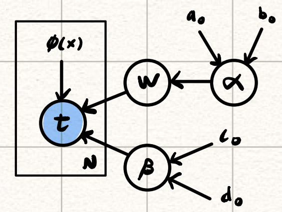
Joint distribution



$$p(\bar{t}, w, \alpha, \beta) = p(\bar{t} \mid w, \beta)\, p(w \mid \alpha)\, p(\alpha)\, p(\beta)$$

$$p(\bar{t} \mid w, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i \mid w^T \phi_i, \beta^{-1})$$

$$p(w \mid \alpha) = \mathcal{N}(w \mid 0, \alpha^{-1})$$

$\mathcal{N}(w \mid 0, \underbrace{\alpha^{-1} I}_{\Lambda^{-1}})$

$= (2\pi)^{-M/2} |\Lambda^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(w-0)^T \Lambda (w-0)\right)$

$$p(\alpha) = \mathrm{Gam}(\alpha \mid a_0, b_0)$$

Gamma distr. is conjugate prior for Gaussian distr.

$\mathrm{Gam}(\alpha \mid a_0, b_0)$

$= \Gamma(a_0)^{-1} b_0^{a_0} \alpha^{a_0 - 1} e^{-b_0 \alpha}$

$$p(\beta) = \mathrm{Gam}(\beta \mid c_0, d_0)$$

Variational distribution

$$q(w, \alpha, \beta) = q(w)\, q(\alpha)\, q(\beta)$$

Re-estimation equations from "General result 10.9"

$\boxed{\log q_j^*(Z_j) = \mathbb{E}_{-j}[\log \overbrace{p(X, Z)}^{p(\bar{t}, w, \alpha, \beta)}] + \text{const.}}$

1

$$\log q^*(\beta) = \log \mathbb{E}_{w,\alpha}\left[p(\bar{t} \mid w, \beta)\, p(w \mid \alpha)\, p(\alpha)\, p(\beta) \right] + \text{const.}$$

$$= \mathbb{E}_{w,\alpha}[\log p(\bar{t} \mid w, \beta)] + \mathbb{E}_{w,\alpha}[\log p(w \mid \alpha)] + \text{const.}$$

$\rightarrow$ const w.r.t. β

$$+ \mathbb{E}_{w,\alpha}[\log p(\alpha)] + \mathbb{E}_{w,\alpha}[\log p(\beta)] + \text{const.}$$

$\rightarrow$ const w.r.t. β

$$= \underbrace{\log p(\beta)}_{1} + \underbrace{\mathbb{E}_{w,\alpha}[\log p(\bar{t} \mid w, \beta)]}_{2} + \text{const.}$$

1 $\log p(\beta) = \log \mathrm{Gam}(\beta \mid c_0, d_0) = \log \Gamma(c_0)^{-1} d_0^{c_0} \beta^{c_0 - 1} e^{-d_0 \beta}$

$$= \log \Gamma(c_0)^{-1} + \log d_0^{c_0} + \log \beta^{c_0 - 1} + \log e^{-d_0 \beta}$$

$\rightarrow$ Const wrt β $\quad$ $\rightarrow$ Const wrt β

$$= \underline{(c_0 - 1) \log \beta - d_0 \beta}$$

② $$\log p(\bar{t} \mid w, \beta) = \log \prod_{i=1}^{N} \left[\left(\frac{\beta}{2\pi} \right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2} (w^T \phi_i - t_i)^2 \right) \right] \qquad \beta = \frac{1}{\sigma^2}$$

$$= \sum_{i=1}^{N} \left[\log \left(\frac{\beta}{2\pi} \right)^{\frac{1}{2}} - \frac{\beta}{2} (w^T \phi_i - t_i)^2 \right]$$

$$= N \cdot \log \left(\frac{\beta}{2\pi} \right)^{\frac{1}{2}} - \frac{\beta}{2} \sum_{i=1}^{N} (w^T \phi_i - t_i)^2$$

$$= \frac{N}{2} \left(\log \beta - \log(2\pi) \right) - \frac{\beta}{2} \sum_{i=1}^{N} (w^T \phi_i - t_i)^2$$

