Variational Autoencoders (VAE)

Mir Murtaza (PhD. Fellow)

Lecture Outline

First Preliminary Topics:

- Maximum Likelihood Estimation (MLE)
- Bayes Theorem
- Generative and Discriminative Models

Variational Autoencoders (VAE):

- Representation
- Learning Mechanism

Maximum Likelihood Estimation (MLE)

- We seek to uncover general laws and principles that govern the behaviour under investigation

- We formulate hypotheses to study these laws and principles

- Such hypotheses are stated in terms of probability distributions called models

Probability Distribution Functions

Illustration on Board ...

MLE Process

● Statistically speaking, a data vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ is a random sample from an unknown population

● Goal is to identify the population **that generated the data**

● Each of the population is identified by a **corresponding probability distribution**

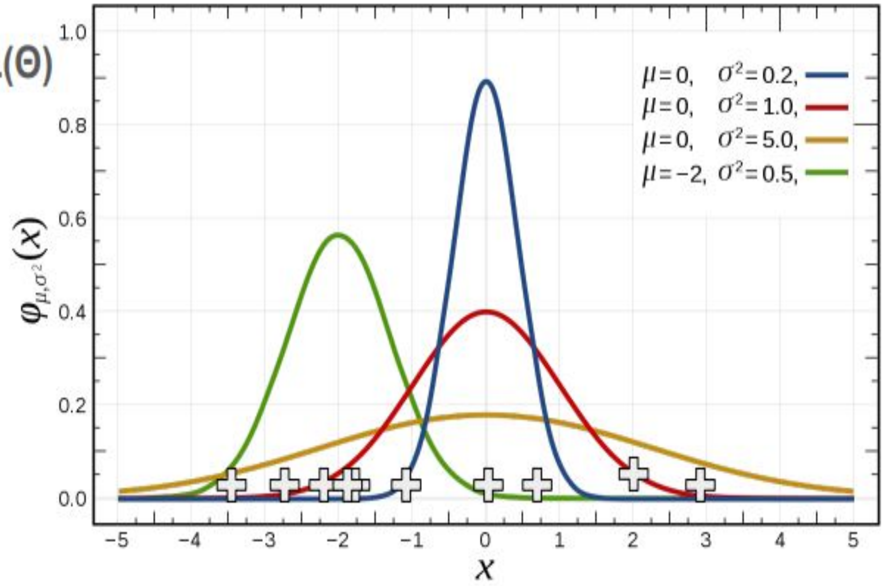
● And **a model** is specified by a family of probability distributions

● For e.g $\mathbf{f(x|\Theta)}$ denotes the probability density function (pdf) of observing data x given the parameters $\mathbf{\Theta}$

Mathematical Formulation of MLE

- The task is find the Θ that maximizes the probability
- $L(\Theta) = P(X_1{=}x_1 , X_2{=}x_2 , \ldots, X_3{=}x_3) = f(x_1 ;\Theta).f(x_2 ,\Theta)\ldots f(x_n ,\Theta) = \Sigma_i f(x_i , \Theta)$
- ΘMLE = argmax $L(\Theta)$
- ΘMLE = argmin - Log $L(\Theta)$
- ∂ $\Theta MLE = 0$

.(Θ)



