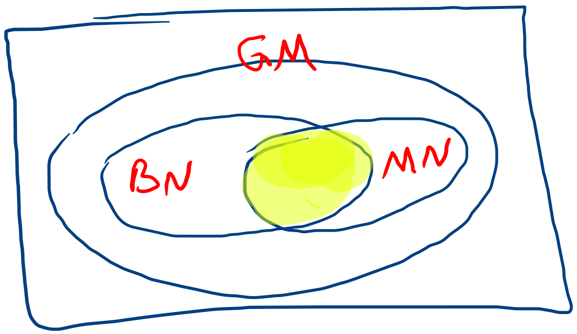
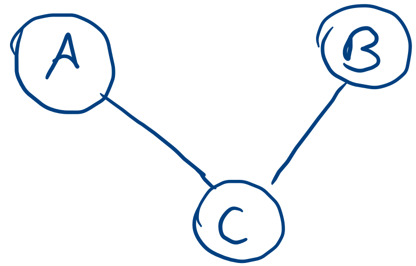
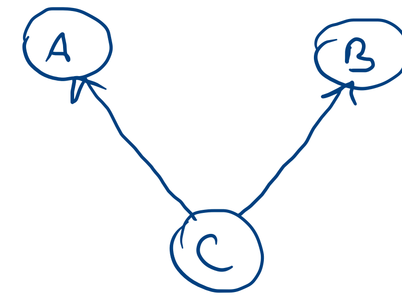


