NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 59: Variational Autoencoder - III

CONCEPTS COVERED

Concepts Covered

- ❑ Generative Model
- ❑ Limitations of usual auto-encoder
- ❑ Intuitions behind VAE
- ❑ Variational Inference
- ❑ Practical Realization of VAE

Variational Autoencoder : Variational Inference

- In VAE, we assume that there is a latent (unobserved) variable, z, generating our observed random variable, x.

- Our aim: To compute the posterior $P(z|x) = \frac{P(X|Z)P(z)}{P(x)}$
- $P(x) = \int P(x|z)P(z)dz$ $\longrightarrow$ Intractable

Variational Autoencoder : Variational Inference

- Let's assume there is a tractable distribution Q, such that $P(z|x) \approx Q(z|x)$
- We want $Q(\cdot)$ to be in the family of tractable distributions (Gaussian for example) such that we can play around with its parameters to match P$(z|x)$
- So, we will aim towards minimizing KL divergence of P$(z|x)$ with respect to $Q(z|x)$
- Our objective: minimize KL($Q(z|x)$ || P$(z|x)$)

KL Divergence

Minimize

$$KL(Q(z|x)||P(z|x))$$

KL Divergence

$$KL(Q(z|x)||P(z|x)) = -\sum_{z} Q(z|x)\log\frac{P(x,z)}{Q(z|x)} + \ logP(x)$$

$$\downarrow$$

$$\log P(x) = KL(Q(z|x)\big|\big|P(z|x)\big) + \sum_{z} Q(z|x)\log\frac{P(x,z)}{Q(z|x)}$$

KL Divergence

$$\log P(x) = KL(Q(z|x)||P(z|x)) + \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$$

- Since, x is given, LHS is constant.
- Aim is to minimize $KL(Q(z|x)||P(z|x))$
- This is same as maximizing $\sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$

KL Divergence

$$\log P(x) = KL(Q(z|x)||P(z|x)) + \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$$

- Since, x is given, LHS is constant.
- Aim is to minimize $KL(Q(z|x)||P(z|x))$
- This is same as maximizing $\sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$

Variational Lower Bound

Variational Lower Bound

$$\log P(x) = KL\big(Q(z|x)\,||\,P(z|x)\big) + \sum_{z} Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$$

$$KL\big(Q(z|x)\,||\,P(z|x)\big) \geq 0$$

$$\sum_{z} Q(z|x) \log \frac{P(x,z)}{Q(z|x)} \leq logP(x)$$

Variational Autoencoder : Variational Inference

- Our initial objective: minimize KL($Q(z|x)$ || P$(z|x)$)
- Which is same as maximizing $\sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)}$

Variational Lower Bound

- So, aim now is: *maximize*

$$\mathrm{L} = \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)} = \sum_z Q(z|x) \log \frac{P(x|z)P(z)}{Q(z|x)}$$

Variational Autoencoder : Variational Inference

Maximize

Variational Autoencoder : Variational Inference

Maximize

$$\mathrm{L} = \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)} = \sum_z Q(z|x) \log \frac{P(x|z)P(z)}{Q(z|x)}$$

Variational Autoencoder : Variational Inference

Maximize

$$L = \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)} = \sum_z Q(z|x) \log \frac{P(x|Z)P(z)}{Q(z|x)}$$

$$= \sum Q(z|x) \log P(x|z) + \sum Q(z|x) \log \frac{P(z)}{Q(z|x)}$$

Variational Autoencoder : Variational Inference

Maximize

$$L = \sum_z Q(z|x) \log \frac{P(x,z)}{Q(z|x)} = \sum_z Q(z|x) \log \frac{P(x|Z)P(z)}{Q(z|x)}$$

$$= \underbrace{\sum Q(z|x) \log P(x|z)}_{E_{Q(z|x)} \log P(x|z)} + \underbrace{\sum Q(z|x) \log \frac{P(z)}{Q(z|x)}}_{-KL(Q(z|x)\,||\,P(z))}$$