$\rightarrow$ Const wrt β

$$= \underline{\frac{N}{2} \log \beta - \frac{\beta}{2} \sum_{i=1}^{N} (w^T \phi_i - t_i)^2} + \text{const.}$$

$$\log q^*(\beta) = (c_0 - 1) \log \beta - d_0 \beta + \mathbb{E}_{w,\alpha} \left[\frac{N}{2} \log \beta - \frac{\beta}{2} \sum_{i=1}^{N} (w^T \phi_i - t_i)^2 \right] + \text{const}$$

$$= (c_0 - 1) \log \beta - d_0 \beta + \frac{N}{2} \log \beta - \frac{\beta}{2} \mathbb{E}_w \left[\sum_{i=1}^{N} (w^T \phi_i - t_i)^2 \right] + \text{const.}$$

$$\sum_{i=1}^{N} (w^T \phi_i - t_i)^2 = \underline{\| \Phi w - \bar{t} \|}^2$$

$$\left(\sqrt{a^2 + b^2} \right)^2 = a^2 + b^2$$

$$= \left(\frac{N}{2} + c_0 - 1 \right) \log \beta - \beta \left[\frac{1}{2} \underline{\mathbb{E}_w \| \Phi w - \bar{t} \|^2} + d_0 \right]$$

$$\mathbb{E}_w \| \Phi w - \bar{t} \|^2 = \mathbb{E}_w \left[(\Phi w)^2 - 2 \bar{t}^T \Phi w + (\bar{t})^2 \right]$$

$$= E_w\left[w^T \Phi^T \Phi w - 2 t^T \Phi w + t^T t\right]$$

*

$$\boxed{w^T \Phi^T \Phi w = \mathrm{Tr}\left(\Phi^T \Phi\, w w^T\right) \quad \text{yes, but why ?}}$$

$$= E_w\left[\mathrm{Tr}(\Phi^T \Phi\, w w^T) - 2 t^T \Phi w + t^T t\right]$$

$$= \mathrm{Tr}\left(\Phi^T \Phi\, E_w[w w^T]\right) - 2 t^T \Phi E_w[w] + t^T t$$

Known expression $E[x^2] = E[x]^2 + Var(x)$

$$= \left(\frac{N}{2} + c_0 - 1\right)\log\beta - \beta\left[\frac{1}{2}\mathrm{Tr}\left(\Phi^T\Phi\left(m_N m_N^T + S_N\right)\right) - t^T \Phi m_N + \frac{1}{2} t^T t + d_0\right]$$

$$\mathrm{Tr}\left[\Phi^T\Phi\left(m_N m_N^T + S_N\right)\right] = \mathrm{Tr}\left[\Phi^T\Phi\, m_N m_N^T + \Phi^T \Phi\, S_N\right]$$

$$= \mathrm{Tr}\left[\Phi^T\Phi\, m_N m_N^T\right] + \mathrm{Tr}\left[\Phi^T\Phi\, S_N\right]$$

Transform "back" using the initial "Trace func"

$$m_N^T \Phi^T \Phi m_N + \mathrm{Tr}\left[\Phi^T\Phi\, S_N\right]$$

$$= \left(\frac{N}{2} + c_0 - 1\right)\log\beta - \beta\left[\frac{1}{2}\mathrm{Tr}\left[\Phi^T\Phi S_N\right] + \frac{1}{2} m_N^T \Phi^T \Phi m_N - \frac{2}{2} t^T \Phi m_N + \frac{1}{2} t^T t + \frac{2}{2} d_0\right]$$

Reverse factorization $\|\Phi m_N - t\|^2$

$$\log q^*(\beta) = \left(\frac{N}{2} + c_0 - 1\right)\log\beta - \frac{\beta}{2}\left[\mathrm{Tr}\left[\Phi^T\Phi S_N\right] + \|\Phi m_N - t\|^2 + 2 d_0\right]$$

$$= \log \beta^{\left(\frac{N}{2} + c_0 - 1\right)} \quad \log e^{-\beta\left[d_0 + \frac{1}{2}\left(\mathrm{Tr}\left[\Phi^T\Phi S_N\right] + \|\Phi m_N - t\|^2\right)\right]}$$

Reckognizing $\log q^*(\beta)$ having form of log Gamma distr.

$$\log \mathrm{Gam}(\alpha \mid a_0, b_0) = \underbrace{\log \Gamma(\alpha_0)^{-1}}_{\to \text{norm.}} + \underbrace{\log b_0^{a_0}}_{\to \text{const.}} + \log \alpha^{a_0 - 1} + \log e^{-b_0 \alpha}$$