MN

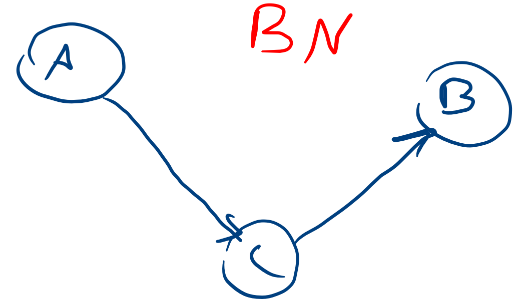


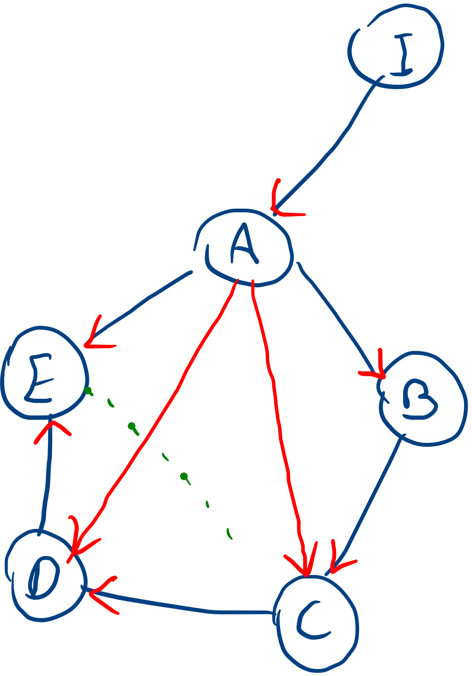
BN

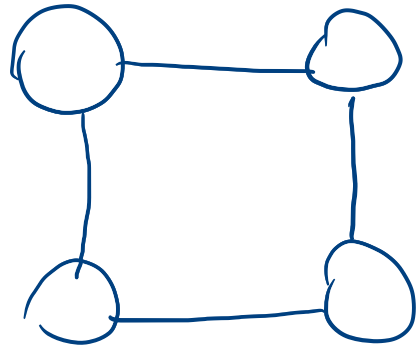


$$A \perp B \mid C$$

BN







MN

Variational Inference

$S = x \cup z$

x → observed

z → Latent/hidden

$$\boxed{p(z|x) = ?}$$

$$\boxed{q(z) \simeq p(z|x)}$$

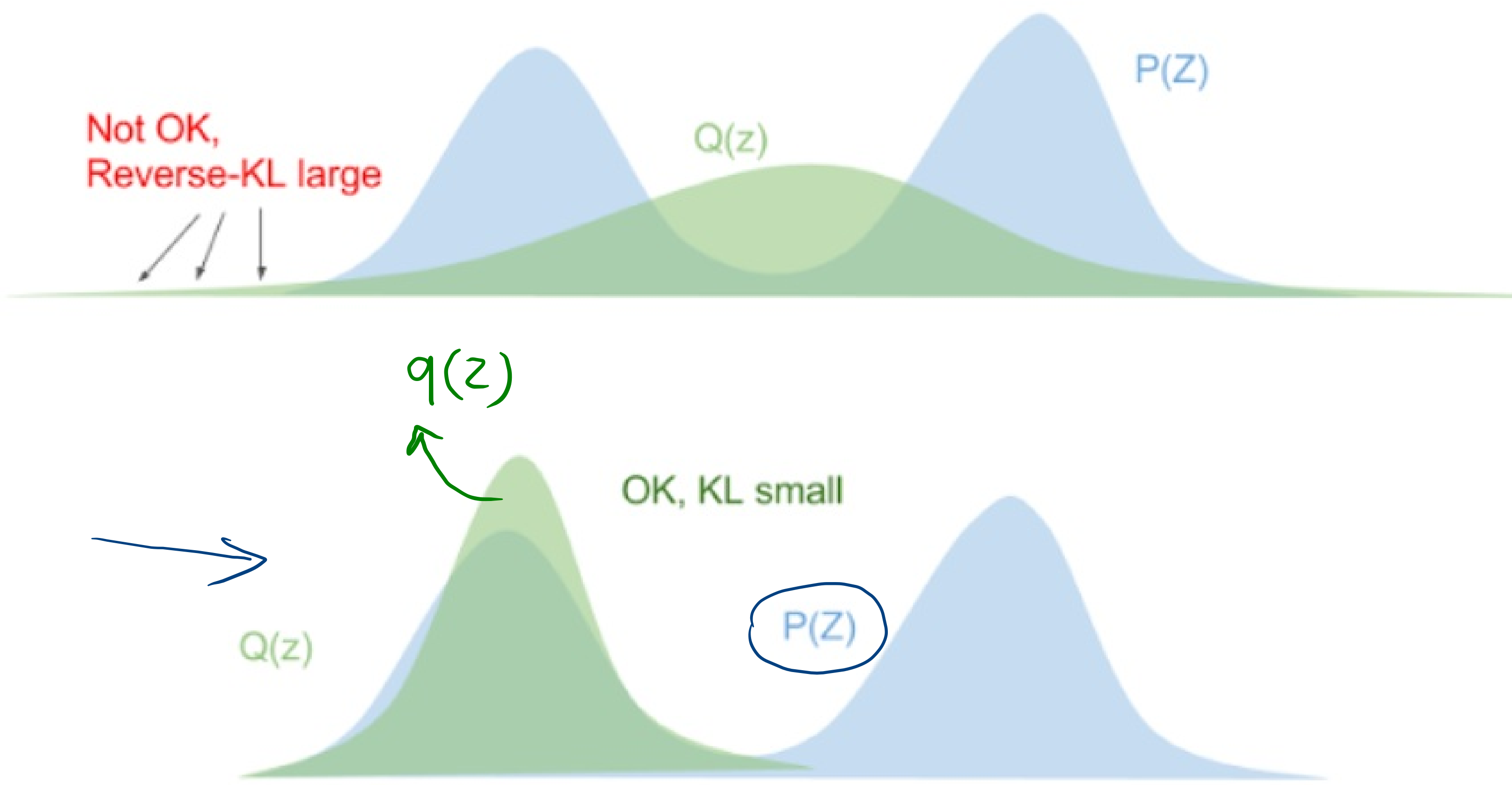
$$\boxed{q^*(z) = \underset{q \in Q}{\arg\min}\ KL\big(q(z) \,\|\, p(z|x)\big)}$$

$$\boxed{q^*(z) = \underset{q \in Q}{\arg\min}\ KL\big(p(z|x) \,\|\, q(z)\big)}$$

