General Idea:

$$z_i \longrightarrow \boxed{\ \theta\ } \longrightarrow x_i$$

$P(z)$ $\quad P_\theta(x|z)$

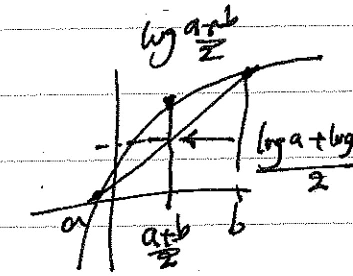
ML:

$$\max_\theta P_\theta(X) = P_\theta(x_1) \cdots P_\theta(x_N)$$

$$\max_\theta \log P_\theta(X) = \sum_i \log P_\theta(x_i)$$

$$\log P(x) = \log \int P(x,z)\,dz$$

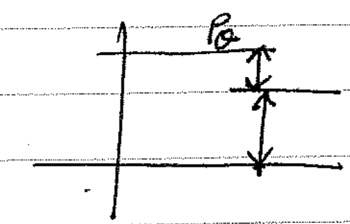
$$= \log \int P(x|z)\,P(z)\,dz$$



$$\geq \int \log\left[P(x|z)\,P(z)\right] dz$$

$$\frac{1}{2}(\log a + \log b) \leq \log \frac{a+b}{2}$$

$$\geq \underbrace{\int \log \frac{P_\theta(x|z)\,P(z)}{q(z)}\, q(z)\,dz}_{\mathbb{E}_q \log \frac{P_\theta(x,z)}{q_\psi(z)}}$$



$$\max_{\theta,\psi} \mathbb{E} \log \frac{P_\theta(x,z)}{q_\psi(z)} \qquad (\text{ELBO})$$

pe2

Another way of saying the same thing:

$X = \{x_1, \ldots, x_N\}$

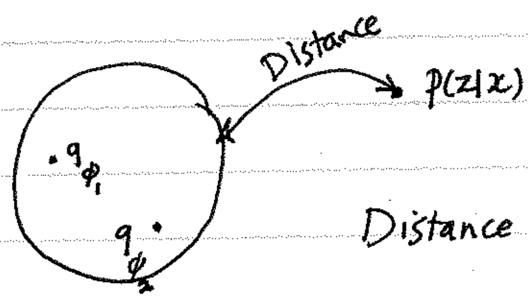
$P(Z|X)$ ← interested

$$\underbrace{P(Z|x)}_{\text{Posterior}} = \frac{P(x,z)}{P(x)} = \frac{\overbrace{P(x|Z)}^{\text{liklihood}}\overbrace{P(Z)}^{\text{prior}}}{\underbrace{P(x)}_{\text{nomalizer}}}$$

$$= \frac{P(x|Z)\,P(Z)}{\underbrace{\int P(x,z)\,dz}_{\text{difficult}}}$$

idea:

$q_\phi(z)$ $\quad \min \text{Distance}\left(q_\phi(z),\ p(z|x)\right)$



Distance $\equiv$ KL

$$KL\left(q_\phi(z),\ P(z|x)\right) = \cancel{\mathbb{E}_{q_\phi}\left[\log \frac{q(z|x)}{p_\phi(z)}\right]} = \mathbb{E}_{q_\phi}\left[\log \frac{q_\phi(z)}{P(z|x)}\right]$$

$$\cancel{= \mathbb{E}\left[\log P(z|x)\right]}$$

$$= \mathbb{E}_{q_\phi}\left[\log \frac{q_\phi(z)\,P(x)}{P(x,z)}\right] = \mathbb{E}_{q_\phi}\log \frac{q_\phi(z)}{P(x,z)} + \log P(x)$$

$$\log P(X) = \underbrace{KL\left(q_\phi(z), p(z|X)\right)}_{\geq 0} - \mathbb{E}_{q_\phi} \log \frac{q_\phi(z)}{P(z,X)}$$

$$\log P(X) \geq \mathbb{E}_q \log \frac{P(z,X)}{q_\phi(z)} \qquad (ELBO)$$

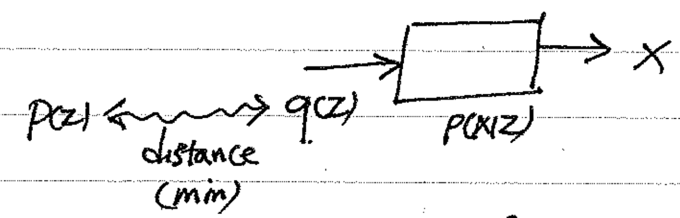
Expand ELBO:

