INFERENCIA VARIACIONAL

→ Necesito conocer $p(z|x)$ (no podemos calcularla)

→ La aproximo por una distribución $q(z)$

(defino $q(z)$ como una familia de distribuciones tratables)

→ Para ajustar $q(z)$ MINIMIZAMOS

$$KL(q\,||\,p) = \int q(z) \log \frac{q(z)}{p(z|x)}\, dz$$

Esta KL es equivalente a :

$$KL(q\,||\,p) = \int q(z) \log \frac{q(z)}{p(x,z)}\, dz + \int q(z) \log p(x) dz$$

$$= + \int q(z) \log \frac{q(z)}{p(x,z)}\, dz + \log p(x)$$

Defino $\quad L(q) = -\int q(z) \log \dfrac{q(z)}{p(x,z)}\, dz$

Tenemos

$$KL(q\,||\,p) = -L(q) + \log p(x)$$

$$L(q) = -KL(q\,||\,p) + \log p(x) \leq \log p(x)$$

↳ $L(q)$ es una COTA INFERIOR DE $\log p(x)$

↳ MAXIMIZAR LA COTA ⇒ MINIMIZAR $KL(q\,||\,p)$

$L(q)$ es MÁXIMA SI $p = q$

THE MEAN FIELD METHOD

↳ Asumimos que $q(z)$ se factoriza sobre todas las variables

$$q(z) = \prod_i q(z_i) = \prod_i q_i$$

↳ Maximizamos $L(q)$ respecto a cada factor:

$$L(q_j) = \int q(z) \log \frac{p(x,z)}{q(z)} \, dz =$$

$$= \int \prod_i q_i \left[\log p(x,z) - \sum_i \log q_i \right] dz =$$

$$= \int q_j \left(\prod_{i \neq j} q_i \right) \log p(x,z) \, dz$$

$$- \int q_j \left(\prod_{i \neq j} q_i \right) \log q_j \, dz - \int q_j \left(\prod_{i \neq j} q_i \right) \left(\sum_{i \neq j} \log q_i \right) dz =$$

$$= \int q_j \left[\int \prod_{i \neq j} q_i \log p(x,z) \, dz_i \right] dz_j$$

$$- \int q_j \log q_j \, dz_j + cte$$

$$= \int q_j \log f_j \, dz_j - \int q_j \log q_j \, dz_j + cte$$

$$= -KL(q_j \| f_j)$$

donde $\log f_j = \mathbb{E}_{-q_j} \left[\log p(x,z) \right] + cte$

→ PARA MAXIMIZAR $L(q_j)$ => MINIMIZAR $KL(q_j \| f_j)$

$$\log q_j = \log q(z_j) = \mathbb{E}_{-q_j} \left[\log p(x,z) \right] + cte$$

$$\ln q^*(W^{(m)}) = E_{Y,\alpha^{1,2},\tau^{1,2}}\left[\ln p(X^{1,2}, W^{1,2}, Y, \alpha^{1,2}, \tau^{1,2})\right] + cte$$

$$= E_{Y,\tau^{(m)}}\left[\ln p(X^{(m)} \mid W^{(m)}, Y, \tau^{(m)})\right] + E_{\alpha^{(m)}}\left[\ln p(W^{(m)} \mid \alpha^{(m)})\right]$$

$$+ cte$$

• TÉRMINO 1 $\quad x_n^{(m)} \sim \mathcal{N}(W^{(m)} y_n, \tau_m^{-1} I)$

$$\ln p(X^{(m)} \mid W^{(m)}, Y, \tau^{(m)}) = \sum_n \ln \mathcal{N}(W^{(m)} y_n, \tau_m^{-1} I)$$

$$= \sum_n -\frac{\tau^{(m)}}{2}(\underline{x}_n^{(m)} - W^{(m)} y_n)^T(\underline{x}_n^{(m)} - W^{(m)} y_n) + cte$$

$$= -\frac{\tau^{(m)}}{2}\sum_n (\underline{x}_n^{(m)T}\underline{x}_n^{(m)} - 2\underline{x}_n^{(m)T} W^{(m)} y_n + y_n^T W^{(m)T} W^{(m)} y_n)$$

$$= \sum_n \left(\tau^{(m)} \underline{x}_n^{(m)T} W^{(m)} y_n - \frac{\tau^{(m)}}{2} \underset{(1\times K_c)}{y_n^T} \underset{(K_c\times D_m)}{W^{(m)T}} \underset{(D_m\times K_c)}{W^{(m)}} \underset{(K_c\times 1)}{y_n}\right) + cte$$

$$= \sum_n \left(\tau^{(m)} \underline{x}_n^{(m)T} W^{(m)} y_n - \frac{\tau^{(m)}}{2}\sum_{d=1}^{D_m} \underset{(1\times K_c)}{y_n^T} \underset{(K_c\times 1)}{W_{d:}^T} \underset{(1\times K_c)}{W_{d:}} \underset{(K_c\times 1)}{y_n}\right) + cte$$