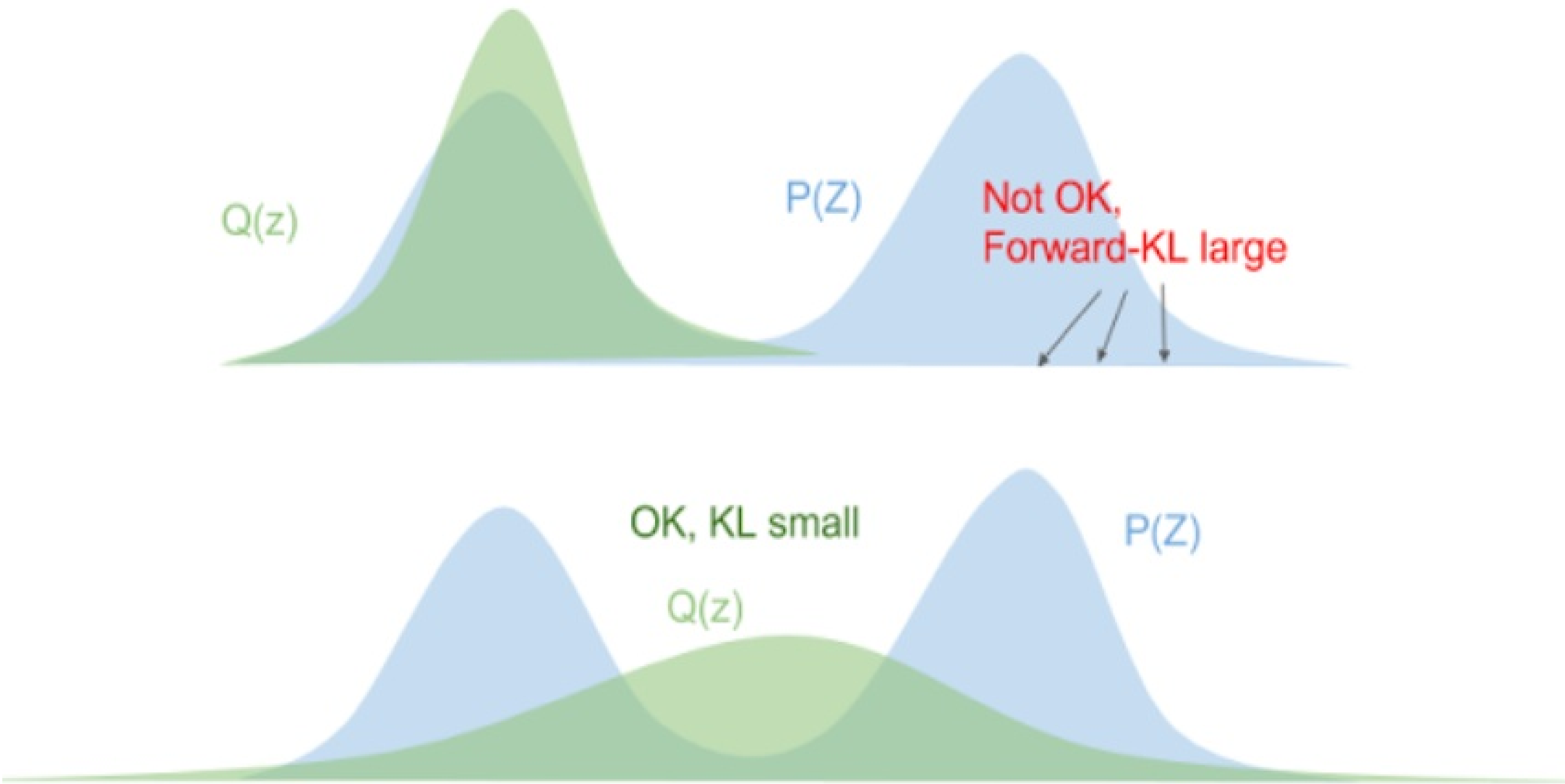
$$\rightarrow KL(p \| q) = \sum_z \underline{\underline{p(z)}} \log \frac{p(z)}{q(z)} = E_p\left[\log \frac{p(z)}{q(z)}\right]$$

