{ab}^{-1} \mathbf{u}_{ab} = \mathbf{x}_a^T (\Delta_{aa} \mathbf{u}_{aa} + \Delta_{ab} (\mathbf{u}_b - \mathbf{x}_b))$$

- Sabemos que $\bar{\Sigma}_{ab}^{-1} = \Delta_{aa}$ y:

$$\bar{\Sigma}_{ab}^{-1} \bar{\Sigma}_{ab}^{-1} \mathbf{u}_{ab} = \sum_{ab} (\Delta_{aa} \mathbf{u}_{aa} + \Delta_{ab} (\mathbf{u}_b - \mathbf{x}_b))$$

$$\mathbf{u}_{ab} = \bar{\Sigma}_{ab}^{-1} \bar{\Sigma}_{ab}^{-1} \mathbf{u}_a + \bar{\Sigma}_{ab}^{-1} \bar{\Sigma}_{ab}^{-1} (\mathbf{u}_b - \mathbf{x}_b)$$

$$\boxed{\mathbf{u}_{ab} = \mathbf{u}_a + \Delta_{aa}^{-1} \Delta_{ab} (\mathbf{u}_b - \mathbf{x}_b)}$$

NOTA: Dados que: $\bar{\Sigma}^{-1} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = M^{-1} \begin{bmatrix} M_{aa} & M_{ab} \\ M_{ba} & M_{bb} \end{bmatrix}$

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 D^{-1} el complemento de Schur.

EJERCICIO: Demuestre que:

Nota: $M\bar{M}^{-1} = I = \bar{M}^{-1}M$

$$M_{ab} = M_{aa} + \Sigma_{ab} \bar{\Sigma}_{bb}^{-1} (\bar{x}_b - \bar{M}_{bb})$$

$$\bar{\Sigma}_{ab} = \Sigma_{aa} - \Sigma_{ab} \bar{\Sigma}_{bb}^{-1} \Sigma_{ba}$$

$$M_{a1b} = M_a + \Delta_{aa}^{-1} \Delta_{ab} (M_b - X_b); \text{ dado que:}$$

$$(\Sigma_{a1b})^{-1} = (\Delta_{aa})^{-1} \rightarrow \Sigma_{a1b} = (\Delta_{aa})^{-1} \quad y \quad \Delta_{aa} = M = (A - BD^T C)^{-1} = (\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1}$$

$$\Sigma_{a1b} = (\Delta_{aa})^{-1} = ((\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1})^{-1}$$

$$\boxed{\Sigma_{a1b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}}$$

Además:

$$M_{a1b} = M_a + \Sigma_{a1b} \Delta_{ab} (M_b - X_b); \quad \text{con} \quad \Delta_{ab} = -MBD^T = -(A - BD^T C)^T B D^T$$

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

$$\Delta_{ab} = -\Sigma_{a1b}^{-1} \Sigma_{ab} \Sigma_{bb}^{-1}$$

$$M_{a1b} = M_a - \Sigma_{a1b} \Sigma_{a1b}^{-1} \Sigma_{ab} \Sigma_{bb}^{-1} (M_b - X_b)$$

$$M_{a1b} = M_a - \Sigma_{ab} \Sigma_{bb}^{-1} (M_b - X_b)$$

$$\boxed{M_{a1b} = M_a + \Sigma_{ab} \Sigma_{bb}^{-1} (X_b - M_b)}$$

Marginalización de Gaussianas

- Bajo el mismo contexto podemos encontrar:

