Generative Modelling

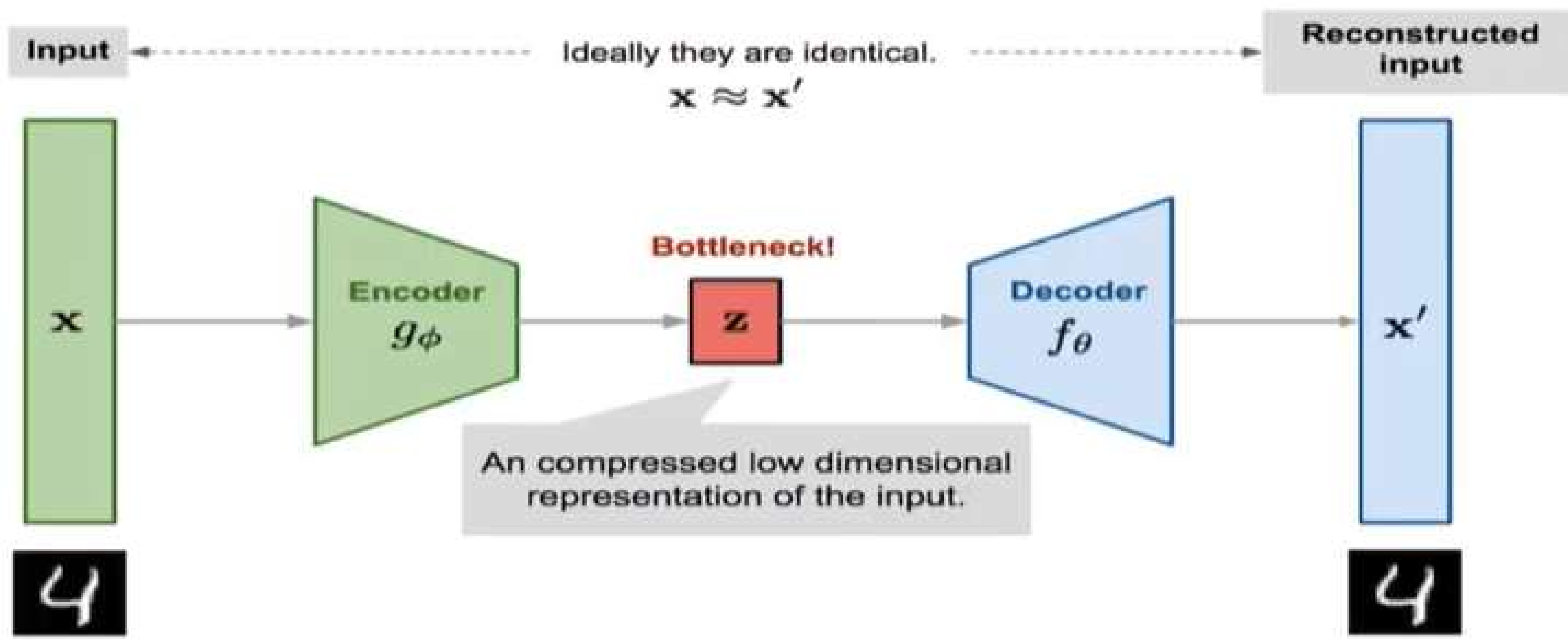
Variational Auto-Encoder

Some slides were adated/taken from various sources, including Andrew Ng's Coursera Lectures, CS231n: Convolutional Neural Networks for Visual Recognition lectures, Stanford University CS Waterloo Canada lectures, Aykut Erdem, et.al. tutorial on Deep Learning in Computer Vision, Ismini Lourentzou's lecture slide on "Introduction to Deep Learning", Ramprasaath's lecture slides, and many more. We thankfully acknowledge them. Students are requested to use this material for their study only and NOT to distribute it.

