$$p = \frac{1}{Z} \exp(f(\theta))$$

$$KL[q \| p] = \int q \log \frac{q}{p} d\theta$$

$$= \int q \left[\log q - \log \frac{1}{Z} \exp(f(\theta)) \right] d\theta.$$

$$KL[p \| q] = \int \frac{1}{Z} \exp(f(\theta)) \log \frac{\frac{1}{Z}\exp(f(\theta))}{q} d\theta$$

?

$$p = \{\theta_1, \dots, \theta_N\} \quad \frac{1}{N}\sum_{n=1}^{N} \delta(\theta - \theta_n)$$

$$\underset{\min}{KL[q(w)\|p(w)]} = \int q(w) \log \frac{q(w)}{p(w)}\, dw =$$

$$= \langle \log q(w) \rangle_q - \langle \log p_u(w) \rangle_q$$

$$\langle -\log Z + \log \mathcal{N}(w|\mu, \Sigma) + \sum_{n=1}^{N} \log \phi_n(w) \rangle_q$$

$$= \langle \log q(w) \rangle_q - \langle \log \mathcal{N}(w|\mu, \Sigma) \rangle_q - \sum_{n=1}^{N} \langle \log \phi_n(w) \rangle_q + \log Z$$

$KL \geq 0$

$$\Rightarrow \underline{\log Z} \geq \overbrace{\underbrace{- \langle \log q(w) \rangle_q}_{\substack{\text{entropy of } q \\ \approx \frac{1}{2}\log|2\pi e \Sigma|}} + \underbrace{\langle \log \mathcal{N}(w|\mu, \Sigma) \rangle_q}_{\substack{\approx \text{ quadratic form} \\ \text{over param of } q}} + \sum_{n=1}^{N} \underbrace{\langle \log \phi_n(w) \rangle_q}_{\langle \phi \rangle_q \;\; w \in \mathbb{R}^D}}^{\max_q} .$$

1. $q(w) = \mathcal{N}(w \mid m, S)$

$$w^T h_n \Rightarrow \mathcal{N}(w^T h_n \mid \cdot, \cdot)$$

$$E_w w^T h_n = m^T h_n \,; \quad E_w (w^T h_n)^2 = E_w \overbrace{w^T h_n}^{h_n^T w} w^T h_n =$$

$$= E_w h_n^T \underbrace{w w^T}_{E_w w w^T = S + m m^T} h_n = \underbrace{h_n^T S h_n + (h_n^T m)^2}$$

$$\sigma^2(w^T h_n) = h_n^T S h_n .$$

2. $\langle \log \phi(w^T h) \rangle_q = \langle \log \phi(y) \rangle_{\mathcal{N}(y \mid h^T m,\, h^T S h)}$

$\mathbb{R}^D$

$\} y = w^T h$

$\mathbb{R}$

$$= \langle \log \phi (h^T m + z \cdot (h^T S h)^{\frac{1}{2}}) \rangle_{\mathcal{N}(z \mid 0, 1)} .$$

$$q = \mathcal{N}(w \mid m, S) \qquad ww^T \qquad w^T w$$

$$\langle \log \mathcal{N}(w \mid \mu, \Sigma) \rangle_{q(w)} =$$

$$= \langle \underbrace{\log |2\pi\Sigma|^{-\frac{1}{2}}} - \frac{1}{2} ((w-\mu)^T \Sigma^{-1} (w-\mu)) \rangle_q =$$

$$= -\frac{1}{2} \log |2\pi\Sigma| - \frac{1}{2} \langle \{ \mathrm{tr}\,(w-\mu)^T \Sigma^{-1} (w-\mu) \} \rangle_q =$$

$$= \frac{1}{2} \langle \mathrm{tr} \{ \Sigma^{-1} (\underbrace{ww^T} + \mu\mu^T - 2w\mu^T) \} \rangle_q =$$

$$= \frac{1}{2} \mathrm{tr} \{ \Sigma^{-1} (S + mm^T + \mu\mu^T - 2m\mu^T) \}$$

$$-\frac{1}{2} \log |2\pi\Sigma| + \frac{1}{2} \mathrm{tr} \{ \Sigma^{-1} S \} + \frac{1}{2} \mu^T \Sigma^{-1} \mu - m^T \Sigma^{-1} \mu .$$