$$= \sum_n \left(\tau^{(m)} \underline{x}_n^{(m)T} W^{(m)} y_n - \frac{\tau^{(m)}}{2}\sum_{d=1}^{D_m} \underset{(1\times K_c)}{W_{d:}} \underset{(K_c\times 1)}{y_n} \underset{(1\times K_c)}{y_n^T} \underset{(K_c\times 1)}{W_{d:}^T}\right) + cte$$

$$E_{Y,\tau^{(m)}}\left[\ln p(X^{(m)} \mid X^{(m)}, Y, \tau^{(m)})\right] =$$

$$= \sum_n \langle \tau^{(m)} \rangle \underline{x}_n W^{(m)} \langle y_n \rangle - \frac{\langle \tau^{(m)} \rangle}{2}\sum_{d=1}^{D_m} \underline{W}_{d:} \langle y_n, y_n^T \rangle \underline{W}_{d:}^T$$

$$+ cte$$

• TERMINO 2 $\quad W_{:k}^{(m)} \sim \mathcal{N}(0, \alpha_k^{(m)^{-1}} I)$

$$\ln p(w^{(m)} \mid \alpha^{(m)}) = \sum_{d=1}^{D_m}\sum_{k=1}^{K_c} \ln p(w_{dk}^{(m)} \mid \alpha_k^{(m)}) =$$

$$= \sum_{d=1}^{D_m}\sum_{k=1}^{K_c} + \frac{1}{2}\ln \alpha_k^{(m)} - \frac{\alpha_k^{(m)}}{2} w_{dk}^{2} + cte$$

$$\mathbb{E}_{\alpha^{(m)}}\left[\ln p(w^{(m)} \mid \alpha^{(m)})\right] = \sum_{d=1}^{D_m}\sum_{k=1}^{K_c}\left(\cancelto{cte}{\frac{1}{2}\mathbb{E}\left[\ln \alpha_k^{(m)}\right]} - \frac{\langle \alpha_k^{(m)}\rangle}{2} w_{dk}^{2}\right) + cte =$$

$$= \sum_{d=1}^{D_m}\sum_{k=1}^{K_c} - \frac{\langle \alpha_k^{(m)}\rangle}{2} w_{dk}^{2} + cte$$

Luego:

$$\ln q^{*}(w^{(m)}) = \sum_{d=1}^{D_m}\sum_{k=1}^{K_c} - \frac{\langle \alpha_k^{(m)}\rangle}{2} w_{dk}^{2} +$$

$$+ \sum_{n}\left(\langle \psi^{(m)}\rangle \underline{x}_n W^{(m)} \langle y_n\rangle - \frac{\langle \psi^{(m)}\rangle}{2}\sum_{d=1}^{D_m} \underline{W}_{d:} \langle y_n, y_n^T\rangle \underline{W}_{d:}^T\right) + cte$$

$$= \sum_{d=1}^{D_m}\Bigg(-\frac{1}{2}\underset{(1\times K_c)}{\underline{W}_{d:}^{(m)}}\ \underset{(K_c\times K_c)}{diag(\langle \alpha^{(m)}\rangle)}\ \underset{(K_c\times 1)}{\underline{W}_{d:}^{(m)T}} +$$

$$+ \sum_{n=1}^{N}\left(\langle \psi^{(m)}\rangle x_{nd}\, \underline{W}_{d:}^{(m)} \langle y_n\rangle - \frac{1}{2}\langle \psi^{(m)}\rangle \underline{W}_{d:}^{(m)} \langle y_n, y_n^T\rangle \underline{W}_{d:}^T\right)\Bigg) + cte$$

$$= \sum_{d=1}^{D_m} -\frac{1}{2}\underline{W}_{d:}^{(m)}\left[diag(\langle \alpha^{(m)}\rangle) + \langle \psi^{(m)}\rangle \Big\langle \underset{(K_c\times N)(N\times K_c)}{\overset{Y,\ Y^T}{\cancel{y_n, y_n^T}}}\Big\rangle\right]\underline{W}_{d:}^T$$

$$+ \langle \psi^{(m)}\rangle\ \underset{(1\times K_c)}{W_{d:}^{(m)}}\ \underset{(K_c\times N)}{\langle Y\rangle}\ \underset{(N\times 1)}{X_{:d}^{(m)T}}$$

$$\ln q^*(W^{(m)}) = \sum_{d=1}^{D_m} \mathcal{N}\Big(\underset{(1\times K_c)}{\underline{W}_{d:}^{(m)}} \,\Big|\, \underset{(1\times K_c)}{\mu_{W_{d:}}^{(m)}},\ \underset{(K_c\times K_c)}{\Sigma_{W^{(m)}}}\Big)$$

$$\Sigma_{W^{(m)}}^{-1} = \mathrm{diag}\left(\langle \alpha^{(m)} \rangle\right) + \langle \tau^{(m)} \rangle \langle Y, Y^T \rangle$$

