

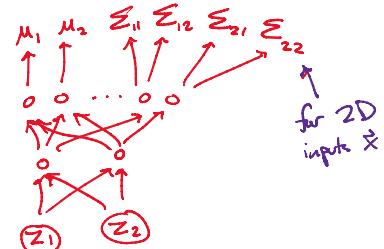
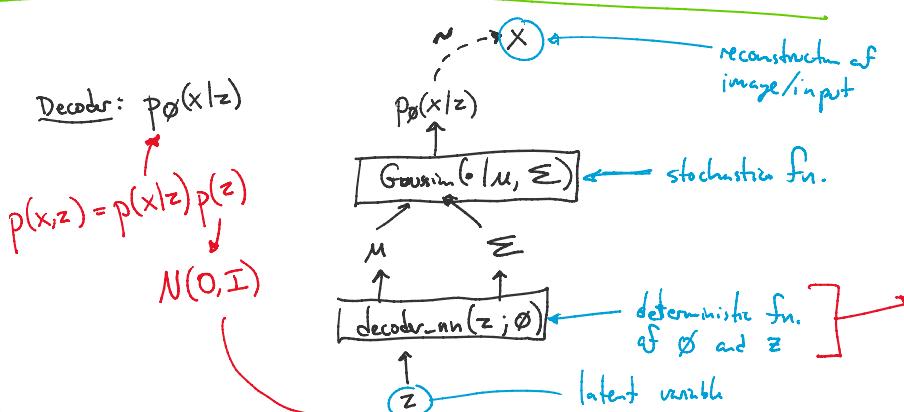
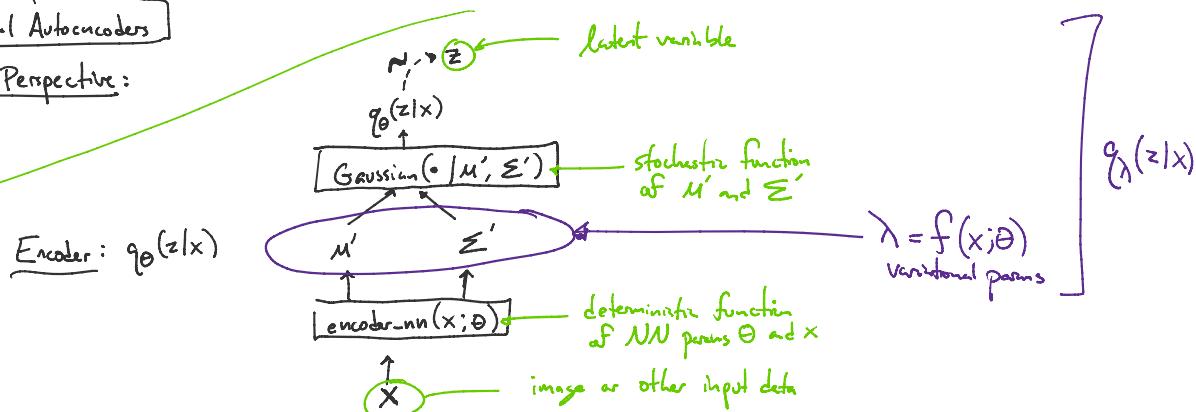
Q: Why not just sample from $p(z|x)$?
Why do we need $q(z|x)$ at all?

A: $p(x)$ is intractable

A: $p(x|z)$ is a NN , we don't know how to invert a NN

Variational Autoencoders

NN Perspective:



Dataset: $D = \{x^{(i)}\}_{i=1}^N$ unlabeled data (e.g. images)

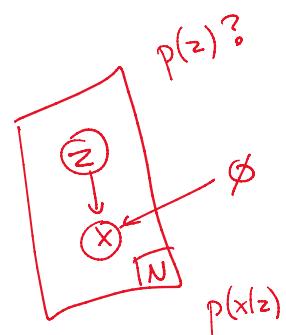
Loss Fn:

$$\ell(\theta, \phi) = \sum_{i=1}^N \ell_i(\theta, \phi)$$

$$\ell_i(\theta, \phi) = - \mathbb{E}_{q_\theta(z|x^{(i)})} [\log p_\phi(x^{(i)}|z)] + KL(q_\theta(z|x^{(i)}) || p(z))$$

reconstruction loss
(just like an autoencoder)

Q1: what does this term accomplish?
A: regularizer



Graphical Model Perspective:

$$\text{Model: } p_\theta(x, z) = p_\theta(x|z)p(z) = p_\theta(x|z)N(0, I)$$

$$\text{Goal: estimate } p_\theta(z|x) = \frac{p_\theta(x|z)p(z)}{p_\theta(x)}$$

intractable, why? b/c
 $p_\theta(x)$ is unknown
 $p_\theta(x|z)$ ~ low dim

$p_\theta(x)$ intractable, why? b/c
decoder- $n(z, \phi)$ is unknown
 $p(x|z)$ a complex dist.

Variational Approx: $q_\lambda(z|x)$ variational distribution
Variational parameters

Two Obj. Fn.s: $\text{ELBO}(\lambda) = E_{q_\lambda(z|x)} [\log p_\theta(x|z)] - E_{q_\lambda(z|x)} [\log q_\lambda(z|x)]$

where $z' = \{z^{(i)}\}_{i=1}^N$
 $x = \{x^{(i)}\}_{i=1}^N$

$$\begin{aligned} &= E_{q_\lambda} [\log p(x|z) + \log p(z) - \log q_\lambda(z'|x)] \\ &= E_{q_\lambda} [\log p(x|z)] - E_{q_\lambda} [\log (q_\lambda(z'|x)/p(z))] \\ &= E_{q_\lambda} [\log p(x|z)] - KL(q_\lambda(z'|x) || p(z)) \\ &= \sum_{i=1}^N \text{ELBO}_i(\lambda) \end{aligned}$$

$$\text{ELBO}_i(\lambda) = E_{q_\lambda(z|x^{(i)})} [\log p_\theta(x^{(i)}|z)] - KL(q_\lambda(z|x^{(i)}) || p(z))$$

given some algebra, we recover the "NN perspective" loss

Recall: Encoder $\lambda = [u', \varepsilon'] = \text{encoder_nn}(x; \theta)$ and $q_\theta(z|x, \lambda) = f(x, \theta)$

Decoder $p_\theta(x|z)p(z)$ N(0, I)

Multivariate Gaussian where $(u, \varepsilon) = \text{decoder_nn}(z; \phi)$

$$\text{ELBO}_i(\theta, \phi) = E_{q_\theta(z|x^{(i)})} [\log p_\theta(x^{(i)}|z)] - KL(q_\theta(z|x^{(i)}) || p(z))$$

just rewritten in terms of θ and ϕ instead of λ
b/c λ is a deterministic function of θ

$$= \sum_{s=1}^S q_\theta(z^{(s)}|x^{(i)}) \log p_\theta(x^{(i)}|z^{(s)}) - KL(q_\theta(z|x^{(i)}) || p(z))$$

Training VAE

where $z^{(s)} \sim q_\theta(z|x^{(i)})$ †s

Monte Carlo Approx.

① PGM Perspective:

just standard Variational EM
where we maximize $\text{ELBO}_i(\theta, \phi)$

② NN Perspective:

just backpropagation with a loss

↪ IVA perspective:

just backpropagation with a loss
consisting of two terms

- a) reconstruction loss
- b) regularizer

