Scaling up Variational EM

$$\underset{w,q}{\text{maximize}} \quad \mathcal{L}(w, q)$$

$$\text{subject to} \quad q_i(t_i) = \widetilde{q}_i(t_{i1}) \ldots \widetilde{q}_i(t_{im})$$

Scaling up Variational EM

$$\underset{w, q_1, \ldots, q_N}{\text{maximize}} \quad \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{subject to} \quad q_i(t_i) = \widetilde{q}_i(t_{i1}) \ldots \widetilde{q}_i(t_{im})$$

Scaling up Variational EM

$$\underset{\substack{w, m_1, \ldots, m_N \\ s_1, \ldots, s_N}}{\text{maximize}} \quad \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

Scaling up Variational EM

$$\begin{aligned} \underset{\substack{w, m_1, \ldots, m_N \\ s_1, \ldots, s_N}}{\text{maximize}} \quad & \mathcal{L}(w, q_1, \ldots, q_N) \\ \text{subject to} \quad & q_i(t_i) = \mathcal{N}(m_i, \text{diag}(s_i^2)) \end{aligned}$$

But this way ~100 parameters for each training object

Scaling up Variational EM

$$\underset{\substack{w, m_1, \ldots, m_N \\ s_1, \ldots, s_N}}{\text{maximize}} \quad \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m_i, \text{diag}(s_i^2))$$

But this way ~100 parameters for each training object

And is not clear what is m, s for test objects

Scaling up Variational EM

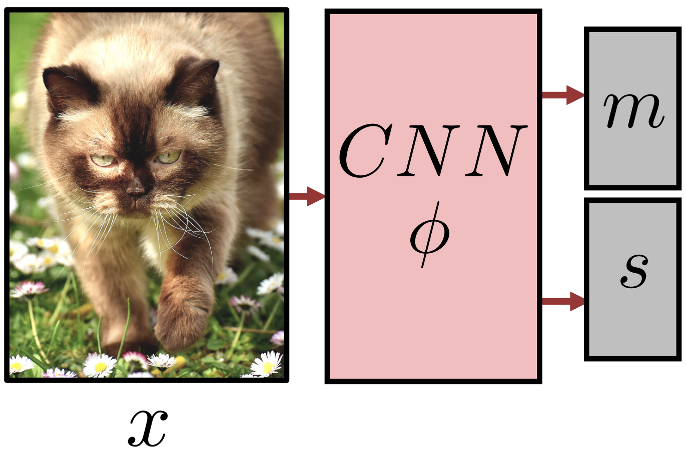
$$\underset{w,\phi}{\text{maximize}} \quad \mathcal{L}(w, q_1, \ldots, q_N)$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \operatorname{diag}(s^2(x_i, \phi)))$$

Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \mathcal{L}(w, q_1, \ldots, q_N)$$

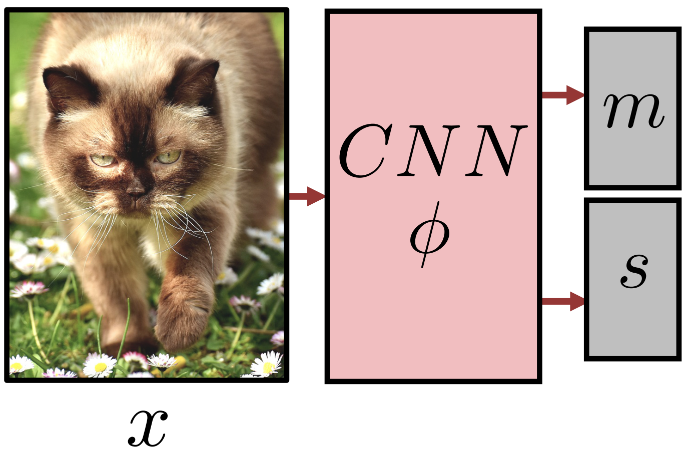
$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \operatorname{diag}(s^2(x_i, \phi)))$$



Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

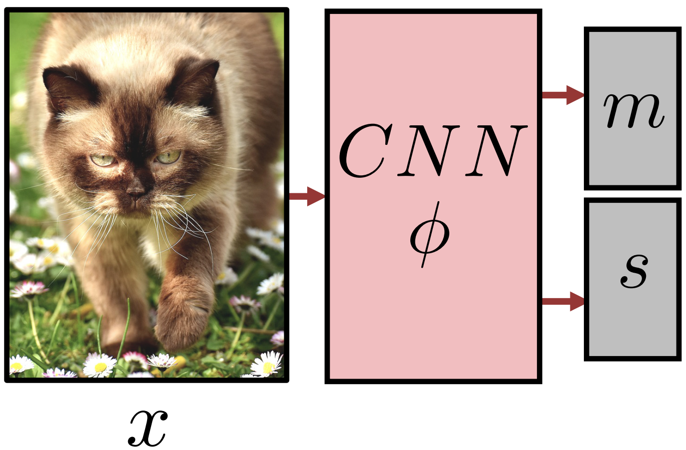
$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \text{diag}(s^2(x_i, \phi)))$$



Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \text{diag}(s^2(x_i, \phi)))$$

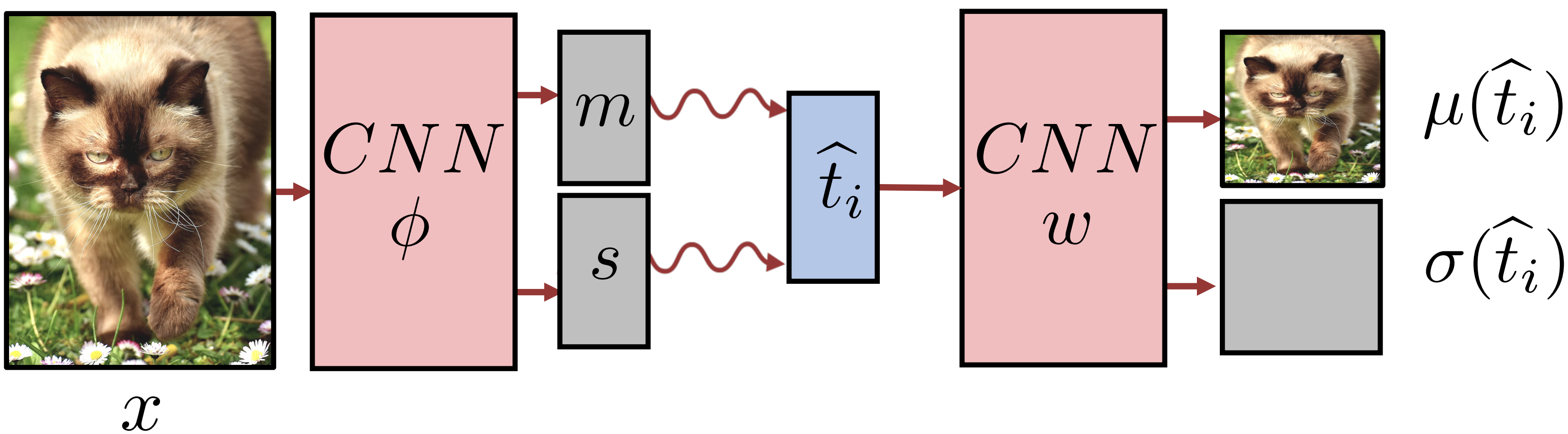


$$\widehat{t_i} \sim \mathcal{N}(m(x_i, \phi), \text{diag}(s^2(x_i, \phi))$$

Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \operatorname{diag}(s^2(x_i, \phi)))$$

