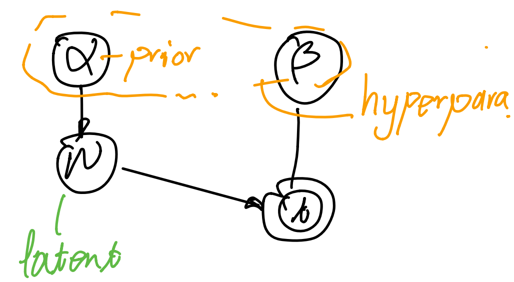
Q1.1

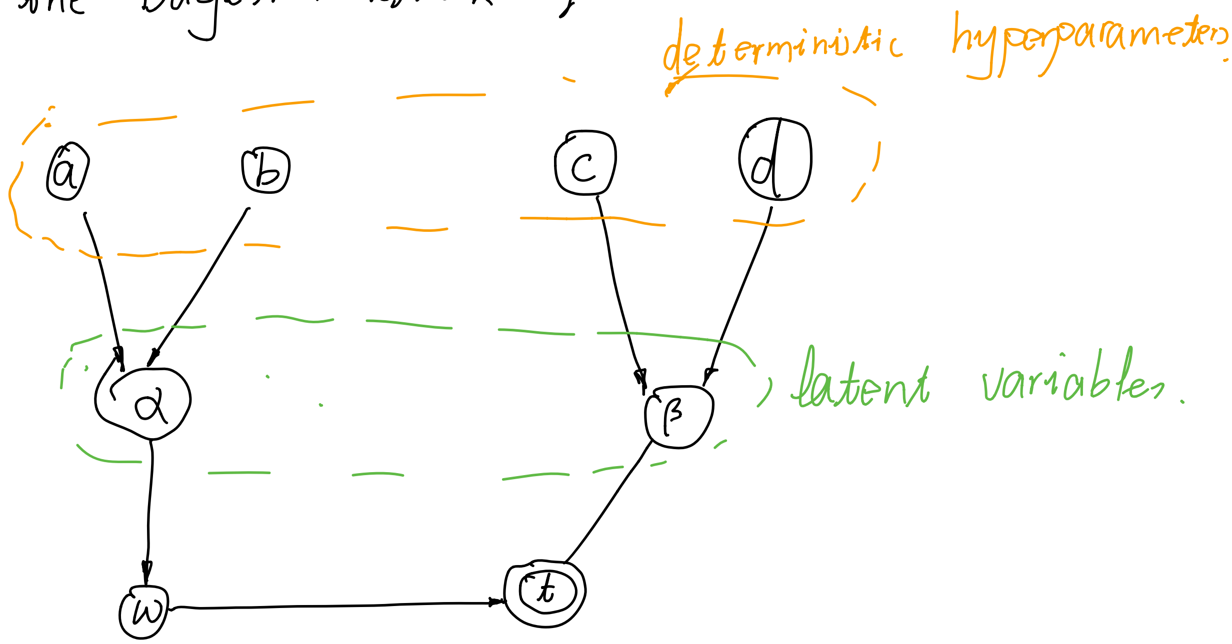
Differences.	SBLR	BLR.
latent variables.	$\alpha \beta$	w
deterministic para.	a,b,c,d	α, β
priors	$\alpha, \beta,$	α

the Bayesian network of the model in the homework



Q1.2

the Bayesian network of SBLR.



$$P(\vec{t}, \vec{w}, \vec{\alpha}, \beta) = \mathcal{N}(\vec{t}; \Phi \vec{w}, \beta^{-1} I)\, \mathcal{N}(\vec{w}; 0, \mathrm{diag}(\vec{\alpha})^{-1}) \cdot \left(\prod_{j=1}^{J} \mathcal{G}(\alpha_j; a, b)\right) \mathcal{G}(\beta; c, d)$$

$$= \left(\prod_{i=1}^{N} \mathcal{N}(t_i; \vec{w}^T \vec{\phi}(x_i), \beta^{-1})\right) \left(\prod_{j=1}^{J} \mathcal{N}(w_j; 0, \alpha_j^{-1})\right) \left(\prod_{j=1}^{J} \mathcal{G}(\alpha_j; a, b)\right) \mathcal{G}(\beta \ldots)$$