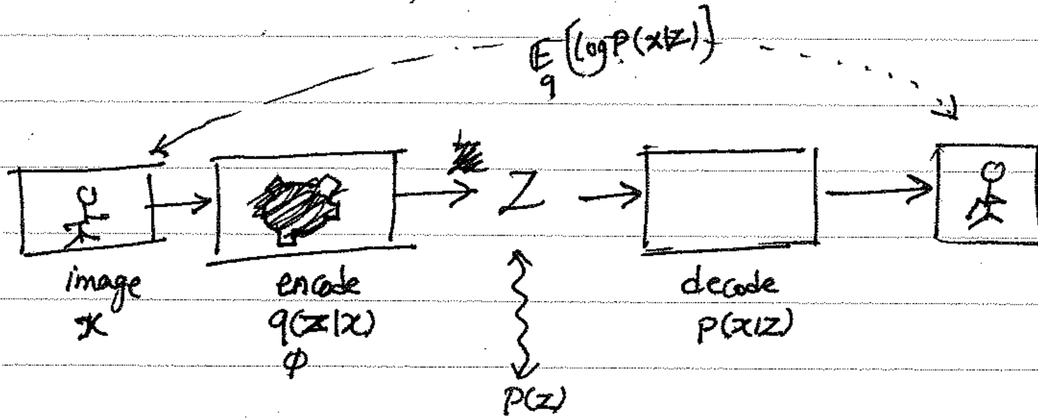
$$\mathbb{E}_q \log \frac{P(z,X)}{q_\phi(z)} = \mathbb{E}_q \log \frac{P(X|z)}{q_\phi(z)} + \mathbb{E}_q \frac{P(z)}{q_\phi(z)}$$

$$= \underbrace{\mathbb{E}_q \log P(X|z)}_{\text{decode -- reconstruction as if } q \text{ is generating}} - \underbrace{KL\left(q_\phi(z) \,\|\, P(z)\right)}_{\text{distance to prior}}$$

decode -- reconstruction
as if q
is generating

distance to prior





Challenges:

We have to find best ϕ and θ that max

$$\max_{\phi,\theta} -D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|x^{i})}\left[\log p_\theta(x^{i}|z)\right]$$

$$\max_{\phi,\theta} - \mathbb{E}_q \log \frac{q(z|x)}{p_\theta(z)} + \mathbb{E}_q\left[\log p_\theta(x^{(i)}|z)\right]$$

∇_θ : easy!

∇_ϕ : No so much!

Let look closer:

$$\mathbb{E}_{q_\phi}\left[f(z)\right] = \int f(z)\, q_\phi(z)\, dz$$

$$\ast\ \nabla_\phi\left(\,''\,\right)$$

$$= \int \nabla f(z)\, q_\phi(z)\, dz = \underbrace{\int f(z)\, \nabla_\phi q_\phi(z)\, dz}_{\text{99\% you cannot do closed form}}$$

$$= \int f(z)\, \frac{\nabla_\phi q_\phi}{q_\phi}\, q_\phi\, dz \quad \ast$$

$$\frac{\nabla_\phi q_\phi}{q_\phi} = \nabla_\phi \log q_\phi$$

$$\circledast \quad \int f(z) \nabla \log q_\phi \; q_\phi \, dz = \mathbb{E}\left[f(z) \nabla \log q_\phi \right]$$

$$\simeq \frac{1}{L} \sum f(z_i) \nabla \log q_\phi(z)$$

(We know how to estimate but it has large variance!)