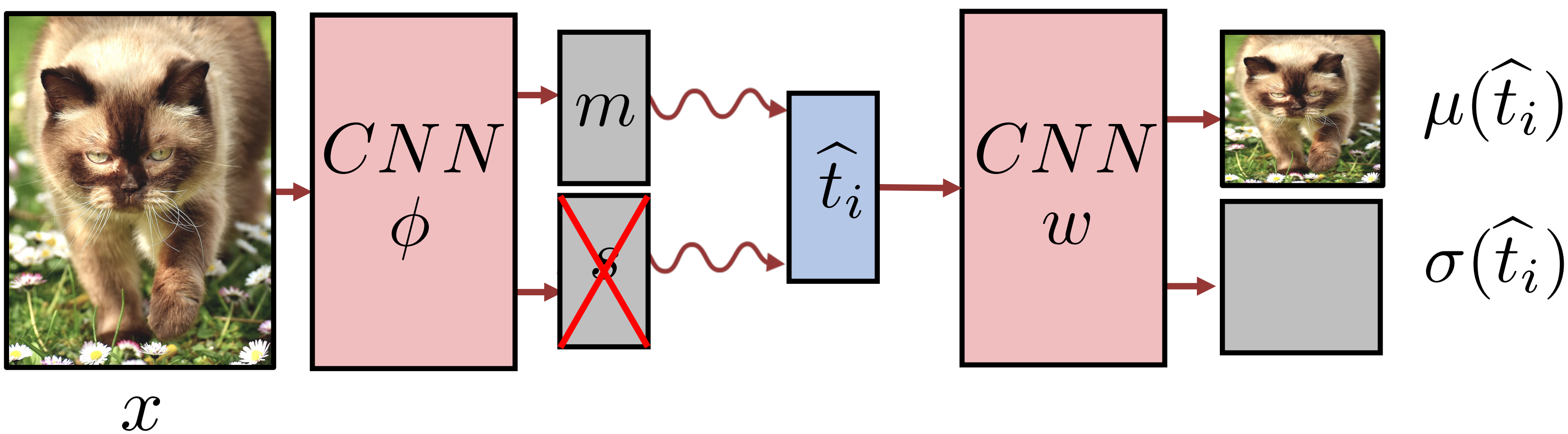


$$\widehat{t}_i \sim \mathcal{N}(m(x_i, \phi), \operatorname{diag}(s^2(x_i, \phi)))$$

Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i,\phi), \operatorname{diag}(s^2(x_i,\phi)))$$