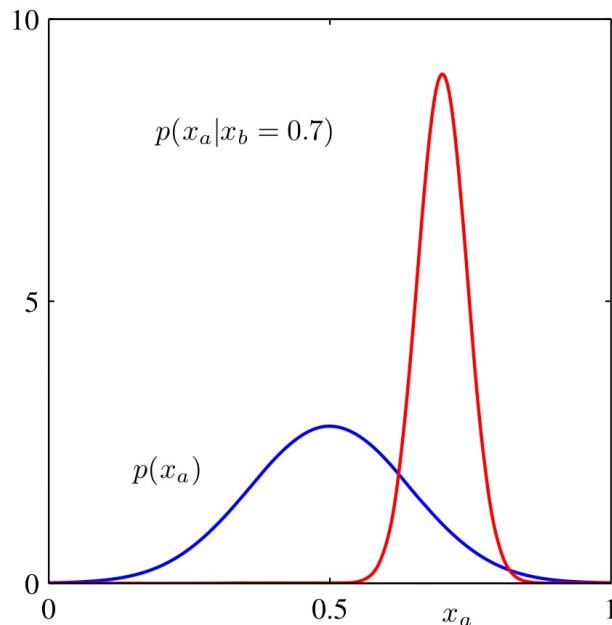
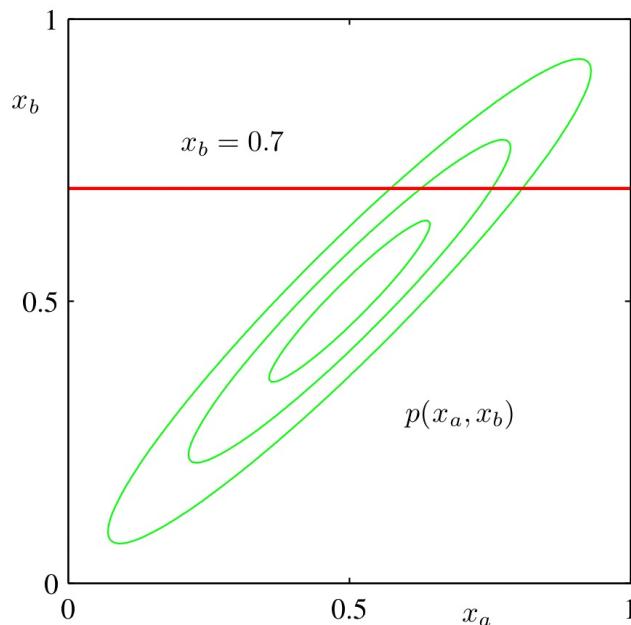
$$p(x_a) = \int p(x_a, x_b) dx_b$$

- Expandiendo y buscando términos relacionados con x_b :

$$p(x_a) = N(x_a | \mu_a, \Sigma_{aa})$$

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 . $\rightarrow \bar{E}_{y|x}$ 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:
 - $p(x) = N(x|\mu, \Sigma')$
 - $p(y|x) = N(y | Ax+b, L^{-1})$

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

$$\Sigma, L \text{ matrices de precisión.}$$
$$\mu \in \mathbb{R}^M, b \in \mathbb{R}^D, A \in \mathbb{R}^{D \times M}, L \in \mathbb{R}^{D \times D}, \Sigma \in \mathbb{R}^{M \times M}$$

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

$$p(\mathbf{z}) = p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

$$\begin{aligned}\log(p(\mathbf{z})) &= \log(p(\mathbf{y}|\mathbf{x})p(\mathbf{x})) = \log(p(\mathbf{y}|\mathbf{x})) + \log(p(\mathbf{x})) \\ &= \log\left[\frac{1}{(2\pi)^{D/2} |\mathbf{L}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{L}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})\right]\right] \\ &\quad + \log\left[\frac{1}{(2\pi)^{n/2} \Delta} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{u})^T \Delta^{-1} (\mathbf{x} - \mathbf{u})\right]\right] \\ &= -\frac{1}{2} (\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{L}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b}) - \frac{1}{2} (\mathbf{x} - \mathbf{u})^T \Delta^{-1} (\mathbf{x} - \mathbf{u}) + \text{cte.}\end{aligned}$$

EJERCICIO: Determine el término cte.

$$\log(p(z)) = \log(p(y, x)) = -\frac{1}{2}(y^T - (Ax)^T - b^T) L (y - Ax - b) \\ - \frac{1}{2}(x^T - u^T) \Delta (x - u) + \text{cte.}$$

$$= -\frac{1}{2} [\underbrace{y^T L y}_{} - \underbrace{y^T L A x}_{} - \underbrace{y^T L b}_{} - \underbrace{x^T A^T L y}_{} + \underbrace{x^T A^T L A x}_{} + \underbrace{x^T A^T L b}_{} \\ - \underbrace{b^T L y}_{} + \underbrace{b^T L A x}_{} + \underbrace{b^T L b}_{}] \\ - \frac{1}{2} [\underbrace{x^T \Delta x}_{} - \underbrace{x^T \Delta u}_{} - \underbrace{u^T \Delta x}_{} + \underbrace{u^T \Delta u}_{}] + \text{cte.}$$

- Factorizamos para encontrar términos cuadráticos de la forma:

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

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

Así:

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

MÁTRIZ DE PRECISIÓN DISTRIBUCIÓN
CONJUNTA

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

$$R^{-1} = \begin{bmatrix} I + A^T L A & -A^T L \\ -L A & L \end{bmatrix}^{-1} = \begin{bmatrix} I' & I' A^T \\ A I' & C' + A I' A^T \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