$$\mu_{W_{d:}}^{(m)} = \left(\langle \tau^{(m)} \rangle\, \Sigma_{W^{(m)}} \langle Y \rangle\, X_{:d}^{(m)\,T}\right)^T$$

$$\underset{D\times K_c}{\langle W^{(m)} \rangle} = \underset{(D\times N)}{X^{(m)}}\ \underset{(N\times K_c)}{\langle Y \rangle^T}\ \underset{(K_c\times K_c)}{\Sigma_{W^{(m)}}}\ \langle \tau^{(m)} \rangle$$

$$\ln q^*(Y) = \mathbb{E}[\ln p(X^{(1)} \mid W^{(1)}, Y, \tau^{(1)})] +$$

$$+ \mathbb{E}[\ln p(X^{(2)} \mid W^{(2)}, Y, \tau^{(2)})] +$$

$$+ \mathbb{E}[\ln p(Y)] + cte$$

- TÉRMINOS 1 y 2

$$\ln p(X^{(m)} \mid W^{(m)}, Y, \tau^{(m)}) = \sum_n \ln \mathcal{N}(W^{(m)} y, \tau^{(m)^{-1}} I) =$$

$$= \sum_n \left(\tau^{(m)} \underline{x}_n^{(m)T} W^{(m)} y_n - \frac{\tau^{(m)}}{2} y_n^T W^{(m)T} W^{(m)} y_n \right) + cte$$

$$\mathbb{E}_{W^{(m)}, \tau^{(m)}} [\ln p(X^{(m)} \mid W^{(m)}, Y, \tau^{(m)})] =$$

$$= \sum_n \left(\langle \tau^{(m)} \rangle \underline{x}_n^{(m)T} \langle W^{(m)} \rangle y_n - \frac{\langle \tau^{(m)} \rangle}{2} y_n^T \langle W^{(m)T}, W^{(m)} \rangle y_n \right)$$

$$+ cte$$

- TERMINO 3 $\quad y_n \sim \mathcal{N}(\underline{0}, I)$

$$\ln p(Y) = \sum_n \mathcal{N}(0, I) = -\frac{1}{2} y_n^T y_n + cte$$

$$\mathbb{E}[\ln p(Y)] = \ln p(Y) = \sum_n -\frac{1}{2} y_n^T y_n + cte$$

- T1 - T2 - T3

$$\ln q^*(Y) = \sum_n \Big(-\frac{1}{2} y_n^T y_n +$$

$$+ \sum_{m=1,2} \left(\langle \tau^{(m)} \rangle x_n^{(m)T} \langle W^{(m)} \rangle y_n - \frac{\langle \tau^{(m)} \rangle}{2} y_n^T \langle W^{(m)T}, W^{(m)} \rangle y_n \right) \Big)$$

$$+ cte$$

$$\ln q^*(Y) = \prod_{n=1}^{N} \mathcal{N}(y_n \mid \underset{K_c \times 1}{\mu_{y_n}}, \underset{(K_c \times K_c)}{\Sigma_y})$$

$$\Sigma_y^{-1} = I + \sum_{m=1,2} \langle \tau^{(m)} \rangle \langle W^{(m)T}, W^{(m)} \rangle$$

$$\mu_{y_n}^T = \sum_{m=1,2} \langle \tau^{(m)} \rangle \; x_n^{(m)T} \langle W^{(m)} \rangle \Sigma_y$$

$$\langle Y \rangle = \sum \langle \tau^{(m)} \rangle \Sigma_y \langle W^{(m)} \rangle X^{(m)}$$

$$\ln q^*(\alpha^{(m)}) = \mathbb{E}_{W^{(m)}}[\ln p(W^{(m)}|\alpha^{(m)})] + \mathbb{E}[\ln p(\alpha^{(m)})]$$
$+ cte$

- TERMINO 1 $\quad W_{:k}^{(m)} \sim N(0, (\alpha_k^{(m)})^{-1} I)$

$\ln p(W^{(m)}|\alpha^{(m)}) =$

$$= \sum_{d=1}^{D_m} \sum_{k=1}^{K_c} \left(+ \frac{1}{2} \ln \alpha_k^{(m)} - \frac{1}{2} \alpha_k^{(m)} w_{dk}^{(m)2} \right) + cte =$$

$$= \sum_{k=1}^{K_c} \left(+ \frac{D_m}{2} \ln \alpha_k^{(m)} - \frac{1}{2} \alpha_k^{(m)} \underset{(1 \times D_m)}{\underline{W}_{k:}^{(m)T}} \cdot \underset{(D_m \times 1)}{\underline{W}_{k:}^{(m)}} \right) + cte$$

$\mathbb{E}_{W^{(m)}}[\ln p(W^{(m)}|\alpha^{(m)})] =$

$$= \sum_{k=1}^{K_c} \left(\frac{D_m}{2} \ln \alpha_k^{(m)} - \frac{1}{2} \alpha_k^{(m)} \langle \underline{W}_{k:}^{(m)T}, \underline{W}_{k:}^{(m)} \rangle \right) + cte$$