$\log q^*(\beta)$ is $\log \mathrm{Gam}(\beta \mid c_0, d_0)$ if

$$\boxed{\begin{aligned} q^*(\beta) &= \mathrm{Gam}(\beta \mid c_N, d_N) \\ c_N &= c_0 + \frac{N}{2} \\ d_N &= d_0 + \frac{1}{2}\left(\|\Phi m_N - t\|^2 + \mathrm{Tr}[\Phi^T \Phi S_N]\right) \end{aligned}}$$

$$c_N \cancel{-1} = \frac{N}{2} + c_0 \cancel{-1}$$

②

$$\log q^*(\alpha) = \log \mathbb{E}_{w,\beta}\left[p(t \mid w, \beta)\, p(w \mid \alpha)\, p(\alpha)\, p(\beta)\right] + \text{const.}$$

Disappears

$$= \log \mathbb{E}_{w,\beta}\left[p(w \mid \alpha)\right] + \log \mathbb{E}_{w,\beta}\left[p(\alpha)\right] + \text{const}$$

Same formulation as without $p(\beta)$

$$\boxed{\begin{aligned} q^*(\alpha) &= \mathrm{Gam}(\alpha \mid a_N, b_N) \\ a_N &= a_0 + \frac{M}{2} \\ b_N &= b_0 + \frac{1}{2}\mathbb{E}[w^T w] \end{aligned}}$$

$$\mathbb{E}[w^T w] = m_N^T m_N + \mathrm{Tr}(S_N)$$

③

$$\log q^*(w) = \log \mathbb{E}_{\alpha,\beta}\left[p(t \mid w, \beta)\, p(w \mid \alpha)\, p(\alpha)\, p(\beta)\right] + \text{const.}$$

const w.r.t. w

$$= \underset{①}{\mathbb{E}_\beta\left[\log p(t \mid w, \beta)\right]} + \underset{②}{\mathbb{E}_\alpha\left[\log p(w \mid \alpha)\right]}$$

①

$$\mathbb{E}_\beta\left[\log p(t \mid w, \beta)\right]$$

$$= \mathbb{E}_\beta\left[\log \prod_{i=1}^{N}\left[\left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta}{2}(w^T \phi_i - t_i)^2\right)\right]\right]$$

$$= E_\beta\left[\sum_{i=1}^{N}\left(\log\left(\frac{\beta}{2\pi}\right)^{1/2} + \log\exp\left(-\frac{\beta}{2}(w^T\phi_i - t_i)^2\right)\right)\right]$$

$$= E_\beta\left[\underbrace{\frac{1}{2}\sum_{i=1}^{N}\log\frac{\beta}{2\pi}}_{\rightarrow \text{ const wrt. } w}\right] - \frac{1}{2}E_\beta\left[\beta \underbrace{\sum_{i=1}^{N}(w^T\phi_i - t_i)^2}_{\|\Phi w - t\|^2}\right]$$

2

$$E_\alpha[\log p(w|\alpha)]$$

$$= E_\alpha\left[\log (2\pi)^{-M/2}|\Lambda^{-1}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(w-0)^T\Lambda(w-0)\right)\right], \quad \Lambda = \alpha I$$

$$= E_\alpha\left[\underbrace{-\frac{M}{2}\log 2\pi}_{\rightarrow \text{ const wrt. } w} \underbrace{-\frac{1}{2}\log|\Lambda^{-1}|}_{\rightarrow \text{ const wrt. } w} - \frac{1}{2}w^T(\alpha I)w\right]$$

$$= E_\alpha\left[-\frac{1}{2}\alpha w^T w\right] = -\frac{1}{2}E[\alpha]w^T w$$

$$\log q^*(w)$$

$$= -\frac{1}{2}E[\beta]\|\Phi w - t\|^2 - \frac{1}{2}E[\alpha]w^T w + \text{const}$$

$$= -\frac{1}{2}E[\beta]\left(w^T\Phi^T\Phi w - 2w^T\Phi^T t + t^T t\right) - \frac{1}{2}E[\alpha]w^T w + \text{const}$$

$$= -\frac{1}{2}E[\beta]w^T\Phi^T\Phi w + E[\beta]w^T\Phi^T t \underbrace{- \frac{1}{2}E[\beta]t^T t}_{\rightarrow \text{ const wrt. } w} - \frac{1}{2}E[\alpha]w^T w + \text{const}$$

$$= -\frac{1}{2}w^T(\underbrace{E[\beta]\Phi^T\Phi + E[\alpha]I}_{\Lambda \text{ term}})w + \underline{E[\beta]w^T\Phi^T t} + \text{const}$$