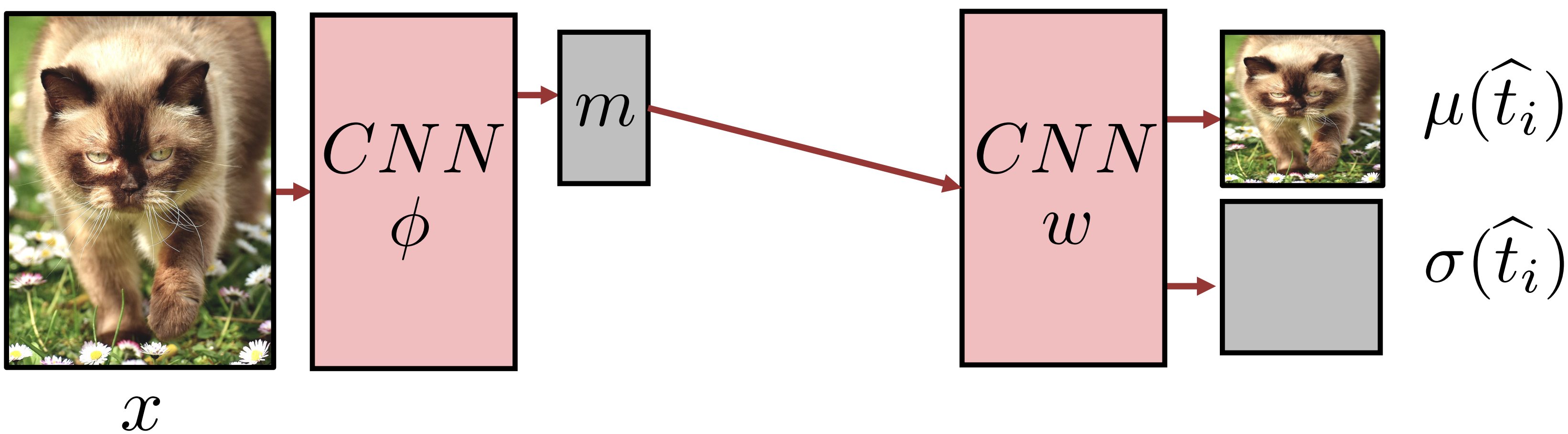


If $s(x) = 0$ then $\widehat{t}_i = m(x_i, \phi)$: usual autoencoder

Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \operatorname{diag}(s^2(x_i, \phi)))$$

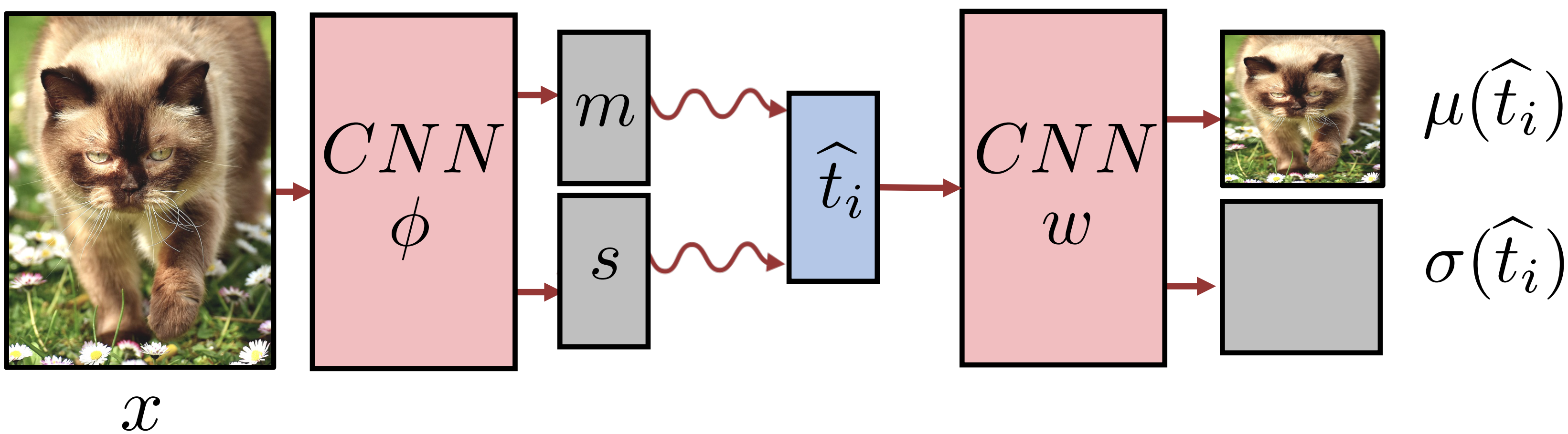


If $s(x) = 0$ then $\widehat{t_i} = m(x_i, \phi)$: usual autoencoder

Scaling up Variational EM

$$\underset{w,\phi}{\text{maximize}} \quad \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m(x_i, \phi), \text{diag}(s^2(x_i, \phi)))$$



$$\widehat{t_i} \sim \mathcal{N}(m(x_i, \phi), \text{diag}(s^2(x_i, \phi)))$$

Objective interpretation

$$\max. \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

Objective interpretation

$$\max. \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

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

Objective interpretation

$$\max. \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

$$= \sum_i \mathbb{E}_{q_i} \log p(x_i \mid t_i, w) + \underbrace{\mathbb{E}_{q_i} \log \frac{p(t_i)}{q_i(t_i)}}_{-\mathcal{KL}(q_i(t_i) \,\|\, p(t_i))}$$