- TERMINO 2 $\quad \alpha_k^{(m)} \sim \text{Gamma}(\alpha_0, \beta_0)$

$$\mathbb{E}[\ln p(\alpha^{(m)})] = \ln p(\alpha^{(m)}) = \sum_{k=1}^{K_c} \ln p(\alpha_k^{(m)}) =$$

$$= \sum_{k=1}^{K_c} \beta_0 \alpha_k^{(m)} + (\alpha_0 - 1) \ln \alpha_k^{(m)} + cte$$

- T1 + T2

$$\ln q^*(\alpha^{(m)}) = \sum_{k=1}^{K_c} \frac{D_m}{2} \ln \alpha_k^{(m)} - \frac{1}{2} \alpha_k^{(m)} \langle \underline{W}_{k:}^{(m)T}, \underline{W}_{k:} \rangle +$$
$$+ \beta_0 \alpha_k^{(m)} + (\alpha_0 - 1) \ln \alpha_k^{(m)} + cte$$

$$q^*(\alpha^{(m)}) = \prod_{k=1}^{K_c} \text{Gamma}(\alpha_k^{(m)} | a_{\alpha_k^{(m)}}, b_{\alpha_k^{(m)}})$$

$$a_{\alpha_k^{(m)}} = \frac{D_m}{2} + \alpha_0$$

$$b_{\alpha_k^{(m)}} = \beta_0 - \frac{1}{2} \langle \underline{W}_{k:}^{(m)T}, W_{k:}^{(m)} \rangle =$$

$$= \beta_0 - \frac{1}{2} \langle W^{(m)T}, W^{(m)} \rangle_{k,k}$$

$$\langle W_{k:}^{(m)T}, W_{k:}^{(m)} \rangle = \sum_{d=1}^{D_m} \langle W_{kd}^2 \rangle = \langle W^T, W \rangle_{k,k}$$
$(K_c \times D_m)\ (D_m \ldots)$

$$\ln q^*(\tau^{(m)}) = \mathbb{E}[\ln p(X^{(m)} | W^{(m)}, Y, \tau^{(m)})] + \mathbb{E}[\ln p(\tau^{(m)})] + cte$$

* TÉRMINO 1 $\quad \underline{x}_n^{(m)} \sim N(W^{(m)} y_n, \tau^{(m)-1} I)$

$$\ln p(X^{(m)} | W^{(m)}, Y, \tau^{(m)}) = \sum_{n=1}^{N} \sum_{d=1}^{D_m} \left(\frac{1}{2} \ln \tau^{(m)} - \right.$$

$$\left. - \frac{1}{2} \tau^{(m)} (x_{dn}^{(m)} - \underline{W}_{d:}^{(m)} y_n)^2 \right) + cte =$$

$$= \frac{N D_m}{2} \ln \tau^{(m)} - \frac{1}{2} \tau^{(m)} \sum_n \sum_d \left(x_{dn}^{(m)2} - 2 x_{dn}^{(m)} \underline{W}_{d:} y_n + \right.$$

$$\left. + (W_{d:}^{(m)} y_n)^2 \right) + cte =$$

$$= \frac{N D_m}{2} \ln \tau^{(m)} - \frac{1}{2} \tau^{(m)} \Big(\sum_n \sum_d x_{dn}^{(m)2} - 2 \sum_d \underset{(1 \times K_c)(K_c \times N)(N \times 1)}{W_{d:} Y X_{d:}^{(m)}} +$$

$$+ \sum_d \underset{(1 \times K_c)(K_c \times N)(N \times K_c)(K_c \times 1)}{W_{d:}^{(m)} Y Y^T W_{d:}^{(m)T}} \Big) + cte =$$

$$= \frac{N D_m}{2} \ln \tau^{(m)} - \frac{1}{2} \tau^{(m)} \Big(\sum_n \sum_d x_{dn}^{(m)2} - 2 \, Tr \{ \underset{(D_m \times K_c)(K_c \times N)(N \times D_m)}{W^{(m)} Y X^{(m)}} \}$$

$$+ Tr \{ W^{(m)} Y Y^T W^{(m)T} \} \Big) + cte =$$

$$\mathbb{E}_{W^{(m)}, Y}[\ln p(X^{(m)} | W^{(m)}, Y, \tau^{(m)})] = \frac{N D_m}{2} \ln \tau^{(m)} -$$

$$- \frac{1}{2} \tau^{(m)} \Big(\sum_n \sum_d x_{dn}^{(m)2} - 2 \, Tr \{ \langle W^{(m)} \rangle \langle Y \rangle X^{(m)} \} +$$

$$+ Tr \{ \langle W^{(m)T}, W^{(m)} \rangle \cdot \langle Y, Y^T \rangle \} \Big)$$