MLE Derivation for Mean (μ) for a Gaussian Distribution

Θ = {μ, $\boldsymbol{\sigma}$ 2 }

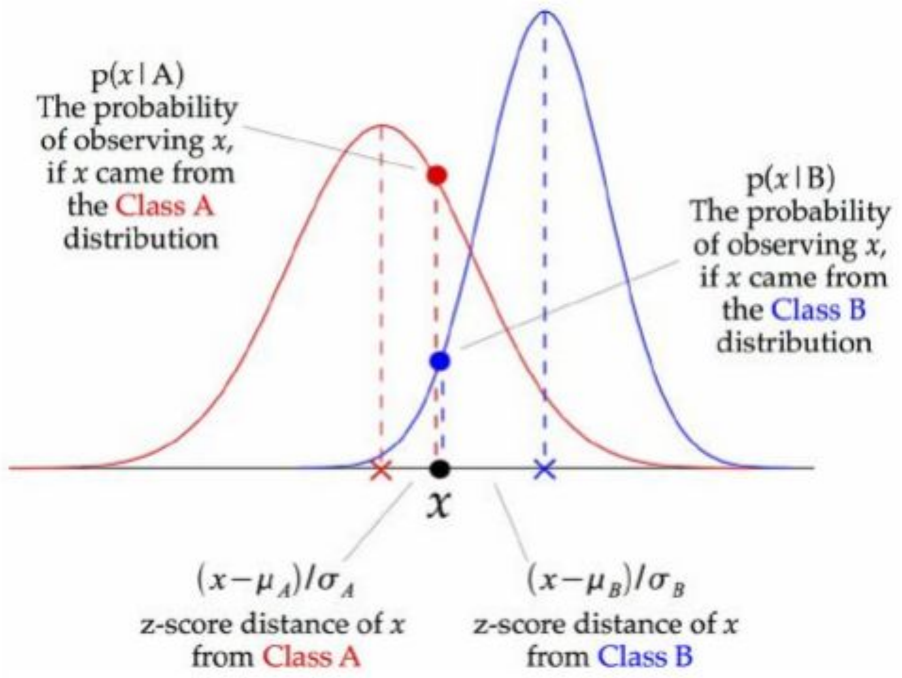
$$P(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\ln P(\mathcal{D} \mid \mu, \sigma) \;=\; \ln\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \prod_{i=1}^{N} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\right]$$

$$\widehat{\mu}_{MLE} \;=\; \frac{1}{N}\sum_{i=1}^{N} x_i$$

Generative Models

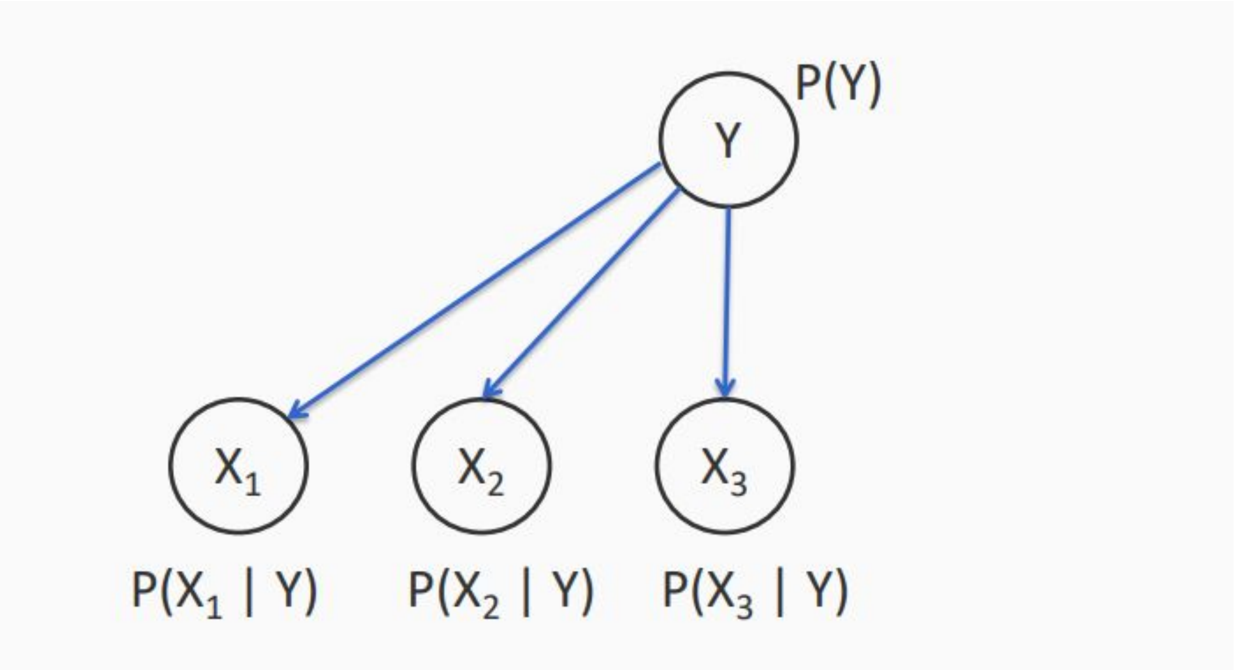
Estimate the model → Define the classifier



