FYS5429, lecture April 23, 2024

FYS 5429/9429, APRIL 23, 2024

Traditional autoencoder

Encoder — Latent variables — Decoder

$$x \rightarrow \boxed{f} \rightarrow h \rightarrow \boxed{g} \rightarrow \tilde{x}$$

$$h = f(x;\phi)$$

$$\tilde{x} = g(h;\theta)$$

$$\underset{\phi,\theta}{\arg\min} \, \|(x-\tilde{x})\|_2^2$$

Interested $p(x;\theta) = p_\theta(x)$

marginal probability,

True probability $p^*(x)$

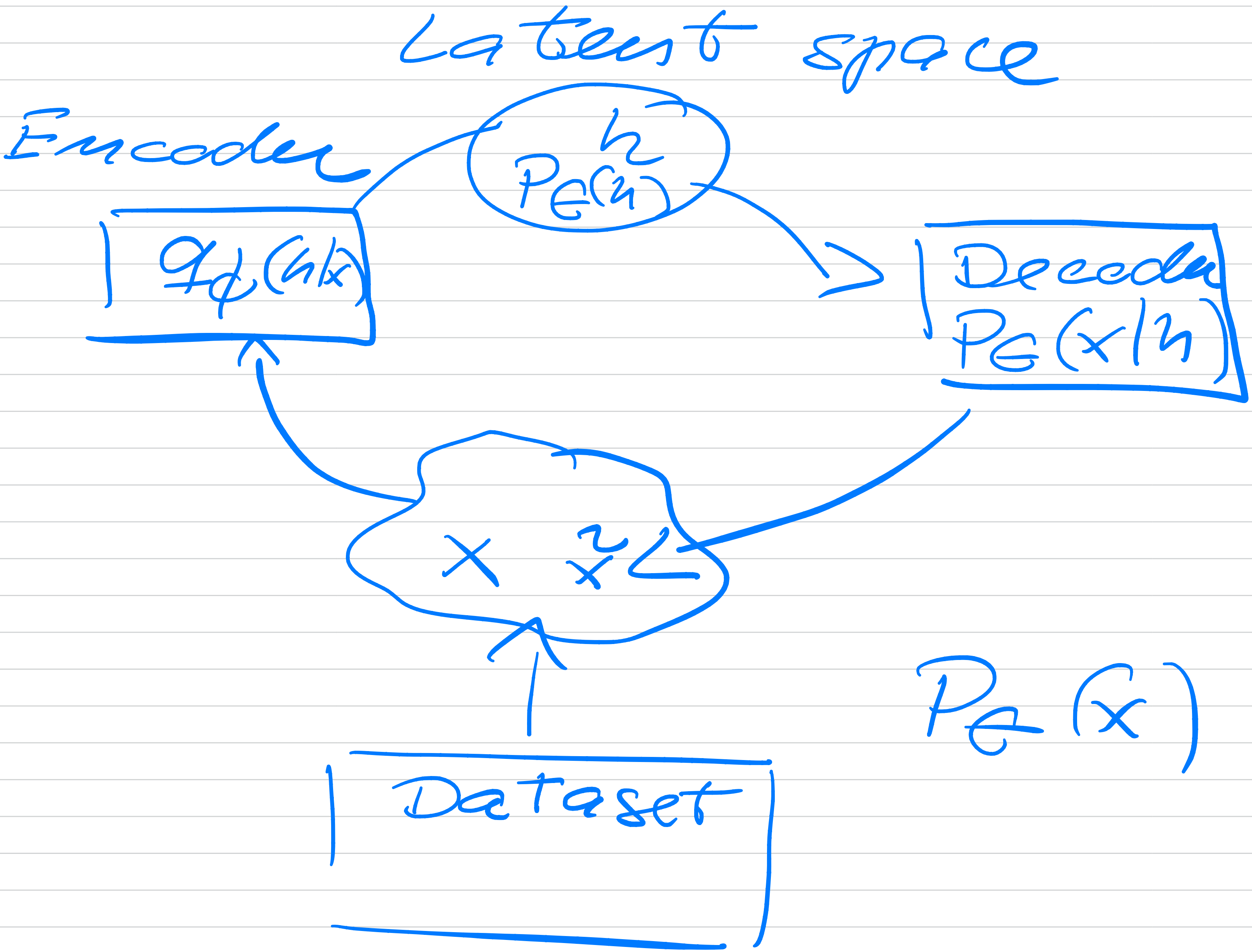
$$p_\theta(x) \simeq p^*(x)$$

we are often interested in the conditional probability

$$p_\theta(y|x) \quad \text{or} \quad p_\theta(x|y)$$

in our case $y \Rightarrow h$ = latent variables

Basic structure of VAE



Bayes' theorem

$$p(x, h) = \underset{\text{likelihood}}{p(x|h)}\,\underset{\text{prior}}{p(h)}$$