* TÉRMINO 2 $\quad \tau^{(m)} \sim Gamma(\alpha_0^{\tau}, \beta_0^{\tau})$

$$\ln p(\tau^{(m)}) = \beta_0^{\tau} \tau^{(m)} + (1 - \alpha_0^{\tau}) \ln \tau^{(m)} + cte$$

$$\mathbb{E}[\ln p(\tau^{(m)})] = \beta_0^{\tau} \tau^{(m)} + (1 - \alpha_0^{\tau}) \ln \tau^{(m)} + cte$$

* T1 + T2

$$\ln q^*(\tau^{(m)}) = \frac{N D_m}{2} \ln \tau^{(m)} - \frac{1}{2} \tau^{(m)} \Big(\sum_n \sum_d x_{dn}^{(m)2}$$

$$- 2 \, Tr \{ \langle W^{(m)} \rangle \langle Y \rangle X^{(m)} \} + Tr \{ \langle W^{(m)T}, W^{(m)} \rangle \cdot \langle Y, Y^T \rangle \} \Big)$$

$$+ \beta_0^{\tau} \tau^{(m)} + (1 - \alpha_0^{\tau}) \ln \tau^{(m)} + cte$$

$$q^*(\tau^{(m)}) = \mathrm{Gamma}(\tau^{(m)} \mid a_{\tau^{(m)}}, b_{\tau^{(m)}})$$

$$a_{\tau^{(m)}} = \alpha_0^{\tau} + \frac{ND_m}{2}$$

$$b_{\tau^{(m)}} = \beta_0^{\tau} - \frac{1}{2}\Big[\sum_n \sum_d x_{dn}^{(m)2} - 2\,\mathrm{Tr}\{\langle W^{(m)}\rangle \langle Y\rangle X^{(m)T}\}$$

$$+ \mathrm{Tr}\{\langle W^{(m)T} W^{(m)}\rangle \langle Y Y^T\rangle\Big]$$

CÁLCULO DE LA COTA

• $$\mathcal{L}(q) = -\int q(Z)\ \ln \frac{q(Z)}{p(X,Z)}\ dZ = \mathbb{E}_q[\ln p(X,Z)] - \mathbb{E}_q[\ln q(Z)]$$

• $$\mathbb{E}_q[\ln p(X,Z)] = \mathbb{E}_q[\ln p(Y)] + \sum_{m=1,2}\Big(\mathbb{E}[\ln p(W^{(m)}|\alpha^{(m)})]$$
$$+\ \mathbb{E}[\ln p(\alpha^{(m)})] + \mathbb{E}[\ln p(X^{(m)}|W^{(m)}, Y, \tau^{(m)})] +$$
$$+\ \mathbb{E}[\ln p(\tau^{(m)})]\Big)$$

TÉRMINOS DE ENTROPÍA

• $$\mathbb{E}_q[\ln q(Z)] = \mathbb{E}_q[\ln q(Y)] + \sum_{m=1,2}\Big(\mathbb{E}[\ln q(W^{(m)})] +$$
$$+\ \mathbb{E}[\ln q(\alpha^{(m)})] + \mathbb{E}[\ln q(\tau^{(m)})]\Big)$$

. . . . .

$$\mathcal{L}(q) = -\frac{NK_c}{2}\ln(2\pi) - \frac{1}{2}\mathrm{Tr}\{\langle YY^T\rangle\} +$$
$$+\sum_{m=1,2}\Big(-\frac{K_c D_m}{2}\ln(2\pi) + K_c(\alpha_0 \ln\beta_0 - \ln\Gamma(\alpha_0)) + \sum_{k=1}^{K_c} a_{\alpha_k}^{(m)} +$$
$$+\left(\frac{D_m}{2} + \alpha_0 - 1\right)\sum_{k=1}^{K_c}\left[\psi(a_{\alpha_k}^{(m)}) - \ln b_{\alpha_k}^{(m)}\right] - 2\beta_0\sum_{k=1}^{K_c}\frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}}\Big) +$$
$$+\sum_{m=1,2}\Big(-\frac{ND_m}{2}\ln(2\pi) + a_\tau^{(m)} + \alpha_0^\tau \ln\beta_0^\tau - \ln\Gamma(\alpha_0^\tau) +$$
$$+\left(\frac{ND_m}{2} + \alpha_0^\tau - 1\right)\left[\psi(a_\tau^{(m)}) - \ln b_\tau^{(m)}\right] - 2\beta_0^\tau\frac{a_\tau^{(m)}}{b_\tau^{(m)}}$$
$$-\left(\frac{NK_c}{2}\ln(2\pi e) + \frac{N}{2}\ln|\Sigma_Y|\right) -$$
$$-\sum_{m=1,2}\left(\frac{D_m K_c}{2}\ln(2\pi e) + \frac{D_m}{2}\ln|\Sigma_{W^{(m)}}|\right)$$
$$-\sum_{m=1,2}\sum_{k=1}^{K_c}\left(a_{\alpha_k}^{(m)} - \ln b_{\alpha_k}^{(m)} + \ln[\Gamma(a_{\alpha_k}^{(m)})] - (1-a_{\alpha_k}^{(m)})\psi(a_{\alpha_k}^{(m)})\right)$$
$$-\sum\left(a_\tau^{(m)} + \ln[\Gamma(a_\tau^{(m)})] - (1-a_\tau^{(m)})\psi(a_\tau^{(m)}) - \ln b_\tau^{(m)}\right)$$