Bayes Rule

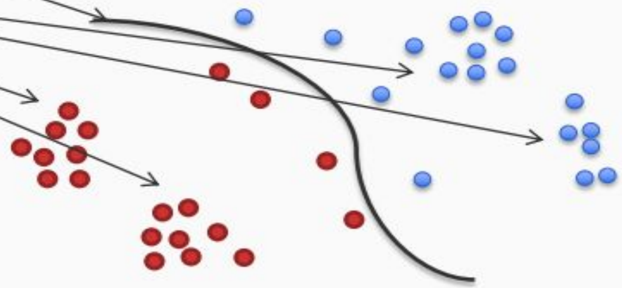
$$\underbrace{P(a|b)}_{\text{Posterior}} = \frac{\overbrace{P(b|a)}^{\text{Likelihood}} * \overbrace{P(a)}^{\text{Prior}}}{\underbrace{P(b)}_{\text{Normalization}}}$$

Probabilistic Graphical Models



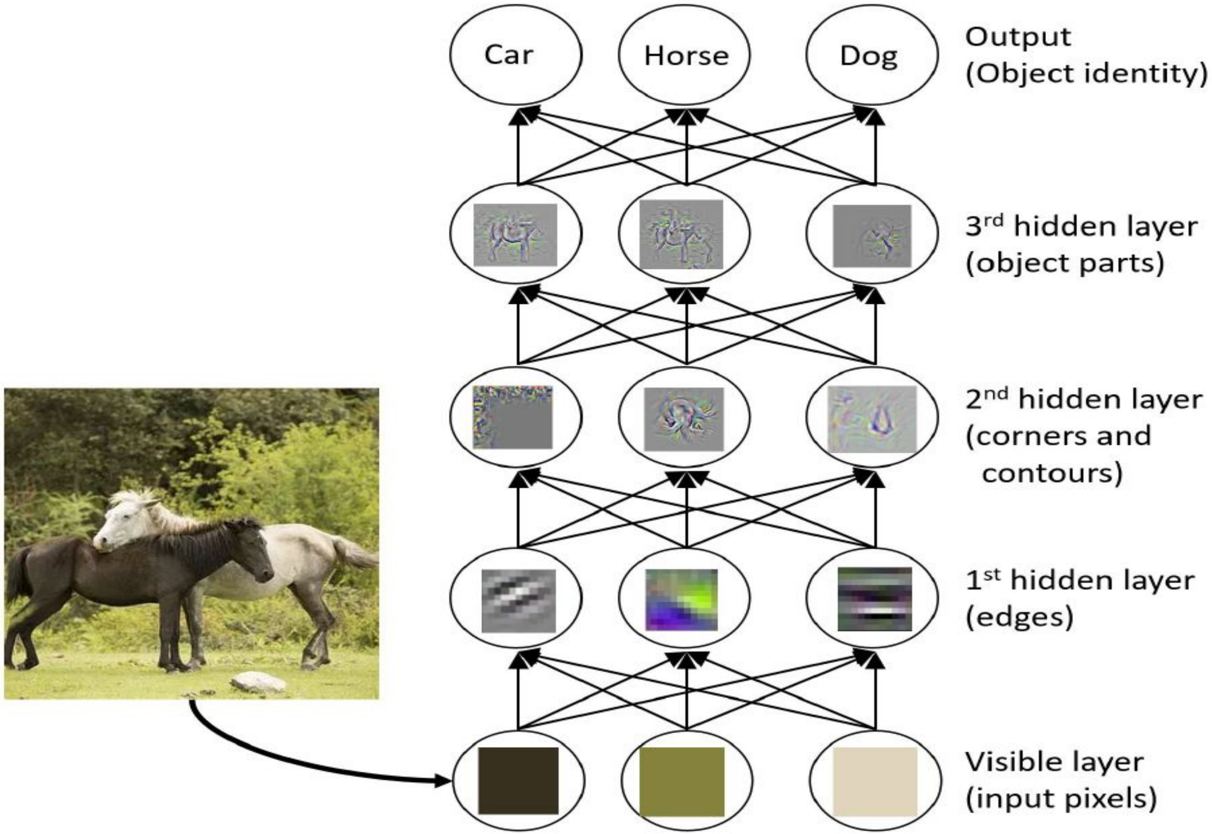
Generative and Discriminative Models

- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model



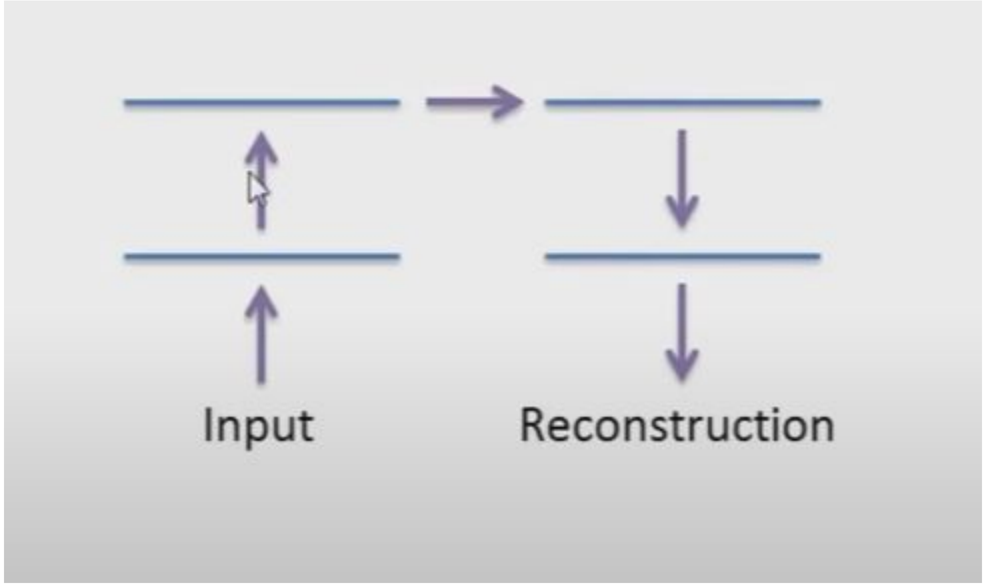
- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models

