$$\log(p(z)) = \log(p(y|x)p(x)) = \log(p(y|x)) + \log(p(x))$$
$$= \log\left[\frac{1}{(2\pi)^{D/2}|L^{-1}|^{1/2}}\exp\left[-\frac{1}{2}(y-Ax-b)^T L(y-Ax-b)\right]\right]$$
$$+ \log\left[\frac{1}{(2\pi)^{M/2}\Lambda}\exp\left[-\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)\right]\right])$$
$$= -\frac{1}{2}(y-Ax-b)^T L(y-Ax-b) - \frac{1}{2}(x-\mu)^T\Lambda(x-\mu) + cte.$$

Ejercicio: Determine el termino cte.

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

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

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

$$= \log\left(\frac{1}{(2\pi)^{D/2}|L^{-1}|^{1/2}} \cdot \exp\left(-\frac{1}{2}(y-Ax-b)^T L(y-Ax-b)\right)\right)$$
$$+ \log\left(\frac{1}{(2\pi)^{M/2}|\Lambda^{-1}|^{1/2}} \cdot \exp\left(\frac{-1}{2}(x-\mu)^T\Lambda(x-\mu)\right)\right)$$

$$= \log\left(\exp\left(-\frac{1}{2}(y-Ax-b)^T L(y-Ax-b)\right)\right)$$
$$+ \log\left(\exp\left(\frac{-1}{2}(x-\mu)^T\Lambda(x-\mu)\right)\right)$$
$$+ \log\left(\frac{1}{(2\pi)^{D/2}|L^{-1}|^{1/2}}\right) + \log\left(\frac{1}{(2\pi)^{M/2}|\Lambda^{-1}|^{1/2}}\right)$$

$$= -\frac{1}{2}(y-Ax-b)^T L(y-Ax-b) - \frac{1}{2}(x-\mu)^T\Lambda(x-\mu)$$
$$- \log\left((2\pi)^{D/2}|L^{-1}|^{1/2}(2\pi)^{M/2}|\Lambda^{-1}|^{1/2}\right)$$

$$= -\frac{1}{2}(y-Ax-b)^T L(y-Ax-b) - \frac{1}{2}(x-\mu)'\Lambda(x-\mu)$$
$$\underbrace{- \log\left((2\pi)^{(D+M)/2}|L^{-1}|^{1/2}|\Lambda^{-1}|^{1/2}\right)}_{\text{constante}}$$

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

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

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

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

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

Para esto vamos a tener en cuenta el teorema de Bayes para variables Gaussianas que nos dice

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

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

entonces

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

donde

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

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

- Primero, hallemos quién es $p(t \mid w^T \Phi, \beta^{-1} I)$

Como $\eta = t_n - w^T \Phi(x_n)$ entonces

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

Por lo tanto,

$$t_n \sim \mathcal{N}(t_n \mid \boldsymbol{\omega}^T \phi(x_n), \beta^{-1}) \rightarrow \mathbf{t} \sim \mathcal{N}(\mathbf{t} \mid \boldsymbol{\omega}^T \boldsymbol{\Phi}, \beta^{-1} \mathbf{I})$$

- Notamos que tenemos las dos hipótesis del teorema

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

$$p(\boldsymbol{\omega}) = \mathcal{N}(\boldsymbol{\omega} \mid \mathbf{M}_0, \mathbf{S}_0)$$

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

$$\text{y} \quad \mathbf{t} = \mathcal{N}(\mathbf{t} \mid \boldsymbol{\omega}^T \boldsymbol{\Phi}, \beta^{-1} \mathbf{I})$$

Lo que está en rojo es para ver el teorema en nuestro ejercicio

Así que lo podemos aplicar

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

$$p(\boldsymbol{\omega} \mid \mathbf{t}) = \mathcal{N}(\boldsymbol{\omega} \mid \mathbf{M}_N, \mathbf{S}_N)$$

donde

$$\boldsymbol{\Sigma}_{x|y} = (\boldsymbol{\Lambda} + \mathbf{A}^T \mathbf{L} \mathbf{A})^{-1} \qquad \text{(Varianza)}$$

$$\boxed{\mathbf{S}_N = \boldsymbol{\Sigma}_{\omega|t} = (\mathbf{S}_0 + \boldsymbol{\Phi}^T \beta \mathbf{I} \boldsymbol{\Phi})^{-1}}$$

$$\boldsymbol{\mu}_{x|y} = (\boldsymbol{\Lambda} + \mathbf{A}^T \mathbf{L} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{L} (\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda} \boldsymbol{\mu}) \qquad \text{(Media)}$$

$$\mathbf{M}_N = \boldsymbol{\mu}_{\omega|t} = (\mathbf{S}_0^{-1} + \boldsymbol{\Phi}^T \beta \mathbf{I} \boldsymbol{\Phi})^{-1} (\boldsymbol{\Phi}^T \beta \mathbf{I} (\mathbf{t} - 0) + \mathbf{S}_0^{-1} \mathbf{M}_0)$$

$$\mathbf{M}_N = \boldsymbol{\mu}_{\omega|t} = (\mathbf{S}_0^{-1} + \boldsymbol{\Phi}^T \beta \mathbf{I} \boldsymbol{\Phi})^{-1} (\beta \boldsymbol{\Phi}^T \mathbf{t} + \mathbf{S}_0^{-1} \mathbf{M}_0)$$

$$\boxed{\mathbf{M}_N = \boldsymbol{\mu}_{\omega|t} = \mathbf{S}_N (\beta \boldsymbol{\Phi}^T \mathbf{t} + \mathbf{S}_0^{-1} \mathbf{M}_0)}$$

Ejercicio :

Demuestre que la predictiva toma la forma:

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

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

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

Vamos a analizar este ejercicio desde la marginalización de la predictiva aplicando un resultado del procedimiento del teorema de Bayes para variables Gaussianas

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

entonces

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

En nuestro ejercicio tenemos que

$$\eta = t^* - \mathbf{w}^T \phi(x^*) \sim \mathcal{N}(t^* - \mathbf{w}^T \phi(x^*) | 0, \beta^{-1})$$

$$\Rightarrow p(t^* | \mathbf{w}) = \mathcal{N}(t^* - \mathbf{w}^T \phi(x^*) | 0, \beta^{-1})$$

$$= \mathcal{N}(t^* | \phi(x^*) \mathbf{w}, \beta^{-1})$$

Además, del ejercicio anterior sabemos que

$$p(\mathbf{w} | \mathbf{t}) = \mathcal{N}(\mathbf{w} | \boldsymbol{\mu}_{M|t}, \boldsymbol{\Sigma}_{M|t}) = \mathcal{N}(\mathbf{w} | M_N, S_N)$$

Por lo tanto,

$$\boxed{p(t^* \mid \mathbf{t}, \boldsymbol{\omega}) = \mathcal{N}\left(t^* \mid \phi(x^*)\mathbf{M}_N, \beta^{-1} + \phi(x^*)^T \mathbf{S}_N \phi(x^*)\right)}$$