Continuous mixture of Gaussians

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

Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t) = \mathcal{N}(\mu(t), \Sigma(t))$$

Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t) = \mathcal{N}(\mu(t), \Sigma(t))$$

If $\mu(t) = Wt + b, \Sigma(t) = \Sigma_0$ get PPCA

(see week 2)

Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t) = \mathcal{N}(\mu(t), \Sigma(t))$$

If $\mu(t) = Wt + b, \Sigma(t) = \Sigma_0$ get PPCA

(see week 2)

But if x is image, why not

$$\mu(t) = \mathrm{CNN}_1(t)$$

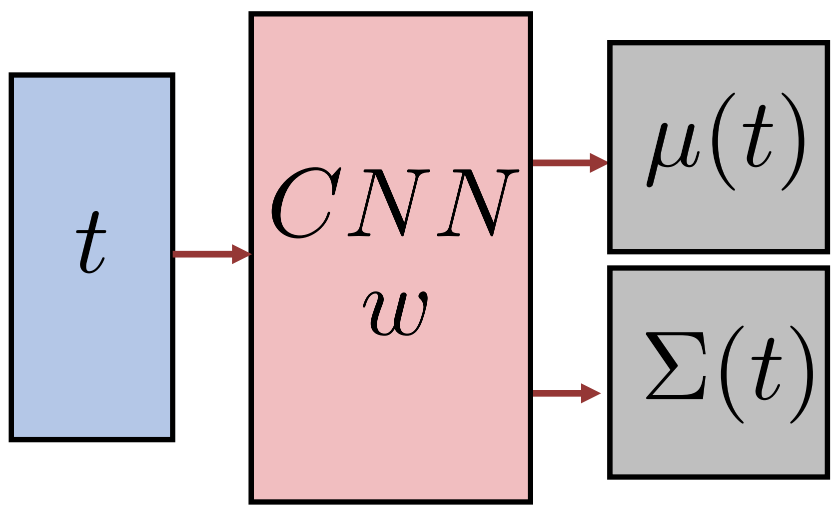
$$\Sigma(t) = \mathrm{CNN}_2(t)$$

Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t) = \mathcal{N}(\mu(t), \Sigma(t))$$