A 2.1. $q^*(\vec{w}) = \mathcal{N}(\vec{w}; \vec{\mu}, \Sigma_N)$.

$$\ln q^*(\vec{w}) = \mathbb{E}_{q(\vec{\alpha})q(\beta)}\left[\ln P(\vec{t}, \vec{w}, \vec{\alpha}, \beta)\right] + cst$$

$$= \mathbb{E}_{q(\vec{\alpha})q(\beta)}\left[\sum_{i=1}^{N} \ln \mathcal{N}(t_i; \vec{w}^T\vec{\phi}(x_i), \beta^{-1}) + \ln \mathcal{N}(\vec{w}; 0, diag(\vec{\alpha})^{-1})\right] + cst$$

$$= \mathbb{E}_{q(\vec{\alpha})q(\beta)}\left[\sum_{i=1}^{N} \left(-\frac{\beta}{2}\right)(t_i - \vec{w}^T\vec{\phi}(x_i))^2 - \frac{1}{2} diag(\vec{\alpha})\vec{w}^2\right] + cst$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{N} \mathbb{E}_{q(\beta)}\beta\left[t_i^2 - 2\vec{w}^T\vec{\phi}(x_i)t_i + \vec{w}^T\vec{\phi}(x_i)\vec{\phi}^T(x_i)\vec{w}\right] + \mathbb{E}_{q(\vec{\alpha})}\left[diag(\vec{\alpha})\vec{w}^2\right]\right) + cst$$

$$= -\frac{1}{2}\left(\mathbb{E}_{q(\beta)}[\beta]\sum_{i=1}^{N}\left[-2\vec{w}^T\vec{\phi}(x_i)t_i + \vec{w}^T\vec{\phi}(x_i)\vec{\phi}^T(x_i)\vec{w}\right] + \mathbb{E}_{q(\vec{\alpha})}[\sim]\vec{w}^2\right) + cst$$

$$= -\frac{1}{2}\left(\underbrace{\mathbb{E}_{q(\beta)}[\beta]\sum_{i=1}^{N}\left(-2t_i\vec{\phi}^T(x_i)\vec{w}\right)}_{-\frac{2\vec{\mu}_N}{\Sigma_n}\vec{w}} + \underbrace{\left\{\mathbb{E}_{q(\beta)}[\beta]\sum_{i=1}^{N}\vec{\phi}^T(x_i)^2 + \mathbb{E}_{q(\vec{\alpha})}\left[diag(\vec{\alpha})\right]\right\}\vec{w}^2}_{\frac{1}{\Sigma_N}\vec{w}^2}\right) + cst$$

$$\Rightarrow \begin{cases} -2\,\dfrac{\vec{\mu}_N}{\Sigma_n} = \mathbb{E}_{q(\beta)}[\beta]\displaystyle\sum_{i=1}^{N}\left(-2t_i\vec{\phi}^T(x_i)\right) \\ \\ \dfrac{1}{\Sigma_N} = \mathbb{E}_{q(\beta)}[\beta]\displaystyle\sum_{i=1}^{N}\left\{\vec{\phi}^T(x_i)^2\right\} + \mathbb{E}_{q(\vec{\alpha})}\left[diag(\vec{\alpha})\right] \end{cases}$$

$$\Rightarrow \begin{cases} \vec{\mu}_N = \mathbb{E}_{q(\beta)}[\beta]\, \Sigma_N \Phi^T \vec{t} \\ \Sigma_N^{-1} = \mathbb{E}_{q(\vec{\alpha})}[\mathrm{diag}(\vec{\alpha})] + \mathbb{E}_{q(\beta)}[\beta]\, \Phi^T \Phi . \end{cases}$$