Topics to be discussed

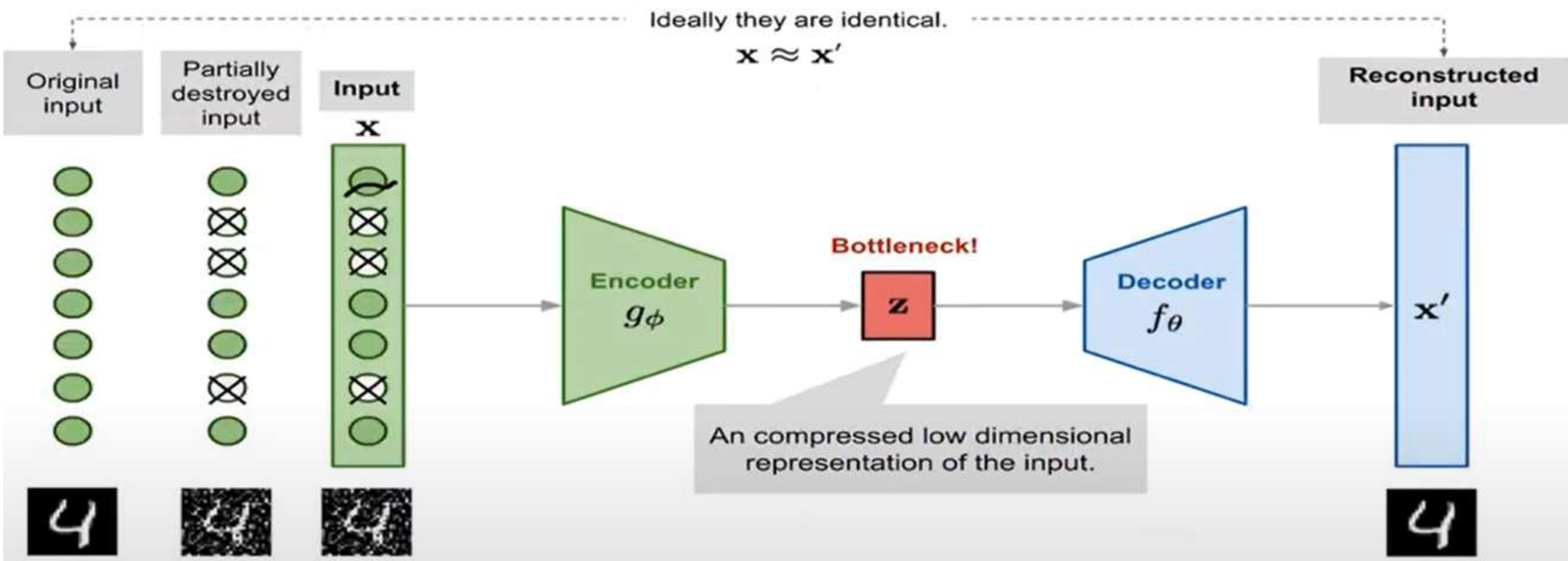
- Auto-Encoder for image reconstruction
- De-noising Auto-Encoder
- Variational Auto-Encoder – An Introduction
- KL Divergence
- Mathematical Formulations of VAE

Auto-Encoder for Image Reconstruction



Cost function: $L(\theta, \phi) = \frac{1}{n}\sum_{1}^{n}[x^{(i)} - f_\theta\left(g_\phi\left(x^{(i)}\right)\right)]^2$