$$\rightarrow KL(q \| p) = \sum_z q(z) \log \frac{q(z)}{p(z)} = E_q\left[\log \frac{q(z)}{p(z)}\right]$$

$$\min_q \; KL(q \,||\, p)$$



KL(p || q)



$$KL(q \| p) = \sum_z q(z) \log \frac{q(z)}{p(z)}$$

$$KL(p \| q) = \sum_z p(z) \log \frac{p(z)}{q(z)}$$

$KL(p \| q)$ $KL(q \| p)$

KL divergence: M-projection vs. I-projection

- Let P is <u>mixture</u> of two 2D Gaussians and Q be a 2D Gaussian distribution with arbitrary covariance matrix:

$$Q^* = \underset{Q}{\operatorname{argmin}} \int P(\mathbf{z}) \log \frac{P(\mathbf{z})}{Q(\mathbf{z})} d\mathbf{z}$$

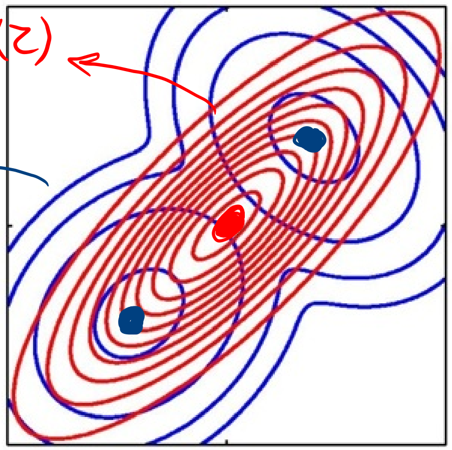
P: Blue
Q^*: Red

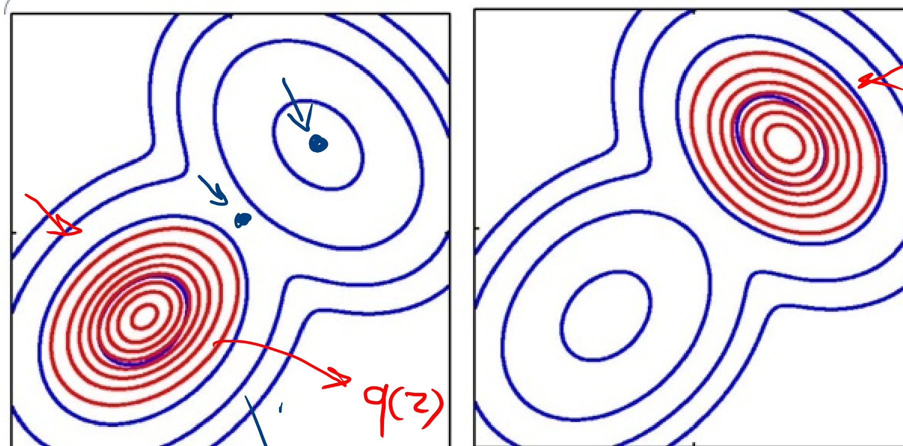
$$Q^* = \underset{Q}{\operatorname{argmin}} \int Q(\mathbf{z}) \log \frac{Q(\mathbf{z})}{P(\mathbf{z})} d\mathbf{z}$$

$q \in Q$

$q(z)$

$p(z)$





$q(z)$

$p(z)$

$\min_q KL(p \| q) = \sum p \log \frac{p}{q}$

$KL(q \| p) = \sum q \log \frac{q}{p}$

$$JSD(p,q) = \frac{1}{2} KL\left(p \,\|\, \frac{p+q}{2}\right) + \frac{1}{2} KL\left(q \,\|\, \frac{p+q}{2}\right)$$

$$\min_{q \in Q} KL(q(z) \| p(z|x)) \qquad p(x,z)$$

$$KL(q(z) \| p(z|x)) = E_q\left[\log \frac{q(z)}{p(z|x)}\right] = E_q\left[\log \frac{q(z)\,p(x)}{p(x,z)}\right]$$