Objective interpretation

$$\max. \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w) p(t_i)}{q_i(t_i)}$$

$$= \sum_i \mathbb{E}_{q_i} \underbrace{\log p(x_i \mid t_i, w)}_{-\|x_i - \mu(t_i)\|^2 + \text{const}} - \mathcal{KL}(q_i(t_i) \parallel p(t_i))$$

If $\sigma(x_i) = 1$ for simplicity

Objective interpretation

$$\max. \sum_i \mathbb{E}_{q_i} \log \frac{p(x_i \mid t_i, w)p(t_i)}{q_i(t_i)}$$

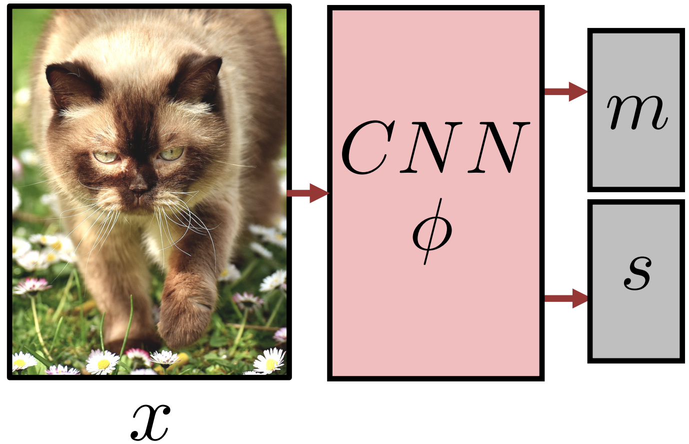
$$= \sum_i \mathbb{E}_{q_i} \underbrace{\log p(x_i \mid t_i, w)}_{-\|x_i - \mu(t_i)\|^2 + \text{const}} - \mathcal{KL}(q_i(t_i) \parallel p(t_i))$$