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

$$\log(p(z)) = \frac{z}{2} Y^T L b - \frac{z}{2} X^T A^T L b + \frac{z}{2} X^T S u + \text{términos cuadráticos} + \text{cte.}$$

$$= X^T (S u - A^T L b) + Y^T L b + \text{términos cuadráticos} + \text{cte.}$$

$$\begin{bmatrix} X \\ Y \end{bmatrix}^T \begin{bmatrix} S u - A^T L b \\ L b \end{bmatrix}$$

NOTA: Recuerda que $-\frac{1}{2}(z - u_z)^T \bar{\Sigma}_z^{-1} (z - u_z) = -\frac{1}{2} z^T \bar{\Sigma}_z^{-1} z + z^T \bar{\Sigma}_z^{-1} u_z + \text{cte.}$

Entonces: $z^T \bar{\Sigma}_z^{-1} u_z = \begin{bmatrix} X \\ Y \end{bmatrix}^T \begin{bmatrix} S u - A^T L b \\ L b \end{bmatrix}$

$$\bar{\Sigma}_z^{-1} u_z = \begin{bmatrix} S u - A^T L b \\ L b \end{bmatrix}$$

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

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

$$z^T R \mu_z = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} \Delta u - A^T L b \\ L b \end{bmatrix}; \Sigma_z = R^{-1} = \begin{bmatrix} \bar{\Lambda}' & \bar{\Lambda}' A^T \\ A \bar{\Lambda}' & L' + A \bar{\Lambda}' A^T \end{bmatrix} = \begin{bmatrix} \bar{\Sigma}_{xx} & \bar{\Sigma}_{xy} \\ \bar{\Sigma}_{yx} & \bar{\Sigma}_{yy} \end{bmatrix}$$

$$R \mu_z = \begin{bmatrix} \Delta u - A^T L b \\ L b \end{bmatrix} \rightarrow \mu_z = \bar{R}^{-1} \begin{bmatrix} \Delta u - A^T L b \\ L b \end{bmatrix}$$

$$\mu_z = \begin{bmatrix} \bar{\Lambda}' & \bar{\Lambda}' A^T \\ A \bar{\Lambda}' & L' + A \bar{\Lambda}' A^T \end{bmatrix} \begin{bmatrix} \Delta u - A^T L b \\ L b \end{bmatrix} = \begin{bmatrix} \bar{\Lambda}' (\Delta u - A^T L b) + \bar{\Lambda}' A^T L b \\ A \bar{\Lambda}' (\Delta u - A^T L b) + (L' + A \bar{\Lambda}' A^T) L b \end{bmatrix}$$

$$\mu_z = \begin{bmatrix} u - \bar{\Lambda}' A^T L b + \bar{\Lambda}' A^T L b \\ A \bar{\Lambda}' \Delta u - A \bar{\Lambda}' A^T L b + L' L b + A \bar{\Lambda}' A^T L b \end{bmatrix} = \begin{bmatrix} u \\ b + A u \end{bmatrix} = \begin{bmatrix} u_x \\ u_y \end{bmatrix}$$

$$\boxed{\begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} u \\ b + A u \end{bmatrix}};$$

$$\begin{aligned} p(x) &= N(x | u, \bar{\Lambda}') \\ p(y|x) &= N(y | Ax + b, \bar{\Sigma}') \\ p(y) &= N(y | u_y, \bar{\Sigma}_{yy}) = N(y | b + Au, L' + A \bar{\Lambda}' A^T) \end{aligned}$$

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

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

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

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

$$p(x_a|x_b) = N(x_a|u_{ab}, \Sigma_{ab})$$

$$u_{ab} = u_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - u_b); \quad u_{ba} = u_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - u_b)$$