$$p(z|x) = \frac{p(x,z)}{p(x)}$$

$$= E_q\left[\log \frac{q(z)}{p(x,z)} + \log p(x)\right] = E_q\left[\log \frac{q(z)}{p(x,z)}\right] + E_q\left[\log p(x)\right]$$

$$= KL(q(z) \| p(x,z)) + \sum_z q(z) \log p(x)$$

$$= KL(q(z) \| p(x,z)) + \log p(x)$$

$$KL\big(q(z)\,\|\,p(z|x)\big) = KL\big(q(z)\,\|\,p(x,z)\big) + \log p(x)$$

$$\Rightarrow \underbrace{\log p(x)}_{\text{const.}} = \underbrace{-KL\big(q(z)\,\|\,p(x,z)\big)}_{\uparrow} + \underbrace{KL\big(q(z)\,\|\,p(z|x)\big)}_{\min\,\downarrow}$$

$$q^{*}(z) = \arg\max_{q\in Q}\; -KL\big(q(z)\,\|\,p(x,z)\big)$$

$$= \arg\min_{q\in Q}\; KL\big(q(z)\,\|\,\underline{p(x,z)}\big)$$

$$\underbrace{\log p(x)}_{\text{evidence}} = \underbrace{-KL\big(q(z)\,\|\,p(x,z)\big)}_{\uparrow} + \underbrace{KL\big(q(z)\,\|\,p(z|x)\big)}_{\geq 0}$$

$$\boxed{G_\theta}$$

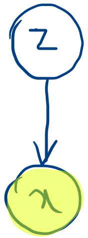
$$p(x) = \sum_z p(x,z)$$

$$\boxed{\hat{\theta}_{ML} = \arg\max_\theta \log \underline{p(x)}}$$

$$\underbrace{-KL\left(q_\theta(z) \,\|\, p(x,z)\right)}_{\text{Evidence lower Bound (ELBO)}} \leq \underbrace{\log p_\theta(x)}_{\text{Evidence}} \qquad \forall\, \theta$$

$$\boxed{\theta}$$

$$\theta^* = \arg\max_\theta \; -KL\left(q(z) \,\|\, p(x,z)\right)$$



$$p(z) = e^{-z} \qquad z \geq 0$$

$$p(x|z) = \mathcal{N}(x|z,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-z)^2}$$

$$p(x,z) = p(z)\,p(x|z)$$

$$\boxed{p(z|x) = ?}$$

$$p(z|x) \simeq q(z)$$

$$\boxed{q(z) = \mathcal{N}(z|\mu,1)}$$

$$q(z) = \mathcal{N}(z|\mu,\sigma^2)$$

$$\mu^* = \arg\min_{\mu} KL\left(q(z)\,\|\,p(x,z)\right)$$

$$KL(q(z) \,\|\, p(x,z)) = E_q\left[\log \frac{q(z)}{p(x,z)}\right] = E_q\left[\log \frac{\cancel{\frac{1}{\sqrt{2\pi}}}\, e^{-\frac{1}{2}(z-\mu)^2}}{e^{-z}\, \cancel{\frac{1}{\sqrt{2\pi}}}\, e^{-\frac{1}{2}(x-z)^2}}\right]$$

$$p(x,z) \swarrow \; p(z)p(x|z)$$

$$= E_q\left[-\frac{1}{2}(z-\mu)^2 + z + \frac{1}{2}(x-z)^2\right]$$

$$= E_q\left[\cancel{-\frac{1}{2}z^2} - \underbrace{\frac{1}{2}\mu^2}_{} + z\mu + z + \frac{1}{2}x^2 + \cancel{\frac{1}{2}z^2} - xz\right]$$

$$= -\frac{1}{2}\mu^2 + \mu E_q[z] + E_q[z] + \frac{1}{2}x^2 - x E_q[z]$$

$$\boxed{= -\frac{1}{2}\mu^2 + \mu^2 + \mu + \frac{1}{2}x^2 - x\mu}$$

$$\frac{dKL}{d\mu} = 0 \Rightarrow \mu + 1 - x = 0 \Rightarrow \boxed{\mu^* = x - 1}$$