$\Sigma = \Lambda^{-1}$

General Gaussian

$$\log \mathcal{N}(x|\mu,\Lambda) \propto -\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)$$

$$= -\frac{1}{2}\left(x^T\Lambda(x-\mu) - \mu^T\Lambda(x-\mu)\right)$$

$$= -\frac{1}{2}\left(x^T\Lambda x - x^T\Lambda\mu - \mu^T\Lambda x + \mu^T\Lambda\mu\right)$$

$$= -\frac{1}{2} x^T \Lambda x + x^T \Lambda \mu + \text{const}$$

→ const. wrt w

↑ Terms unrelated to w

$$\Lambda = \mathbb{E}[\beta] \Phi^T \Phi + \mathbb{E}[\alpha] I , \quad \Lambda^{-1} = \Sigma$$

$$\Sigma = (\mathbb{E}[\beta] \Phi^T \Phi + \mathbb{E}[\alpha] I)^{-1}$$

$$w^T \Lambda \mu = \mathbb{E}[\beta] w^T \Phi^T \mathbf{t}$$

$$\Lambda \mu = \mathbb{E}[\beta] \Phi^T \mathbf{t}$$

$$\mu = \mathbb{E}[\beta] \Lambda^{-1} \Phi^T \mathbf{t} , \quad \Lambda^{-1} = \Sigma$$

$$\mu = \mathbb{E}[\beta] \Sigma \Phi^T \mathbf{t}$$

*

$$\boxed{\begin{aligned} q^*(w) &= \mathcal{N}(w \mid m_N, S_N) \\ m_N &= \mathbb{E}[\beta] S_N \Phi^T \mathbf{t} \\ S_N &= (\mathbb{E}[\beta] \Phi^T \Phi + \mathbb{E}[\alpha] I)^{-1} \end{aligned}}$$

$$\mathbb{E}[\alpha] = \frac{a_N}{b_N}$$

$$\mathbb{E}[\beta] = \frac{c_N}{d_N}$$

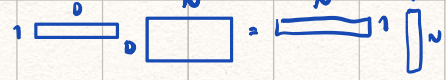
Variational lower bound $\mathcal{L}(q)$

Joint prob.

$$\mathcal{L}(q) = \mathbb{E}[\log p(\mathbf{t}, w, \alpha, \beta)] - \mathbb{E}[\log q(w, \alpha, \beta)]$$

$$= \underset{①}{\mathbb{E}[\log p(\mathbf{t} \mid w)]} + \underset{②}{\mathbb{E}[\log p(w \mid \alpha)]} + \underset{③}{\mathbb{E}[\log p(\alpha)]} + \underset{④}{\mathbb{E}[\log p(\beta)]}$$

$$- \underset{⑤}{\mathbb{E}[\log q(w)]} - \underset{⑥}{\mathbb{E}[\log q(\alpha)]} - \underset{⑦}{\mathbb{E}[\log q(\beta)]}$$



① $\mathbb{E}[\log p(\mathbf{t}|\mathbf{w})]$

$$= \frac{N}{2}\log\left(\frac{\mathbb{E}[\beta]}{2\pi}\right) - \frac{\mathbb{E}[\beta]}{2} t^T t + \mathbb{E}[\beta]\, \mu_N^T\, \Phi^T t - \frac{\mathbb{E}[\beta]}{2} \mathrm{Tr}\left[\Phi^T \Phi (\mu_N \mu_N^T + \Sigma_N)\right]$$

② $\mathbb{E}[\log p(\mathbf{w}|\alpha)]$

$$= -\frac{M}{2}\log(2\pi) + \frac{M}{2}(\psi(a_N) - \log b_N) - \frac{a_N}{2b_N}\left[\mu_N^T \mu_N + \mathrm{Tr}(\Sigma_N)\right]$$

③ $\mathbb{E}[\log p(\alpha)]$

$$= a_0 \log b_0 + (a_0 - 1)\left[\psi(a_N) - \log b_N\right] - b_0 \frac{a_N}{b_N} - \log \Gamma(a_N)$$

④ $\mathbb{E}[\log p(\beta)]$

$$= c_0 \log d_0 + (c_0 - 1)\left[\psi(c_N) - \log d_N\right] - d_0 \frac{c_N}{d_N} - \log \Gamma(c_N)$$

⑤ $\mathbb{E}[\log q(\mathbf{w})]$

$$= -\frac{1}{2}\log|\Sigma_N| - \frac{M}{2}(1 + \log 2\pi)$$

⑥ $\mathbb{E}[\log q(\alpha)]$

$$= -\log \Gamma(a_N) + (a_N - 1)\psi(a_N) + \log b_N - a_N$$

⑦ $\mathbb{E}[\log q(\beta)]$

$$= -\log \Gamma(c_N) + (c_N - 1)\psi(c_N) + \log d_N - c_N$$