Remember ELBO:

$$\mathcal{L} = \int \log \frac{P(x,z)}{q(z|x)}\, q(z|x)\, dz = \mathbb{E}\left[\log P(x,z) - q(z|x)\right]$$

$$\simeq \frac{1}{L} \sum_{l} \log(P(x,z^l) - q(z^l|x)) \qquad (I)$$

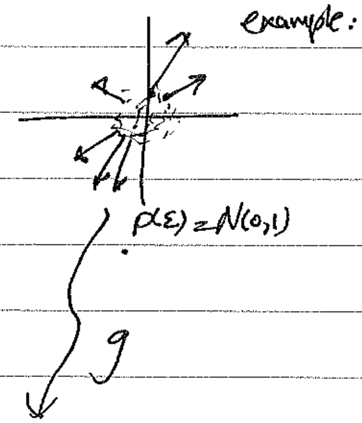
s.t. $\underbrace{x^l \sim q(z|x)}$

$\downarrow$ How to generate this?

Sample $\varepsilon^l \sim p(\varepsilon)$

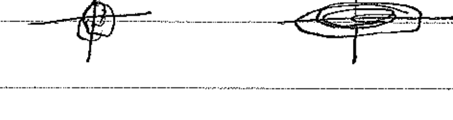
$z^l = g(\varepsilon^l, x)$

$\uparrow$ a deterministic function



A. Let's look at example of g

$$\underset{2D}{g}(\varepsilon) = \begin{bmatrix} \cancel{20}\ 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$



~~Approx~~ Estimation of Eq (I) is fine but can have high variance

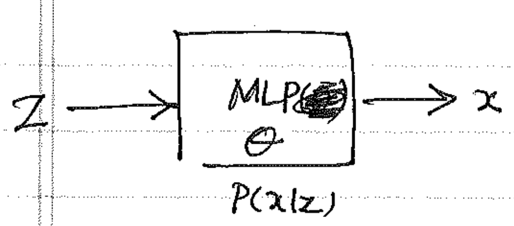
Let's compute anything that can be computed closed form!

B. $z = g(\varepsilon) = \mu + \sigma \varepsilon$. What is $p(z) = ?$

C. $z = g(\varepsilon) = \mu + A\varepsilon$ " " $p(z) = ?$

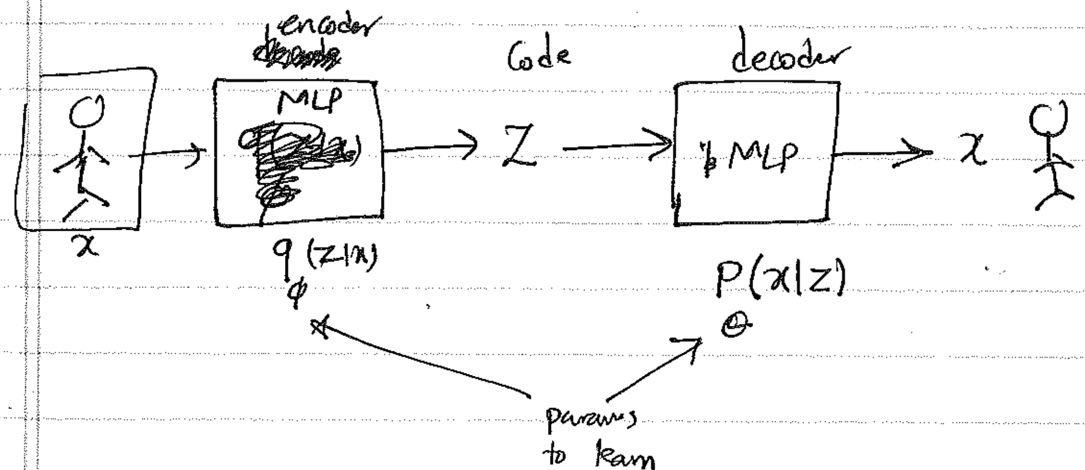
Page 7

Our machine



$P(Z|x)$ almost impossible to compute closed form

Remember decoder-encoder



How ~~we~~ we are going to learn ϕ, θ ?

- We set ϕ to something, we pass the guy to $q_\phi(Z|x)$ & how? remember $Z = g_\phi(x, \varepsilon)$!

  Sample ε, subst. ε, x get Z !

- Pass Z to decoder get new $\tilde{x}$.

- Is $x, \tilde{x}$ similar? $\mathbb{E}_z \log P(x|z) \nearrow$

~~log~~ ~~logP~~ example ~~B~~, $\log P(x|z) = \mu(x;\theta) + A(x;\theta)Z$