Deep Learning or Deep Representation Learning

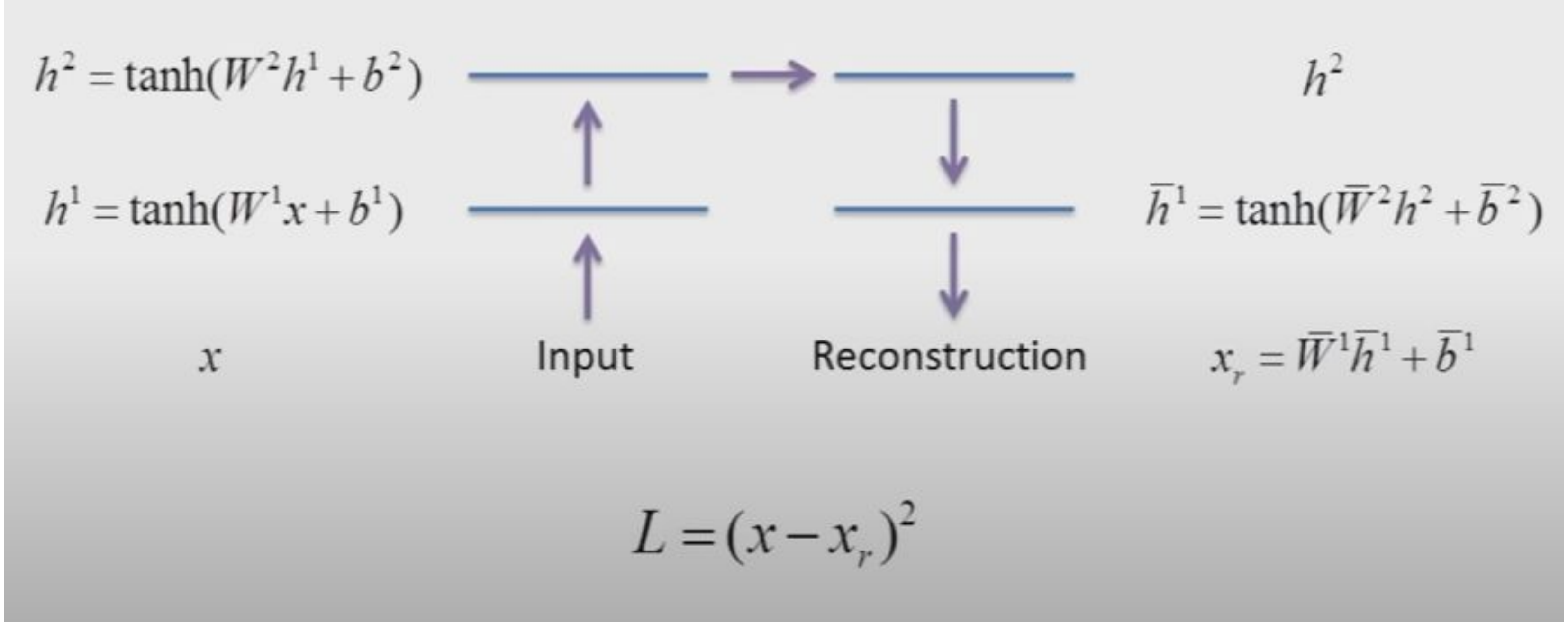


Autoencoders

Features Learned?



Autoencoders



Basic Principle ...

The simpler the explanation of data, the more likely it is correct.

Goal is to maximally compress the data …

As we did in the MLE example: parameterized distribution for a population, dataset ...