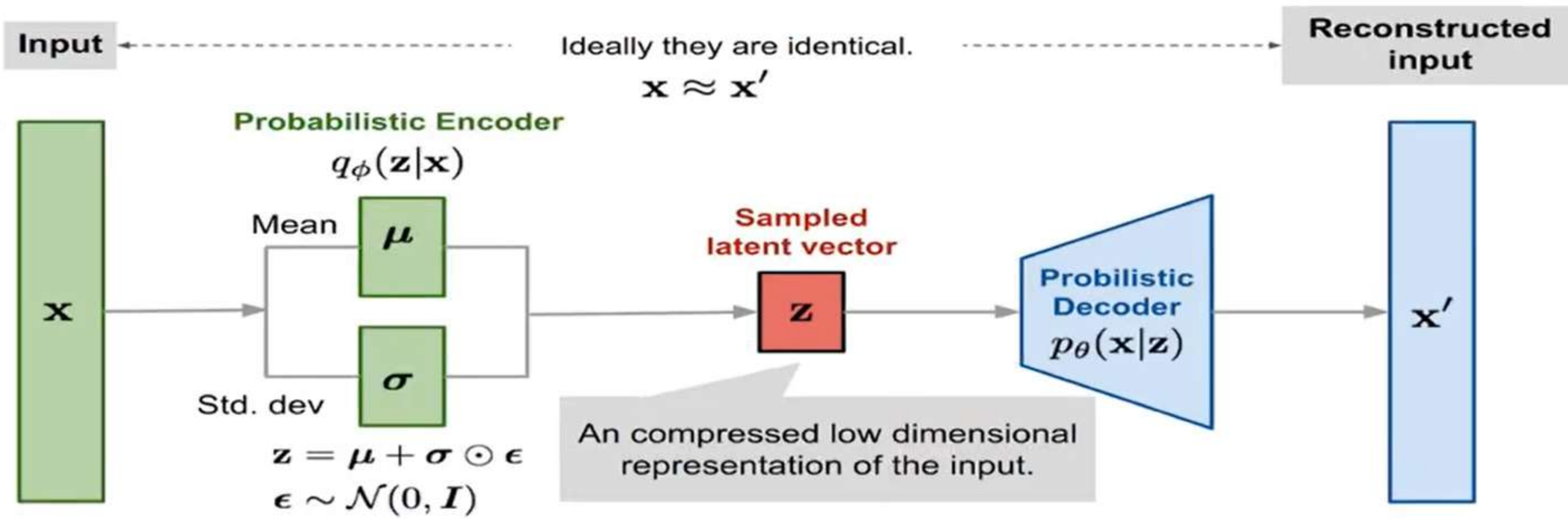
De-noising Auto-Encoder



$$\tilde{x}^i \sim \chi(\tilde{x}^i | x^i)$$

Loss function: $L(\theta, \phi) = \frac{1}{n}\sum_1^n [x^{(i)} - f_\theta\left(g_\phi\left(\tilde{x}^{(i)}\right)\right)]^2$

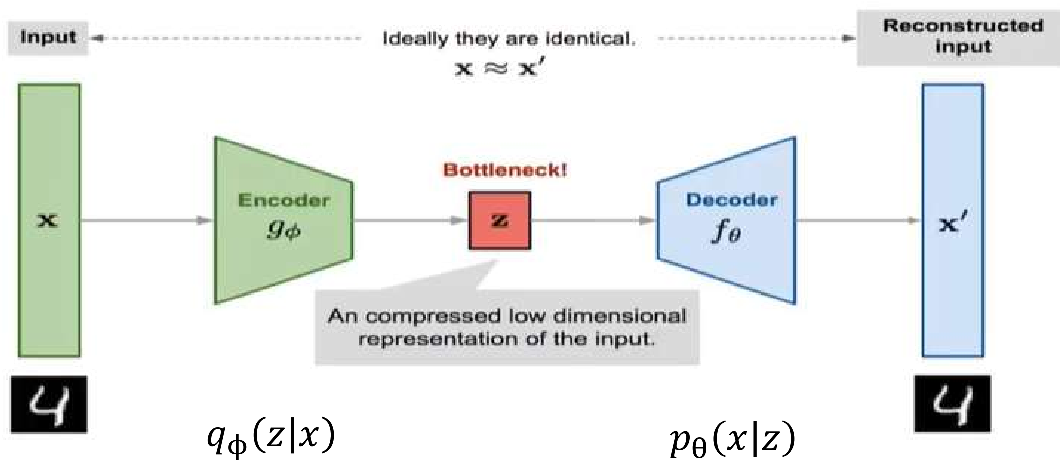
Variational Auto-Encoder



Loss Function $L(\theta, \phi) = -E_{z \sim q_\phi(z|x)}[p_\theta(x|z)] + D_{KL}(q_\phi(z|x)||p_\theta(z))$