Podemos simplificar $L(q)$, eliminando todos los términos que son constantes (no varían de iteración en iteración)

$\hookrightarrow$ $a_{\alpha_k}^{(m)}$ y $a_\psi^{(m)}$ son constantes

$$L(q) = -\frac{1}{2} Tr\{\langle Y Y^T \rangle\} + \sum_{m=1,2} \left(\left(\frac{D_m}{2} + \alpha_0 - 1 \right) \sum_{k=1}^{K_c} \left(-\ln b_{\alpha_k}^{(m)} \right) - 2 \beta_0 \sum_{k=1}^{K_c} \frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}} \right) + \sum_{m=1,2} \left(\left(\frac{N D_m}{2} + \alpha_0^\psi - 1 \right) \left(-\ln b_\psi^{(m)} \right) \right.$$

$$\left. - 2 \beta_0^\psi \frac{a_\psi^{(m)}}{b_\psi^{(m)}} \right) - \frac{N}{2} \ln |\Sigma_Y| - \sum_{m=1,2} \frac{D_m}{2} \ln |\Sigma_{W^{(m)}}|$$

$$+ \sum_{m=1,2} \sum_{k=1}^{K_c} \ln b_{\alpha_k}^{(m)} + \sum_{m=1,2} \ln b_\psi^{(m)}$$

TÉRMINOS DE ENTROPÍA

SOME REMARKS

- Entropía Gauss $\rightarrow \frac{1}{2} \ln (2\pi e)^d |\Sigma| = \frac{d}{2} \ln(2\pi e) + \frac{1}{2}\ln|\Sigma|$
- Entropía Gamma $\rightarrow a_0 - \ln \beta_0 + \ln [\Gamma(a_0)] - (1-a_0)\psi(a_0)$

$$\mathbb{E}_q [\ln q(Y)] \underset{\uparrow}{=} \sum_n \left(\frac{K_c}{2} \ln(2\pi e) + \frac{1}{2} \ln |\Sigma_y| \right) = \frac{N K_c}{2} \ln(2\pi e) + \frac{N}{2} \ln |\Sigma_y|$$

$$q(Y) = \prod_{n=1}^{N} N(y_n \mid \mu_{y_n}, \Sigma_y)$$

$$\mathbb{E}_q [\ln q(W^{(m)})] \underset{\uparrow}{=} \sum_d \left(\frac{K_c}{2} \ln(2\pi e) + \frac{1}{2} \ln |\Sigma_{W^{(m)}}| \right) =$$

$$q(W^{(m)}) = \prod_{d=1}^{D_m} N(W_{d:}^{(m)} \mid \mu_{W_{d:}^{(m)}}, \Sigma_{W^{(m)}})$$

$$= \frac{D_m K_c}{2} \ln(2\pi e) + \frac{D_m}{2} \ln |\Sigma_{W^{(m)}}|$$

$$\mathbb{E}_q [\ln q(\alpha^{(m)})] \underset{\uparrow}{=} \sum_{k=1}^{K_c} \left(a_{\alpha_k}^{(m)} - \ln b_{\alpha_k}^{(m)} + \ln[\Gamma(a_{\alpha_k}^{(m)})] - (1-a_{\alpha_k}^{(m)})\psi(a_{\alpha_k}^{(m)}) \right)$$

$$q(\alpha^{(m)}) = \prod_{k=1}^{K_c} \mathrm{Gamma}(a_{\alpha_k}^{(m)}, b_{\alpha_k}^{(m)})$$

$$= K_c \left(a_{\alpha_k}^{(m)} + \ln [\Gamma(a_{\alpha_k}^{(m)})] - (1-a_{\alpha_k}^{(m)}) \psi(a_{\alpha_k}^{(m)}) \right) - \sum_{k=1}^{K_c} \ln b_{\alpha_k}^{(m)}$$

$$\mathbb{E}_q [\ln q(\tau^{(m)})] \underset{\uparrow}{=} a_\tau^{(m)} + \ln [\Gamma(a_\tau^{(m)})] - (1-a_\tau^{(m)}) \psi(a_\tau^{(m)}) - \ln b_\tau^{(m)}$$

$$q(\tau^{(m)}) = \mathrm{Gamma}(a_\tau^{(m)}, b_\tau^{(m)})$$

TÉRMINOS ASOCIADOS A $\mathbb{E}_q[p(X,Z)]$

- $\mathbb{E}[\ln p(Y)] = \sum_n \mathbb{E}\left[-\frac{K_c}{2}\ln(2\pi) - \frac{1}{2} y_n^T y_n\right]$