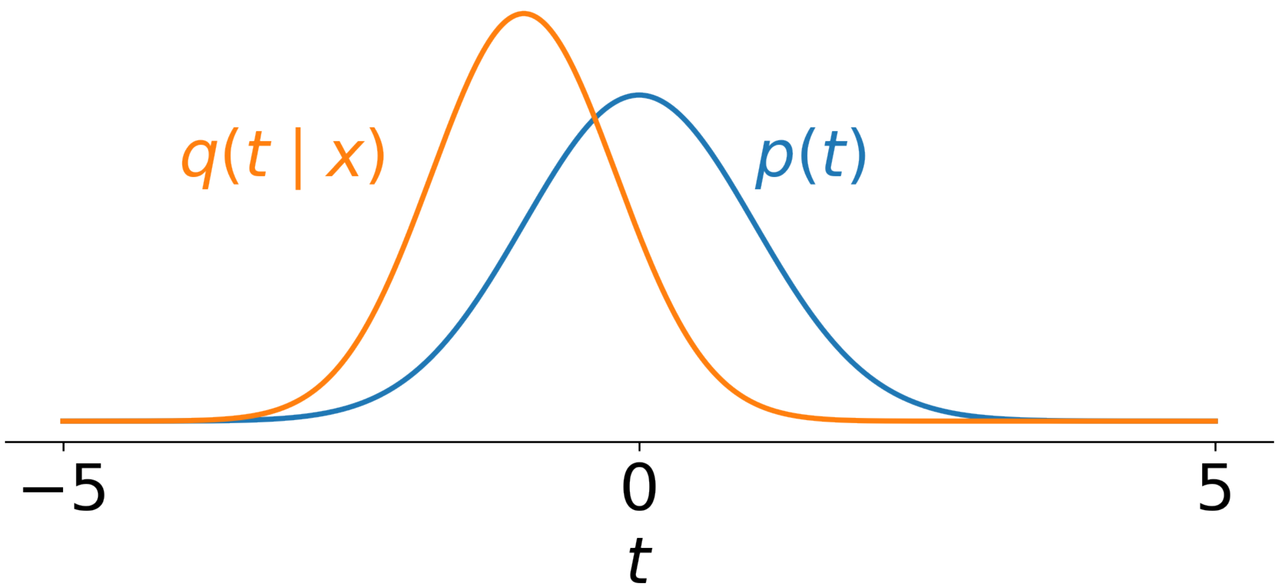
Reconstruction loss

Approximate posterior
$q(t_i) \approx p(t_i \mid x_i, w)$

Regularization

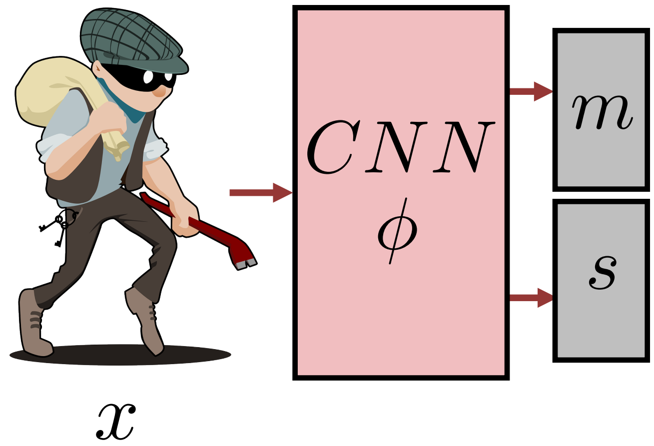
Detecting outliers



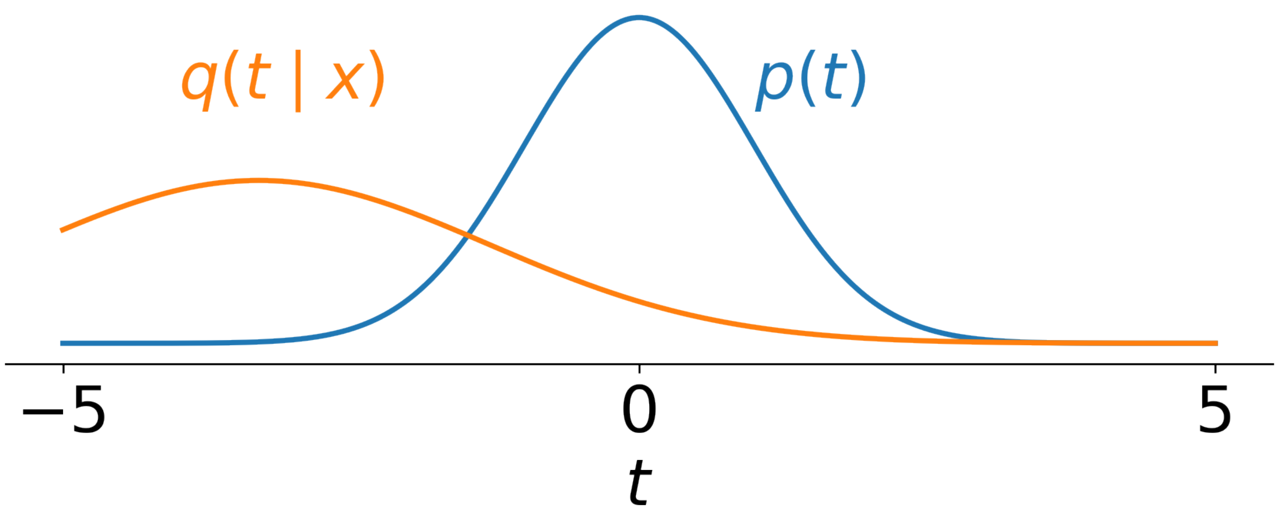


$$\mathcal{KL}(q(t \mid x) \,\|\, p(t)) \approx 0.54$$

Detecting outliers



For other methods to detect outliers see Supplementary



$$\mathcal{KL}(q(t \mid x) \,\|\, p(t)) \approx 6.25$$

Generating new samples

$$p(x \mid w) = \int p(x \mid t, w)\, p(t) dt$$

$$\widehat{t}_i \sim \mathcal{N}(0, I)$$