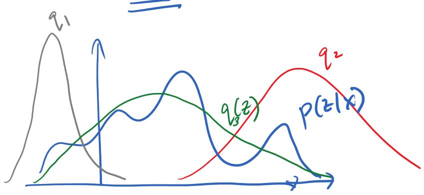
Tutorial 8

Thursday, 23 March 2023 6:21 PM

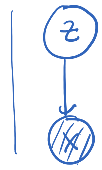
Variational inference

Key idea: $p(z|x)$

inference $\longrightarrow$ optimization.



Example



$p(z)$

θ "fixed"

$p(x|z)$

$\int p(x|z)\,p(z)\,dz$

"Distance" - Divergence btw p and q.

$$\mathbb{D}_{KL}[q\|p] = -\int q(z) \lg \frac{p(z)}{q(z)} dz$$

$KL(q,p)$

$$= \int q(z) \lg \frac{q(z)}{p(z)} dz$$

$$q^* = \underset{q \in Q}{\arg\min}\ \mathbb{D}_{KL}[q \| p(z|x)]$$

$$\phi^* = \underset{\phi = \{\mu, \Sigma\} \in \Phi}{\arg\min}\ \mathbb{D}_{KL}[q_\phi \| p(z|x)] \Leftarrow \text{how to compute.}$$

Recall EM.

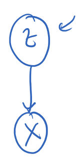
$$\lg p(x) = \mathcal{L}(q) + \mathbb{D}_{KL}[q \| \overset{p(z|x)}{p}]$$

const. C wrt. q.

$$\mathcal{L}(q) = \int q(z) \lg \frac{p(x,z)}{q(z)} dz.$$

$$\mathbb{D}_{KL}[q\|p] = \underbrace{-\mathcal{L}(q)}_{\text{maximize } \mathcal{L}(q)} + C.$$

$$\mathcal{L}(q) = \int q(z) \lg \frac{p(x,z)}{q(z)} dz.$$



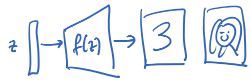
Example:

$$= \int q(z) \lg p(x|z)\,dz + \int q(z) \lg \frac{p(z)}{q(z)} dz.$$

$$= \mathbb{E}_{q(z)}[\lg p(x|z)] - \mathbb{D}_{KL}[q(z) \| \overset{\text{prior}}{p(z)}].$$

---

learning with Variational Inf



z $\quad p(z) = N(0, I)$

$\downarrow$

x $\quad p(x|z) = p(x \mid \overset{\text{deep NN.}}{f_\theta(z)})$

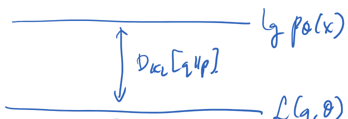
θ is not fixed

learn θ via MLE.

$$\theta^* = \underset{\theta}{\arg\max}\ \lg p_\theta(x)$$

$$\lg p_\theta(x) = \mathcal{L}(q,\theta) + \mathbb{D}_{KL}[q(z) \,\|\, p_\theta(z|x)].$$

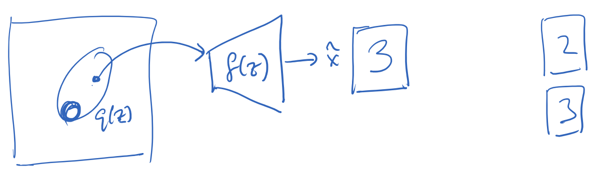
$$\mathcal{L}(q,\theta) = \underbrace{\boxed{\lg p_\theta(X)}}_{\text{evidence.}} - \underbrace{\mathbb{D}_{KL}[q(z) \,\|\, p_\theta(z|x)]}_{} \Leftarrow$$



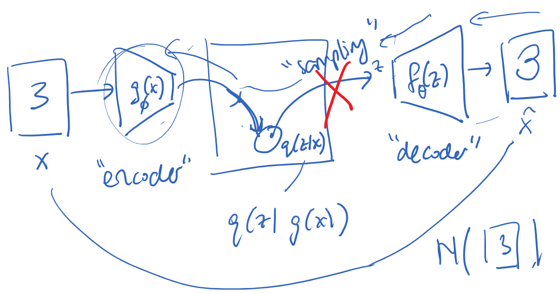
$$\mathcal{L}(q,\theta) = \underbrace{\boxed{\mathbb{E}_{q(z)}[\lg \underbrace{p(X|z)}_{f(z)}]}}_{\text{sampling.}} - \underbrace{\mathbb{D}_{KL}[q(z)\|p(z)]}_{\text{easy}}$$

Pick $q(z)$:

- $q(z) = \mathcal{N}(\mu, \Sigma).$



- $q(z) = \mathcal{N}(z \mid g(x)).$ → $g^{\mu}_{\phi}(x)$, $g^{\Sigma}_{\phi}(x)$.



$$\mathcal{N}(\boxed{3}, \boxed{3}) >> 0$$

$$\lg \mathcal{N} = c + (x - f(z))^T \Sigma^{-1} (x - f(z))$$

$$\underset{\theta,\phi}{\text{maximize}}\ \mathcal{L}(\phi,\theta) = \underbrace{\mathbb{E}_{q_\phi}[\lg p_\theta(x|z)]}_{\text{sampling.}} - \underbrace{\mathbb{D}_{KL}[q_\phi(z|x) \,\|\, p(z)]}_{\text{closed form}}.$$

gradient ascent

$$\phi_{t+1} = \phi_t + \eta \nabla_\phi \mathcal{L}(\phi,\theta)$$

$$\theta_{t+1} = \theta_t + \eta \nabla_\theta \mathcal{L}(\phi,\theta).$$

Problem ② no gradients!

Problem (2) no gradients! :(

$z \sim N(\mu, \Sigma)$

$\epsilon \sim N(0, I)$

$\mu + \epsilon$

idea: don't sample z directly.

sample $\epsilon \sim N(0, I)$

$$z = g_\phi^\mu(x) + \left(g_\phi^\Sigma(x)\right)^{1/2} \epsilon .$$