Variational Auto-Encoder

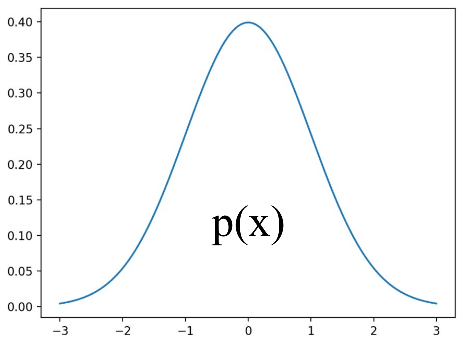
- Goals:
  - Find a distribution $q_\phi(z|x)$ of some latent variables which we can sample from $z \sim q_\phi(z|x)$ to generate new samples $x' \sim p_\theta(x|z)$.

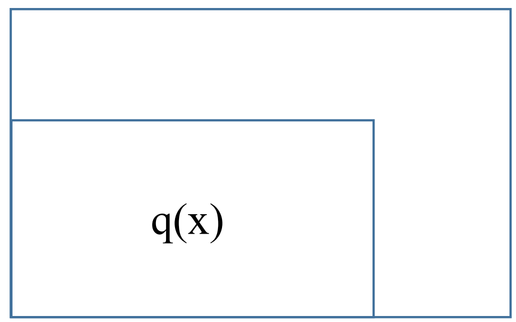


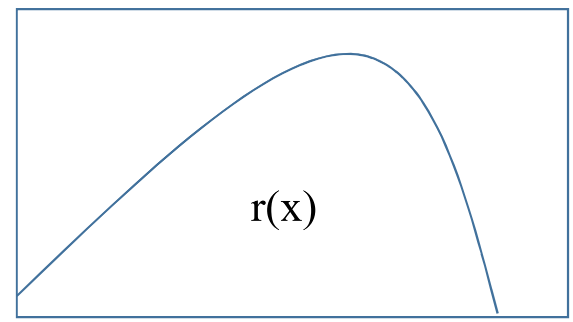
Typical Auto Encoder

KL Divergence

KL Divergence







$D_{KL}(p(x)||q(x) = high\ value$

$D_{KL}(p(x)||r(x) = small\ (\sim 0)$

KL Divergence

- Kullback-Leiller Divergence (D_{KL}) is a measure of how one probability distribution is different from the second. For the discrete probability distribution P and Q, the Kullback-Leiller divergence (D_{KL}) between P and Q is defined as

$$D_{KL}(P||Q) = \sum_{x} P(X = x)\log[\frac{P(X = x)}{Q(X = x)}] = \sum_{x} P(x)\log[\frac{P(x)}{Q(x)}]$$

- Properties of Kullback-Leiller Divergence (D_{KL}):

1. $D_{KL}(P||Q) or D_{KL}(Q||P) \geq 0$
2. $D_{KL}(P||Q) \neq D_{KL}(Q||P)$

KL Divergence

Suppose we have two multivariate normal distributions defined as

p(x) = $N(x; \mu_1, \Sigma_1)$ and q(x) = $N(x; \mu_2, \Sigma_2)$

where μ_1 and μ_2 are means and Σ_1 and Σ_2 are covariance (matrix)