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

Entonces, a partir de la matriz de precisión Λ :

$$R = \begin{bmatrix} \Lambda + A^T \Lambda A & -A^T \Lambda \\ -\Lambda A & \Lambda \end{bmatrix} = \begin{bmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{bmatrix}, \quad u_2 = \begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} u \\ b + Au \end{bmatrix}$$

$$\Sigma_{xy} = \Lambda_{xx}^{-1} = (\Lambda + A^T \Lambda A)^{-1}$$

$$u_{x|y} = u_x - \Lambda_{xx}^{-1}\Lambda_{xy}(y - u_y) = u + (\Lambda + A^T \Lambda A)^{-1}A^T \Lambda (y - Au - b)$$

$$\Sigma_{x_1y} = (\lambda + A^T L A)^{-1}$$

$$u_{x_1y} = u + (\lambda + A^T L A)^{-1} (A^T L \gamma - A^T L b - A^T L A u)$$

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

$$= (\lambda + A^T L A)^{-1} [A^T L (y - b) - \underline{A^T L A u} + \lambda u + \underline{A^T L A u}]$$

$$u_{x_1y} = (\lambda + A^T L A)^{-1} [A^T L (y - b) + \lambda u]$$

$$p(x_1|y) = N(x_1 | u_{x_1y}, \Sigma_{x_1y})$$

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

con $w \sim p(w) = N(w | m_0, S_0)$ $\rightarrow \eta \sim p(\eta) = N(\eta | 0, B^{-1})$

Demuestre que $p(w | t) = N(w | m_N, S_N)$

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

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

$$\phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)]^T \in \mathbb{R}^{N \times Q}, \quad t = [t_1, t_2, \dots, t_N]^T \in \mathbb{R}^N$$
$$w \in \mathbb{R}^Q \quad \beta \in \mathbb{R}^+$$

Demuestre que la predictiva toma la forma:

$$p(t^* | t, \phi, w) = \int p(t^* | w) p(w | t) dw$$

$$p(t^* | t, \phi, w) = N(t^* | m_N^T \phi(x^*), \sigma_N^2(x^*))$$

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

$$S_i: p(x) = N(x | \mu, \Sigma); \quad p(y|x) = N(y | Ax + b, L^{-1})$$

$$\text{con } y = f(x | A, b) = Ax + b.$$

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

$$\mu_{xy} = (\Lambda + A^T L A)^{-1} [A^T L (y - b) + \Lambda u]$$

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

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

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

$$\text{Entonces: } p(t_n | w^T \phi(x_n), \beta^{-1}) = N(t_n | w^T \phi(x_n), \beta^{-1}) \rightarrow p(t | w) = N(t | \phi w, \beta' I)$$

$$\text{Ahora: } w \sim p(w) = N(w | m_0, S_0) \rightarrow \mu = m_0; \quad \Lambda' = S_0$$

$$\text{POSTERIOR: } p(w | t) = N(w | \mu_{w|t}, \Sigma_{w|t}) \quad S_0^{-1} = \Lambda$$

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

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

$$M_N = \mu_{w|t} = S_N (S_0^{-1} m_0 + \beta \phi^T t)$$

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

$$\begin{matrix} \downarrow \\ A & X & L' \end{matrix}$$

$$b = 0$$

$$L' = \bar{\beta}' I$$

$$L = \beta I$$

Para la predictiva: $t_* = w^\top \phi(x_*) + \eta$; x_* : entrada nueva
 ↓ predictiva t_* : salida nueva → predicción

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

Dado que: $\eta = t_* - w^\top \phi(x_*) \sim N(t_* - w^\top \phi(x_*), 10, \beta^{-1})$

$$p(t^* | w) = N(t^* | w^\top \phi(x_*), \beta^{-1}) \rightarrow \text{verosimilitud para } t^*$$

$$p(w | t) = N(w | M_{w|t}, \Sigma_{w|t}) = N(w | M_N, S_N) \rightarrow \text{posterior}$$

$$p(t_* | t, \phi, w) = \int N(t_* | w^\top \phi(x_*), \beta^{-1}) N(w | M_N, S_N) dw.$$

NOTA: Dado que se cuenta con pdf Gaussianas:

$$\mathbb{E}_{w|t} \{ t_* \} = \mathbb{E}_{w|t} \{ w^\top \phi(x_*) + \eta \} = \mathbb{E}_{w|t} \{ w^\top \phi(x_*) \} + \mathbb{E}_{w|t} \{ \eta \} = (\mathbb{E}_{w|t} \{ w \})^\top \phi(x_*)$$

$$\boxed{\mathbb{E}_{w|t} \{ t_* \} = M_N^\top \phi(x_*) = M_{t^*|w}}$$

$$\mathbb{E}_{w|t} \{ (t_* - \mathbb{E}_{w|t} \{ t_* \ })^\top (t_* - \mathbb{E}_{w|t} \{ t_* \ }) \} = \Sigma_{t^*|w}$$

