Gradients

$$\max_{w,\phi} \sum_i \mathbb{E}_{q_i} \log p(x_i \mid t_i, w) - \mathcal{KL}(q_i(t_i) \parallel p(t_i))$$

Gradients

$$\max_{w,\phi} \sum_i \mathbb{E}_{q_i} \log p(x_i \mid t_i, w) - \mathcal{KL}(q_i(t_i) \parallel p(t_i))$$

Easy and analytical

$$\mathcal{KL}(q_i(t_i) \parallel p(t_i))$$

$$= \sum_j \left(-\log \sigma_j(t_i) + \frac{\sigma_j^2(t_i) + \mu_j^2(t_i) - 1}{2} \right)$$

Gradients

$$f(w, \phi) = \sum_i \mathbb{E}_{q_i} \log p(x_i \mid t_i, w)$$

Gradients

$$f(w, \phi) = \sum_i \mathbb{E}_{q_i} \log p(x_i \mid t_i, w)$$

$$q_i(t_i) = q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \mathrm{diag}(s_i^2))$$

Gradients

$$f(w, \phi) = \sum_i \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$q_i(t_i) = q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \mathrm{diag}(s_i^2))$$

Gradients

$$\nabla_w f(w, \phi) = \nabla_w \sum_{i=1}^{N} \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

Gradients

$$\nabla_w f(w, \phi) = \nabla_w \sum_{i=1}^{N} \int q(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w) dt_i$$

Gradients

$$\nabla_w f(w, \phi) = \sum_i \nabla_w \int q(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w) dt_i$$

Gradients

$$\nabla_w f(w, \phi) = \sum_i \int \nabla_w q(t_i \mid x_i, \phi) \log p(x_i \mid t_i, w) dt_i$$

Gradients

$$\nabla_w f(w, \phi) = \sum_i \int q(t_i \mid x_i, \phi) \nabla_w \log p(x_i \mid t_i, w) dt_i$$

Gradients

$$\nabla_w f(w, \phi) = \sum_i \int q(t_i \mid x_i, \phi) \nabla_w \log p(x_i \mid t_i, w) dt_i$$

$$= \sum_i \mathbb{E}_{q(t_i \mid x_i, \phi)} \nabla_w \log p(x_i \mid t_i, w)$$

Gradients

$$\nabla_w f(w, \phi) = \sum_i \int q(t_i \mid x_i, \phi) \nabla_w \log p(x_i \mid t_i, w) dt_i$$

$$= \sum_i \mathbb{E}_{q(t_i \mid x_i, \phi)} \nabla_w \log p(x_i \mid t_i, w)$$

$$\approx \sum_i \nabla_w \log p(x_i \mid \widehat{t_i}, w)$$

$$\widehat{t_i} \sim q(t_i \mid x_i, \phi)$$

Gradients

$$\nabla_w f(w, \phi) \approx \sum_{i=1}^{N} \nabla_w \log p(x_i \mid \widehat{t}_i, w)$$

Gradient of standard NN

$$\widehat{t}_i \sim q(t_i \mid x_i, \phi)$$

Gradients

$$\nabla_w f(w, \phi) \approx \sum_{i=1}^{N} \nabla_w \log p(x_i \mid \widehat{t}_i, w)$$

$$\approx \frac{N}{n} \sum_{s=1}^{n} \nabla_w \log p(x_{i_s} \mid \widehat{t}_{i_s}, w)$$

Stochastic gradient of standard NN

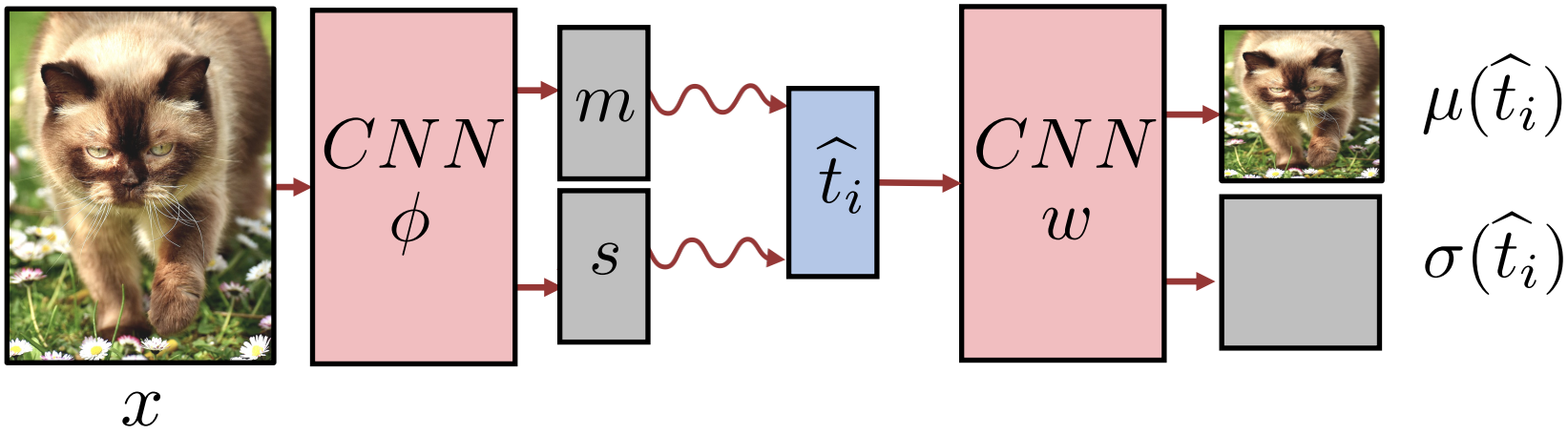
$$\widehat{t}_i \sim q(t_i \mid x_i, \phi)$$

$$i_s \sim \mathcal{U}\{1, \ldots, N\}$$

Gradients

$$\nabla_w f(w,\phi) \approx \sum_{i=1}^{N} \nabla_w \log p(x_i \mid \hat{t}_i, w)$$

$$\approx \frac{N}{n} \sum_{s=1}^{n} \nabla_w \log p(x_{i_s} \mid \hat{t}_{i_s}, w)$$



$$\hat{t}_i \sim \mathcal{N}(m(x_i,\phi), \mathrm{diag}(s^2(x_i,\phi))$$