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

Bayes' theorem combines these two equations

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

$$p_\theta(x) = \underbrace{\int p_\theta(x,h)\,dh}$$

$$= \int p_\theta(x|h)\, p_\theta(h)\, dh$$

$$p_\theta(h|x) = \frac{p_\theta(x,h)}{p_\theta(x)}$$

A VAE learns stochastic mapping between an observed x-space whose distribution is $\boxed{q_\phi(h|x)}$ and a

Latent space h whose distribution ($p_\theta(h)$) can be simple.

$$p_\theta(x, h) = p_\theta(x|h)\boxed{p_\theta(h)}$$

Need a model

$$\boxed{q_\phi(h|x)} \simeq p_\phi(h|x)$$

From last time

we want to optimize

$$\arg\max_{\theta, \phi} \log P_\theta(x)$$

$$x \sim P_\theta(x)$$

$$\log P_\theta(x) = \mathbb{E}_{q_\phi(z|x)}\left[\log P_\theta(x)\right]$$

$$\log P_\theta(x) = \log\left[\int P_\theta(x|z) P_\theta(z)\, dz\right]$$

$$= \log\left[\int P_\theta(x, z)\, dz\right]$$

insert $1 = \dfrac{q_\phi(h|x)}{q_\phi(h|x)}$

$$\log p_\theta(x) = \log\left[\int \frac{p_\theta(x,h)\, q_\phi(h|x)}{q_\phi(h|x)}\, dh\right]$$

$$= E_{q_\phi(h|x)}\left[\log \frac{p_\theta(x,h)}{q_\theta(h|x)}\right]$$

$$= E_{q_\phi(h|x)}\left[\log\left[\frac{p_\theta(x,h)}{q_\phi(h|x)} \; \frac{q_\phi(h|x)}{p_\theta(h|x)}\right]\right]$$

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

$$+ \underbrace{\mathbb{E}_{q_\phi(z|x)}\left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)}\right]}_{\text{KL-divergence}}$$

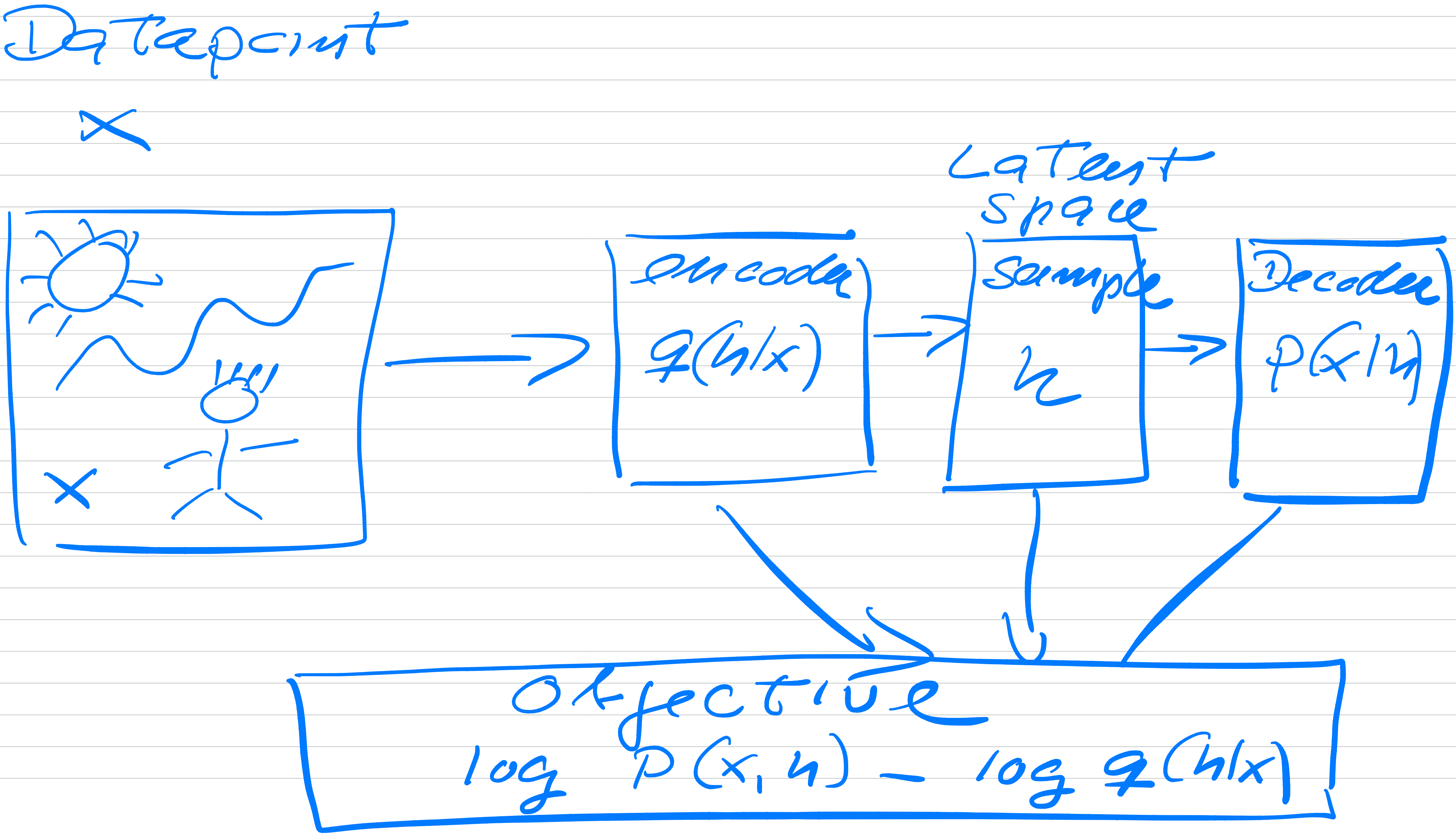
$$= KL\left(q_\phi(z|x) \,\|\, p_\theta(z|x)\right)$$