$$\underset{(K_c \times 1)}{y_n} \sim N(0, I)$$

$$= -\frac{N K_c}{2}\ln(2\pi) - \frac{1}{2} Tr\{\langle Y Y^T \rangle\}$$

$$(D_m \times 1)\ W_{:k}^{(m)} \sim N(0, (\alpha_k^{(m)})^{-1} I)$$

- $\mathbb{E}[\ln p(W^{(m)} | \alpha^{(m)})] =$

$$= \sum_{k=1}^{K_c} \mathbb{E}\left[-\frac{D_m}{2}\ln(2\pi) + \frac{D_m}{2}\ln \alpha_k^{(m)} - \frac{1}{2}\alpha_k^{(m)} W_{:k}^{(m)T} W_{:k}^{(m)}\right] =$$

$$= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_k \mathbb{E}[\ln \alpha_k^{(m)}] - \frac{1}{2}\sum_k \mathbb{E}[\alpha_k^{(m)}]\,\mathbb{E}[W_{:k}^{(m)T} W_{:k}^{(m)}] =$$

$$E_{q(\alpha)}[\ln \alpha_k^{(m)}] = -\ln(b_{\alpha_k}^{(m)}) + \psi(a_{\alpha_k}^{(m)})$$

$$E_{q(\alpha)}[\alpha_k^{(m)}] = a_{\alpha_k}^{(m)} / b_{\alpha_k}^{(m)}$$

$$E_{q(W)}[W_{:k}^{(m)T} W_{:k}^{(m)}] = \langle W_{:k}^T, W_{:k} \rangle = \langle W^{(m)T}, W^{(m)} \rangle_{k,k}$$

De $q(\alpha^{(m)}) \rightarrow b_{\alpha_k}^{(m)} = \beta_0 - \frac{1}{2}\langle W^{(m)T}, W^{(m)} \rangle_{k,k}$

$$= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_k \left(\psi(a_{\alpha_k}^{(m)}) - \ln(b_{\alpha_k}^{(m)})\right) +$$

$$+ \sum_k \frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}}\left(b_{\alpha_k}^{(m)} - \beta_0\right) =$$

$$= -\frac{K_c D_m}{2}\ln(2\pi) + \frac{D_m}{2}\sum_k \left(\psi(a_{\alpha_k}^{(m)}) - \ln(b_{\alpha_k}^{(m)})\right)$$

$$+ \sum_k a_{\alpha_k}^{(m)} - \beta_0 \sum_k \frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}}$$

$$\circ \ \mathbb{E}\left[\ln p(\alpha^{(m)})\right] = \sum_{k=1}^{K_c} \mathbb{E}\left[\ln\left(\mathrm{Gamma}(\alpha_0, \beta_0)\right)\right] =$$

$$\alpha_k^{(m)} \sim \mathrm{Gamma}(\alpha_0, \beta_0)$$

$$\ln p(\alpha_k^{(m)}) = \ln \beta_0 - \beta_0 \alpha_k^{(m)} + (\alpha_0 - 1) \ln \beta_0 + (\alpha_0 - 1) \ln \alpha_k^{(m)} - \ln \Gamma(\alpha_0)$$

$$= \sum_{k=1}^{K_c} \alpha_0 \ln \beta_0 - \beta_0 \mathbb{E}_q\left[\alpha_k^{(m)}\right] + (\alpha_0 - 1) \mathbb{E}_q\left[\ln \alpha_k^{(m)}\right] - \ln \Gamma(\alpha_0) =$$

$$= K_c\left(\alpha_0 \ln \beta_0 - \ln \Gamma(\alpha_0)\right) + \sum_{k=1}^{K_c}\left(-\beta_0 \frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}} + (\alpha_0 - 1)\left(\psi(a_{\alpha_k}^{(m)}) - \ln b_{\alpha_k}^{(m)}\right)\right)$$

→ Combinamos $\mathbb{E}\left[\ln p(w^{(m)} | \alpha^{(m)})\right]$ y $\mathbb{E}\left[\ln p(\alpha^{(m)})\right]$

$$\mathbb{E}\left[\ln p(w^{(m)}, \alpha^{(m)})\right] = -\frac{K_c D_m}{2} \ln(2\pi) +$$

$$+ K_c\left(\alpha_0 \ln \beta_0 - \ln \Gamma(\alpha_0)\right) + \sum_{k=1}^{K_c} a_{\alpha_k}^{(m)} +$$

$$+ \left(\frac{D_m}{2} + \alpha_0 - 1\right) \sum_{k=1}^{K_c}\left[\psi(a_{\alpha_k}^{(m)}) - \ln b_{\alpha_k}^{(m)}\right] +$$

$$- 2 \beta_0 \sum_{k=1}^{K_c} \frac{a_{\alpha_k}^{(m)}}{b_{\alpha_k}^{(m)}}$$

$(D_m \times 1) \quad \underline{x}_n^{(m)} \sim N(W^{(m)} \underline{y}_n, \ \psi_m^{-1} I)$