A2.2. $q^*(\vec{\alpha}) = \prod_{j=1}^{M} \mathcal{G}(\alpha_j ; \hat{a}, \hat{b}_j)$.

$$\ln q^*(\vec{\alpha}) = \mathbb{E}_{q(\vec{w})q(\beta)}[\ln p(\vec{t}, \vec{w}, \vec{\alpha}, \beta)] + const.$$

$$= \mathbb{E}_{q(\vec{w})q(\beta)}\left[\ln \prod_{j=1}^{M} \mathcal{N}(w_j ; 0, \alpha_j^{-1}) + \ln \prod_{j=1}^{M} \mathcal{G}(\alpha_j ; a, b)\right] + const$$

$$= \mathbb{E}_{q(\vec{w})} \sum_{j=1}^{M} \left(\frac{1}{2} \ln \alpha_j - \frac{\alpha_j}{2} w_j^2 + (a-1) \ln \alpha_j - b \alpha_j \right) + const$$

$$= \sum_{j=1}^{n} \left(\underbrace{(\alpha + \tfrac{1}{2} - 1)}_{\hat{\alpha}-1} \ln \alpha_j - \underbrace{\left(b + \tfrac{1}{2} \mathbb{E}_{q(\vec{w})}[w_j^2]\right)}_{\tilde{b}} \alpha_j \right) + const$$

$$\Rightarrow \begin{cases} \hat{\alpha} = \alpha + \frac{1}{2} \\ \hat{b} = b + \frac{1}{2} \mathbb{E}_{q(\vec{w})}[w_j^2] . \end{cases}$$

A2.3. $q^*(\beta) = \mathcal{G}(\beta; \tilde{c}, \tilde{d})$.

$$\ln q^*(\beta) = \mathbb{E}_{q(\vec{w})q(\vec{\alpha})}[\ln P(\vec{t}, \vec{w}, \vec{\alpha}, \beta)] + \text{cst}$$

$$= \mathbb{E}_{q(\vec{w})q(\vec{\alpha})}\left\{\sum_{i=1}^{N}\left[\ln \mathcal{N}(t_i; \vec{w}^T\vec{\phi}(x_i), \beta^{-1})\right] + \ln \mathcal{G}(\beta; c, d)\right\} + \text{cst}$$

$$= \mathbb{E}_{q(\vec{w})}\left[\sum_{i=1}^{N}\left(\frac{1}{2}\ln\beta - \frac{\beta}{2}(t_i - \vec{w}^T\vec{\phi}(x_i))^2\right) + (c-1)\ln\beta - d\beta\right] + \text{cst}$$

$$= \sum_{i=1}^{N}\frac{1}{2}\ln\beta - \frac{\beta}{2}\mathbb{E}_{q(\vec{w})}\left[\sum_{i=1}^{N}(t_i - \vec{w}^T\vec{\phi}(x_i))^2\right] + (c-1)\ln\beta - d\beta + \text{cst}$$

$$= -\frac{\beta}{2}\mathbb{E}_{q(\vec{w})}\left[\|\vec{t} - \Phi\vec{w}\|_2^2\right] + (c - 1 + \frac{N}{2})\ln\beta - d\beta + \text{cst}$$

$$= -\underbrace{\left(d + \frac{1}{2}\mathbb{E}_{q(\vec{w})}\left[\|\vec{t} - \Phi\vec{w}\|_2^2\right]\right)}_{\tilde{d}}\beta + (\underbrace{c + \frac{N}{2} - 1}_{\tilde{c} - 1})\ln\beta + \text{cst}$$

$$\Rightarrow \begin{cases} \tilde{d} = d + \frac{1}{2}\mathbb{E}_{q(\vec{w})}\left[\|\vec{t} - \Phi\vec{w}\|_2^2\right] \\ \tilde{c} = c + \frac{N}{2}. \end{cases}$$