Variational Autoencoders (VAE)

Graphical Representation:

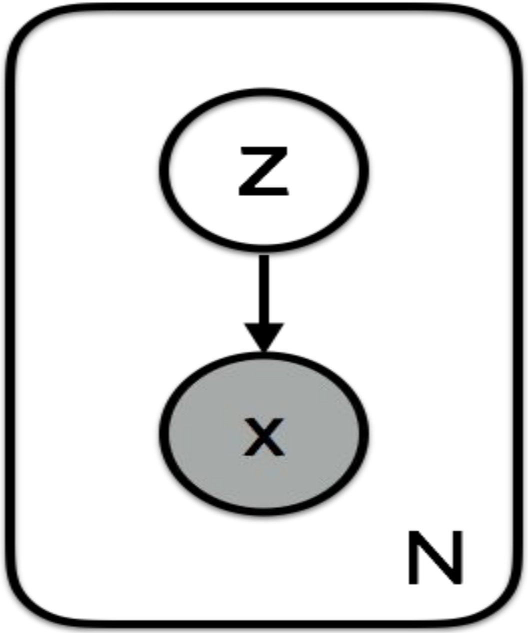
Prior: P(z) **Latent Representation**

Generate Data: P(x|z)

Bayes Rule: P(x|z) ~ P(z|x) P(x) / evidence

P(z|x) is computationally intractable

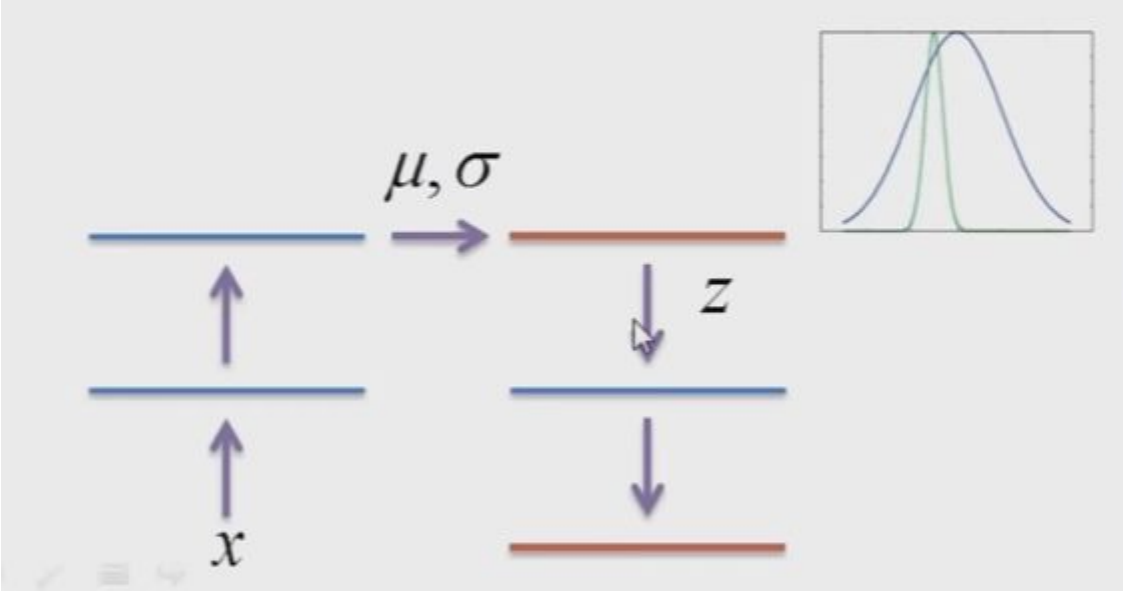
Approximate P(z|x) ~ q(z|x)