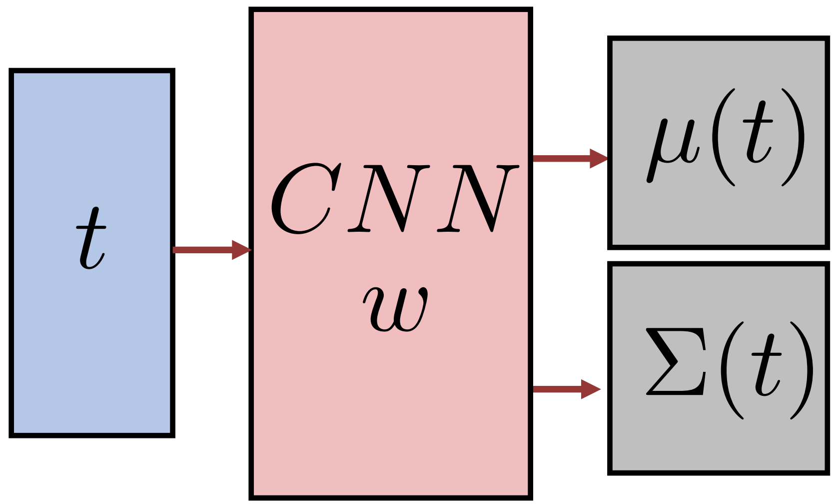


Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), \Sigma(t, w))$$

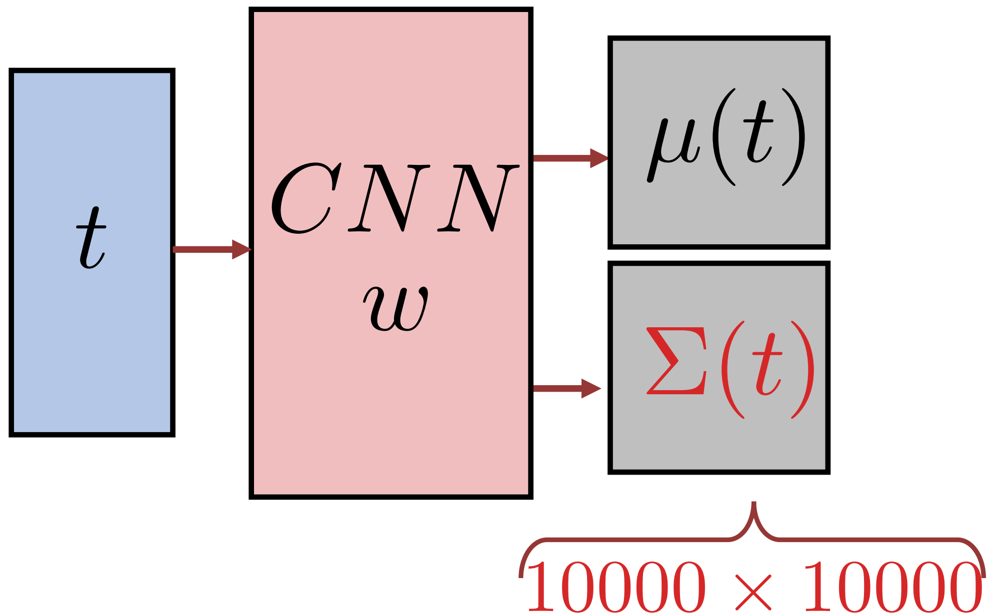


Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), \Sigma(t, w))$$

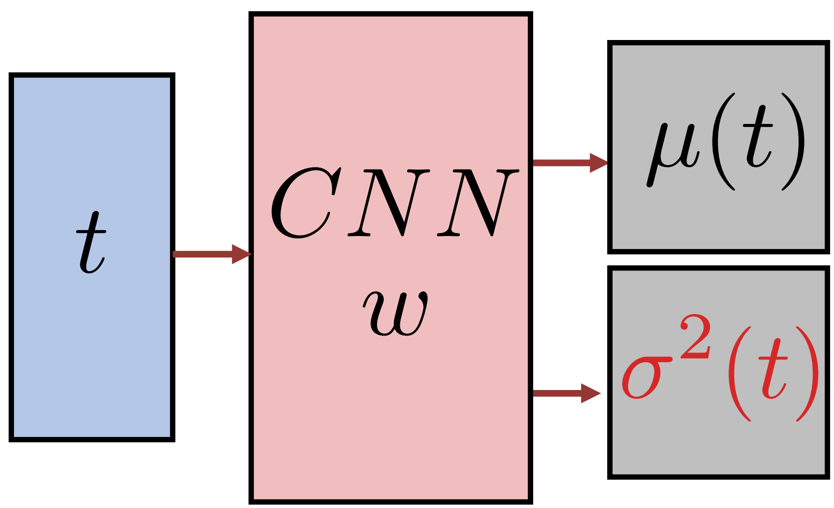


Continuous mixture of Gaussians

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

$$p(t) = \mathcal{N}(0, I)$$

$$p(x \mid t, w) = \mathcal{N}(\mu(t, w), \mathrm{diag}(\sigma^2(t, w)))$$



Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM!

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

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

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

Need to compute $p(T \mid X, w)$

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

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

Scaling up Expectation Maximization

$$\max_w . \; p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC?

Scaling up Expectation Maximization

$$\max_{w} . \; p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC?

$$\mathbb{E}_q \log p(X, T \mid w) \approx \frac{1}{M} \sum_{s=1}^{M} \log p(X, T_s \mid w)$$

$$T_s \sim q(T)$$

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC? An option, but we can do better

$$\mathbb{E}_q \log p(X, T \mid w) \approx \frac{1}{M} \sum_{s=1}^{M} \log p(X, T_s \mid w)$$

$$T_s \sim q(T)$$

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC? An option, but we can do better

Then Variational EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

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

Scaling up Expectation Maximization

$$\max_{w} . \; p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC? An option, but we can do better

Then Variational EM!

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

$$\begin{aligned} \underset{w,q}{\text{maximize}} \quad & \mathcal{L}(w, q) \\ \text{subject to} \quad & q_i(t_i) = \widetilde{q}(t_{i1}) \ldots \widetilde{q}(t_{im}) \end{aligned}$$

Scaling up Expectation Maximization

$$\max_{w}.\ p(X \mid w) = \int p(X \mid T, w)\, p(T) dt$$

Latent Variable model — use EM! But E-step is intractable

MCMC? An option, but we can do better

Then Variational EM! But again intractable.

$$\log p(X \mid w) \geq \mathcal{L}(w, q)$$

$$\begin{aligned} \underset{w,q}{\text{maximize}} \quad & \mathcal{L}(w, q) \\ \text{subject to} \quad & q_i(t_i) = \widetilde{q}(t_{i1}) \dots \widetilde{q}(t_{im}) \end{aligned}$$