$$KL(q_\phi \,\|\, p_\theta) \geq 0$$

First term is normally called variational lower bound (Evidence lower bound, ELBO)

$$\mathcal{L}_{\theta\phi}(x) = \mathbb{E}_{q_\phi(x)}\left[\log \frac{p_\theta(x,h)}{p_\theta(x|h)\,p_\theta(h)} - \log q_\phi(h|x)\right]$$

Computational flow of VAE



$q_\phi(x|z)$ is an intermediate step and defines the encoder

$$\mathbb{E}_{q_\phi(z|x)}\left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\right]$$

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

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

$$= \mathbb{E}_{q_\phi}\left[\log p_\theta(x|h)\right]$$

$$- KL\left(q_\theta(h|x) \,\|\, p_\theta(h)\right)$$

$(MC)^2$ procedure.

How do we proceed with second term?

Common choice

$$q_\phi(h|x) = \mathcal{N}\left(h;\, \mu_\phi(x),\, \sigma_\phi^2(x)\,\mathbb{1}\right)$$

$$p_\theta(z) = \mathcal{N}(z; 0, 1)$$

variance

mean value

We can compute the KL-divergence analytically, and the first term (reconstruction) is estimated by Monte Carlo sampling

$$\underset{\theta,\phi}{\arg\max}\left(\mathbb{E}_{q_\phi(h|x)}\left[\log p_\theta(x|h)\right] - KL\left(q_\phi(h|x)\,\|\,p_\theta(h)\right)\right]$$

$$\simeq \underset{\theta,\phi}{\arg\max}\left\{\frac{1}{MCS}\sum_{i=1}^{MCS}\log\left(p_\theta(x|h^{(i)})\right) - KL\left(q_\phi(h|x)\,\|\,p_\theta(h)\right)\right\}$$

The latent variables $h^{(i)}$ are sampled from $q_\phi(h|x)$ for every observation x in the dataset.

Reparameterization trick

- rewrites a random variable as a deterministic function of a noise variable

$$x \sim \mathcal{N}(x; \mu, \sigma^2)$$

$$x = \mu + \sigma \cdot \varepsilon \qquad \varepsilon \sim \mathcal{N}(\varepsilon, 0, 1)$$

in a VAE:

$$z = \mu_\phi(x) + \sigma_\phi(x) \cdot \varepsilon$$

$$\varepsilon \sim \mathcal{N}(\varepsilon, 0, 1)$$