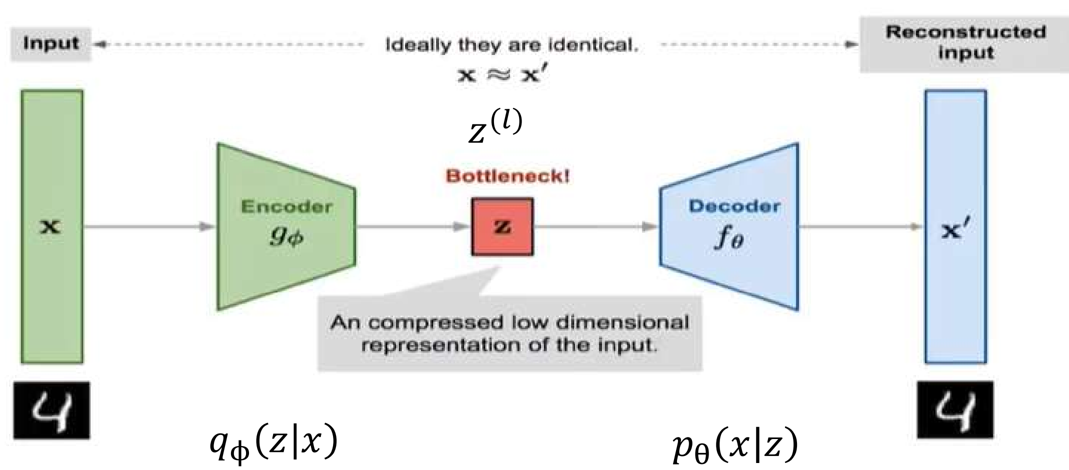
and the multivariate normal density is defined as

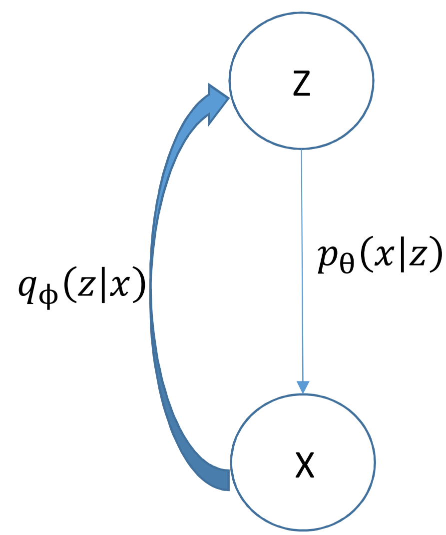
$N(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}}$ exp($-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$) with both the distributions have same dimension k then

$$D_{KL}(p(x)||q(x)) = \frac{1}{2}[log \frac{|\Sigma_2|}{|\Sigma_1|} - d + tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1)]$$

Goal of VAE

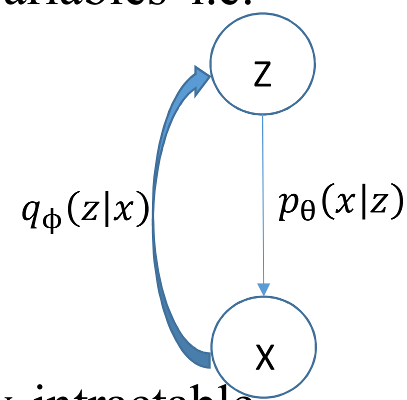
Find a distribution $q_{\phi}(z|x)$ of some latent variables which we can sample from $z \sim q_{\phi}(z|x)$ to generate new samples $\tilde{x} \sim p_{\theta}(x|z)$.





The problem of approximate inference

- Let x be a set of observed variables and z be a set of latent variables with joint probability distribution p(z,x). Then the inference problem is to compute the conditional distribution of the latent variables given the observed variables i.e. p(z|x).
- From Bayes theorem, we can write it as $p(z|x) = \frac{p(X|Z)p(z)}{p(x)}$ -- (A)
- Computation of (A) is hard as p(x) is mathematically intractable.
- Reason: $p(x) = \int p(x|z)p(z)dz = \int (p(x,z)dz$
- This integral is not available in the closed form or is mathematically intractable due to multiple integral is involved for latent variable dimensinality.



Alternative

- The alternative is to approximate the $p(z|x)$ by another distribution $q(z|x)$ which is defined in such a way that it has tractable solution. This is done by Variational inference. The main idea of Variational inference is to pose the inference problem as optimization problem.
- How? By modelling $P(z|x)$ using $Q(z|x)$ where $Q(z|x)$ has a simple distribution such as Gaussian.
- Now, we calculate the difference between $P(z|x)$ and $Q(z|x)$ using KL divergence and try to make it as close as possible to zero.

Rest is from board work