VAE

P(q|x):

Constraint P(q|x) to be a gaussian family of distributions…



Neural Net Approximator

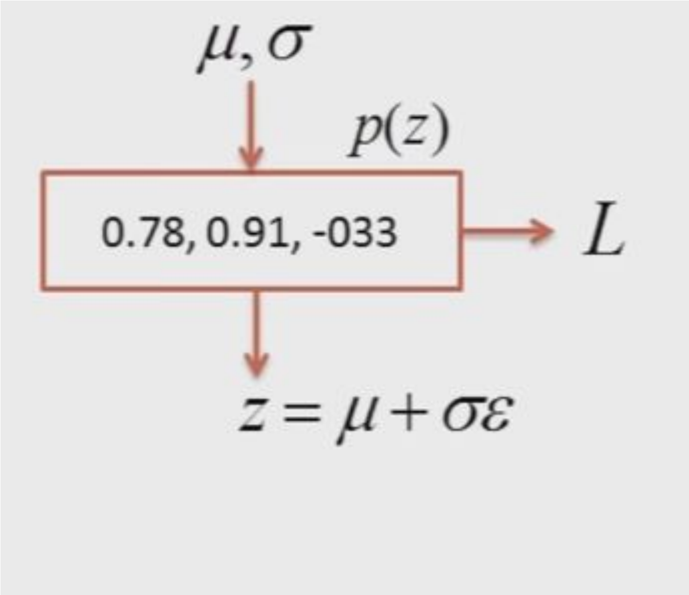
Loss: L = -log p(z) / q(z|x)

-- Learn parameters: compute function gradients

with **stochastic gradient descent** +

-- Update hidden layer weights via

backpropagation

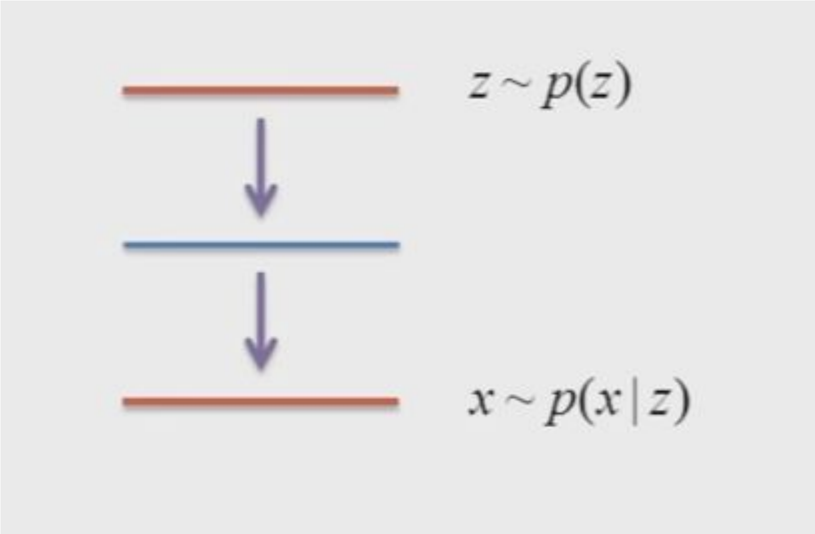


Generation (Imagination)

1) Choose z~p(z)
2) Calculate p(x|z)
3) Choose x ~ p(x|z)

Single pass through VAE

Very Efficient !!!



VAE Generation Results