$$\Sigma_{t^*|w} = \mathbb{E}_{w|t} \{ (w^\top \phi(x_*) + \eta - M_N^\top \phi(x_*))^\top (w^\top \phi(x_*) + \eta - M_N^\top \phi(x_*)) \}$$

$$\begin{aligned}
\Sigma_{t^*|w} &= \sum_{w|t} \left\{ \left[(\omega^\top \phi(x_+))^\top + \eta - (M_N^\top \phi(x_+))^\top \right] [\omega^\top \phi(x_+) + \eta - M_N^\top \phi(x_+)] \right\} \\
&= \sum_{w|t} \{ [\phi^\top(x_+) \omega + \eta - \phi^\top(x_+) M_N] [\omega^\top \phi(x_+) + \eta - M_N^\top \phi(x_+)] \} \\
&= \sum_{w|t} \{ [\phi^\top(x_+) (\omega - M_N) + \eta] [\underbrace{(\omega^\top - M_N^\top)}_{\text{red}} \phi(x_+) + \eta] \} \\
&= \sum_{w|t} \{ \phi^\top(x_+) (\omega - M_N) \underbrace{(\omega - M_N)^\top}_{\text{red}} \phi(x_+) + \underbrace{\phi^\top(x_+) (\omega - M_N) \eta}_{\text{purple}} \\
&\quad + \underbrace{\eta (\omega - M_N)^\top \phi(x_+)}_{\text{purple}} + \eta \eta \} \\
&= \sum_{w|t} \{ \phi^\top(x_+) (\omega - M_N) (\omega - M_N)^\top \phi(x_+) \} + \sum_{w|t} \{ 2\eta (\omega - M_N)^\top \phi(x_+) \} \\
&\quad + \sum_{w|t} \{ \eta^2 \} \quad \xrightarrow{\text{green}} S_N: \text{covarianza posterior } p(w|t) \\
\Sigma_{t^*|w} &= \phi^\top(x_+) \underbrace{\sum_{w|t} \{ (\omega - M_N) (\omega - M_N)^\top \}}_{\text{green}} \phi(x_+) \\
&\quad + \underbrace{(\sum_{w|t} \{ \eta^2 \})}_{\text{purple}} \rightarrow \beta^{-1}: \text{varianza del ruido} \rightarrow N(\eta | 0, \beta^{-1}) \\
&\quad + 2 \underbrace{\sum_{w|t} \{ \eta (\omega - M_N)^\top \phi(x_+) \}}_{\text{orange}} \rightarrow \phi \quad m_\eta = 0
\end{aligned}$$

NOTA:

$$\begin{aligned}
\sum_{w|t} \{ \eta (\omega - M_N)^\top \phi(x_+) \} &= \sum_{w|t} \{ \eta (\omega^\top - M_N^\top) \phi(x_+) \} = \sum_{w|t} \{ \eta \omega^\top \phi(x_+) \} - \sum_{w|t} \{ \eta M_N^\top \phi(x_+) \} \\
&= \sum_{w|t} \{ \eta \omega^\top \phi(x_+) \} - \sum_{w|t} \{ \eta \} M_N^\top \phi(x_+)
\end{aligned}$$

Si a independiente de b : $p(a,b) = p(a)p(b)$

$$\Sigma_{a,b} \{a,b\} = \int a.b p(a,b) da db = \int a.b p(a)p(b) da db$$

$$\Sigma_{a,b} \{a,b\} = \int a p(a) da \int b p(b) db = \Sigma_a \{a\} \Sigma_b \{b\}$$

Entonces: $\mathbb{E}\{\eta w^T\} = \mathbb{E}\{\eta\} \mathbb{E}\{w^T\}$

$$\begin{aligned} \mathbb{E}_{w|t} \{\eta (w - M_N)^T \phi(x_*)\} &= \sum_{w|t} \{\eta\} \sum_{w|t} \{w^T\} \phi(x_*) - \sum_{w|t} \{\eta\} M_N^T \phi(x_*) \\ &= \sum_{w|t} \{\eta\} M_N^T \phi(x_*) - \sum_{w|t} \{\eta\} M_N^T \phi(x_*) = 0 \end{aligned}$$

Por lo tanto:

$$\Sigma_{t+1|w} = \phi^T(x_*) S_N \phi(x_*) + \beta^{-1} \in \mathbb{R}^+$$

$$M_{t+1|w} = M_N^T \phi(x_*) \in \mathbb{R}$$

$$p(t^+|t, w, \phi, x_*) = N(t_+ | M_{t+1|w}, \Sigma_{t+1|w})$$