- $$\mathbb{E}[\ln p(X^{(m)} \mid W^{(m)}, Y, \psi^{(m)})] =$$

$$= \sum_{n=1}^{N} \Big(-\frac{D_m}{2} \ln(2\pi) + \frac{D_m}{2} \mathbb{E}_q[\ln \psi^{(m)}] - \frac{1}{2} \mathbb{E}_q[\psi^{(m)}] \cdot$$

$$\cdot \mathbb{E}_q[(\underline{x}_n^{(m)} - W^{(m)} y_n)^T (\underline{x}_n^{(m)} - W^{(m)} y_n)] \Big) =$$

$$= -\frac{N D_m}{2} \ln(2\pi) + \frac{N D_m}{2} \left(\psi(a_\psi^{(m)}) - \ln(b_\psi^{(m)}) \right) -$$

$$- \frac{1}{2} \frac{a_{\psi^{(m)}}}{b_{\psi^{(m)}}} \Big[\sum_n \underline{x}_n^{(m)T} \underline{x}_n^{(m)} - 2 \operatorname{Tr}\{ \langle W^{(m)} \rangle \langle Y \rangle X^{(m)T} \} +$$

$$+ \operatorname{Tr}\{ \langle W^{(m)T} W^{(m)} \rangle \langle Y Y^T \rangle \} \Big] =$$

DEL CALCULO DE $q(\psi^{(m)})$

$$b_{\psi^{(m)}} = \beta_0^\psi - \frac{1}{2} \Big[\sum_n \sum_d x_{nd}^{(m)2} - 2 \operatorname{Tr}\{ \ \} + \operatorname{Tr}\{ \ \} \Big]$$

$$= -\frac{N D_m}{2} \ln(2\pi) + \frac{N D_m}{2} \left(\psi(a_\psi^{(m)}) - \ln(b_\psi^{(m)}) \right) + \frac{a_{\psi^{(m)}}}{b_{\psi^{(m)}}} \left(b_{\psi^{(m)}} - \beta_0^\psi \right)$$

$$= -\frac{N D_m}{2} \ln(2\pi) + \frac{N D_m}{2} \left(\psi(a_\psi^{(m)}) - \ln(b_\psi^{(m)}) \right) + a_{\psi^{(m)}} - \frac{a_{\psi^{(m)}}}{b_{\psi^{(m)}}} \beta_0^\psi$$

- $\mathbb{E}[\ln p(\psi^{(m)})] = \quad \psi^{(m)} \sim \text{Gamma}(\alpha_0^\psi, \beta_0^\psi)$

$$= \alpha_0^\psi \ln \beta_0^\psi - \beta_0^\psi \mathbb{E}_q[\psi^{(m)}] + (\alpha_0^\psi - 1) \mathbb{E}_q[\ln \psi^{(m)}] - \ln \Gamma(\alpha_0^\psi) =$$

$$= \alpha_0^\psi \ln \beta_0^\psi - \ln \Gamma(\alpha_0^\psi) - \beta_0^\psi \frac{a_{\psi^{(m)}}}{b_{\psi^{(m)}}} + (\alpha_0^\psi - 1) \left[\psi(a_\psi^{(m)}) - \ln(b_\psi^{(m)}) \right]$$

$\rightarrow$ JUNTAMOS $\mathbb{E}[\ln p(X^{(m)} \mid \dots)]$ y $\mathbb{E}[\ln p(\psi^{(m)})]$

$$\mathbb{E}[\ln p(X^{(m)}, \psi^{(m)} \mid W^{(m)}, Y)] = -\frac{N D_m}{2} \ln(2\pi) + a_{\psi^{(m)}} +$$

$$+ \alpha_0^\psi \ln \beta_0^\psi - \ln \Gamma(\alpha_0^\psi) + \left(\frac{N D_m}{2} + \alpha_0^\psi - 1 \right) \left[\psi(a_\psi^{(m)}) - \ln(b_\psi^{(m)}) \right]$$

$$- 2 \beta_0^\psi \frac{a_\psi^{(m)}}{b_{\psi^{(m)}}}$$

ALGUNAS UTILIDADES

$x \sim \mathcal{N}(\underline{m}, V)$ $\qquad p(\underline{x}) = \frac{1}{(2\pi)^{D/2} |V|^{1/2}} \exp\left(-\frac{1}{2} (\underline{x}-\underline{m})^T V^{-1} (\underline{x}-\underline{m})\right)$

$$\ln p(\underline{x}) = -\frac{D}{2} \ln (2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2} (\underline{x}-\underline{m})^T V^{-1} (\underline{x}-\underline{m})$$

$\alpha \sim \text{Gamma}(a_0, b_0)$ $\qquad p(\alpha) = b_0 e^{-b_0 \alpha} \frac{(b_0 \alpha)^{a_0 - 1}}{\Gamma(a_0)}$

$$\ln p(\alpha) = \ln b_0 - b_0 \alpha + (a_0 - 1) \ln b_0 + (a_0 - 1) \ln \alpha - \ln \Gamma(a_0)$$