$$D = (x_i, y_i)_{i=1}^{N} \qquad p(D|w) = \prod_{i=1}^{N} p(y_i|x_i, w); \quad w \in \mathbb{R}^D$$

1. Learn parameters w
2. Select τ for prior

ARD-prior (automatic relevance determ.) $\quad p(w|\tau) = \prod_{i=1}^{D} \mathcal{N}(w_i|0, \tau_i^{-1})$

$$\tau^* = \arg\max_{\tau} p(D|\tau) = \arg\max_{\tau} \underbrace{\int \overbrace{p(D|w)}^{NN} p(w|\tau)}_{\text{EVIDENCE}} dw$$

$$\log p(D|\tau) \geq \mathcal{L}(\phi, \tau) = \langle \log p(D|w) \rangle_{q(w;\phi)} - \underbrace{D_{KL}[q(w;\phi) \| p(w;\tau)]}$$

$$q(w;\phi) = \prod_{i=1}^{D} \mathcal{N}(w_i|\mu_i, \sigma_i^2)$$

2. τ Selection

$$\tau_i \qquad D_{KL}[q\|p] = \sum_{i=1}^{D} D_{KL}[\mathcal{N}(w_i|\mu_i, \sigma_i^2) \| \mathcal{N}(w_i|0, \tau_i^{-1})]$$

FOC for τ_i

$$\nabla_{\tau_i}\left(\langle -\frac{1}{2}\log \tau_i^{-1} - \frac{1}{2}\tau_i w_i^2 \rangle_{q(w_i)}\right) := \frac{1}{2}\frac{1}{\tau_i} - \frac{1}{2}\langle w_i^2 \rangle_q = 0$$

$$\Rightarrow \boxed{\tau_i = \frac{1}{\langle w_i^2 \rangle_q} = \frac{1}{\mu_i^2 + \sigma_i^2}} \quad \text{ARM}$$

$$D_{KL}[\underbrace{\mathcal{N}(w_i|\mu_i,\sigma_i)}\,||\,\underbrace{\mathcal{N}(w_i|0,\underbrace{\tau_i^{-1}})}] =$$

$$= \langle \log \mathcal{N}(w_i|0,\tau_i^{-1}) \rangle_{\mathcal{N}(w_i|\mu_i,\sigma_i)} = 0$$

$$= \langle -\frac{1}{2}\log \tau_i^{-1} - \frac{1}{2} w_i^2 \tau_i \rangle_{\mathcal{N}(w_i|\mu_i,\sigma_i)}$$

$$D_{KL}[q||p] =$$

$$= \int q \log \frac{q}{p}\, d\theta =$$

$$= \underbrace{\int q \log q\, d\theta}_{-H[q]}$$

$$- \int q \log p\, d\theta$$

$$\mathcal{N}(w_i|0,\tau_i^{-1}) = |2\pi\tau_i^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} w_i^2 \tau_i\right)$$

$$\mathcal{L}^*(\mu,\sigma) = \sum_{i=1}^{N} \langle \log p(y_i|x_i,w)\rangle_{q(w|\mu,\sigma)} - \frac{1}{2}\sum_{i=1}^{d} \log\left(1+\frac{\mu_i^2}{\sigma_i^2}\right) \to \max_{\mu,\sigma}$$

$$\mathcal{L}_{SVDO}(\mu,\alpha) = \sum_{i=1}^{N} \langle \log p(y_i|x_i,w)\rangle_{q(w|\mu,\sigma)} + \sum_{i=1}^{D}\left[k_1\sigma(k_2+k_3\log\alpha_j) - \frac{1}{2}\log\left(1+\frac{1}{\alpha}\right)\right]$$

$$\boxed{\alpha = \frac{\sigma_i^2}{\mu_i^2}}$$

$$p(w) \propto \frac{1}{|w|}$$

$$KL\left[q(w)\,\|\,\prod_{i=1}^{D}\frac{1}{|w_i|}\right]$$

$$\mathcal{N}(w|\mu,\sigma^2)$$

$$\mathcal{N}(w|\mu,\alpha\mu^2) \Rightarrow \alpha = \frac{\sigma^2}{\mu^2}$$

$$q(w_i)=\mathcal{N}(w_i|\mu_i,\sigma_i) \longrightarrow w_i=\mu_i+\sigma_i\varepsilon \qquad \varepsilon\sim\mathcal{N}(0,1)$$

$$q(w_i)=\mathcal{N}(w_i|\mu_i,\alpha_i\mu_i^2) \longrightarrow w_i=\mu_i(1+\underbrace{\sqrt{\alpha_i}\,\varepsilon})$$

$$\frac{\partial}{\partial\mu_i}\mathcal{L}(w)=\vec{1}^T\cdot\frac{\partial\mathcal{L}}{\partial w} \quad ; \quad \frac{\partial}{\partial\mu_i}=\vec{1}^T\cdot\frac{\partial\mathcal{L}}{\partial w}+\underbrace{\sqrt{\alpha_i}\,\varepsilon^T}\frac{\partial\mathcal{L}}{\partial w}$$

$$B = A\, W^{d \times d_0}$$

B: $B \times d_0$, A: $B \times d$

$$\prod_{i,j} \mathcal{N}(w_{ij} \mid \mu_{ij}, \sigma_{ij}^2)$$

$$B \sim \mathcal{N}(A\mu, ?)$$

$$B = A\mu + \varepsilon \odot \sqrt{A \odot A\ \sigma^2}$$

$$A \cdot (\varepsilon \sigma + \mu)$$

$$\mathrm{Var}(b_{nk}) = A_{nj}^2 \cdot \sigma_j^2$$

$$A_{nj} A_{mi} \delta_{ij} \delta_{kl} \sigma_{jk}^2$$

$$A_{nj} A_{mi} \overbrace{\mathrm{cov}(W_{jk}, W_{il})}$$

$$b_{nk} = A_{nj} W_{jk}$$

$$\langle b_{nk} \rangle = A_{nj} \langle W_{jk} \rangle = A_{nj} \mu_{jk}$$

$$\langle b_{nk}\, b_{ml} \rangle = \langle A_{nj} W_{jk} A_{mi} W_{il} \rangle =$$

$$= A_{nj} A_{mi} \langle W_{jk} W_{il} \rangle$$

$$\mathrm{cov}(b_{nk}, b_{ml}) = A_{nj} A_{mi} \left[\langle W_{jk} W_{il} \rangle - \langle W_{jk} \rangle \times \langle W_{il} \rangle \right]$$

$$B = A @ \mu + \varepsilon \odot \left[A \odot A @ \sigma^2 \right]^{\frac{1}{2}} \longrightarrow \text{element wise.}$$

pre-activ.

samples same size as B

Objective:

$$N \times \frac{1}{N} \sum_{i=1}^{N} \langle \log p(D|w) \rangle_{q(w;\phi)} - \frac{1}{2} \sum_{i=1}^{d} \log\left(1 + \overbrace{\frac{\mu_i^2}{\sigma_i^2}}^{\alpha^{-1}}\right)$$

$$x^0 \to xW^1 \to f^1 \to x^1$$

$$x^1 \to xW^2 \to f^2 \to x^2$$

$$\int q(W^1)q(W^2)\ f^2\left[f^1(xW^1)W^2\right] dW^1 dW^2 =$$

$$W^1, W^2$$

$$=$$