Variational Autoencoder : Variational Inference

Maximize

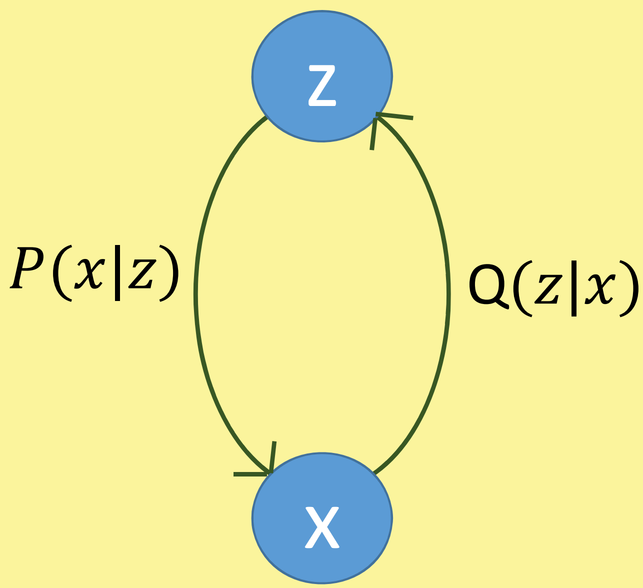
$$\mathrm{L} = \sum_z Q(z|x)\log\frac{P(x,z)}{Q(z|x)} = \sum_z Q(z|x)\log\frac{P(x|Z)P(z)}{Q(z|x)}$$

$$= \underbrace{\sum Q(z|x)\log P(x|z)}_{E_{Q(z|x)}\log P(x|z)} + \underbrace{\sum Q(z|x)\log\frac{P(z)}{Q(z|x)}}_{-KL(Q(z|x)\,||\,P(z))}$$

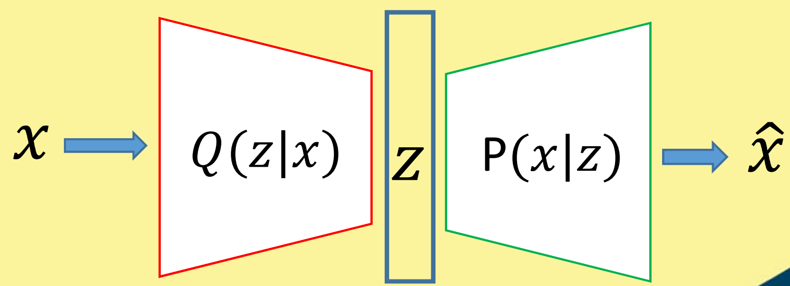
- Translate the loss functions into an auto-encoder architecture.

Variational Autoencoder : Network Realization

- We have the following graphical model



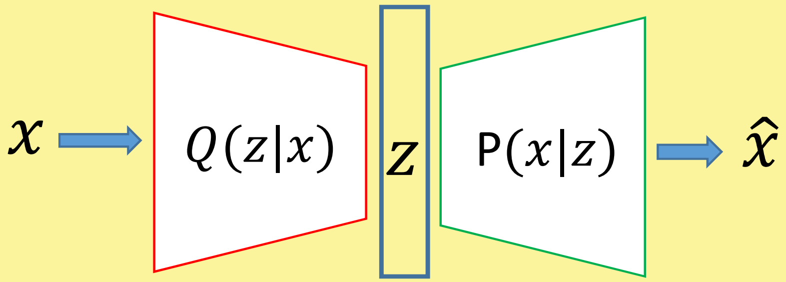
- Realize both $P(\cdot)$ and $Q(\cdot)$ with neural networks



Variational Autoencoder : Network Realization

- The z codes we get here should match with the distribution of $P(z)$ and we can decide what prior distribution to choose for $P(z)$.

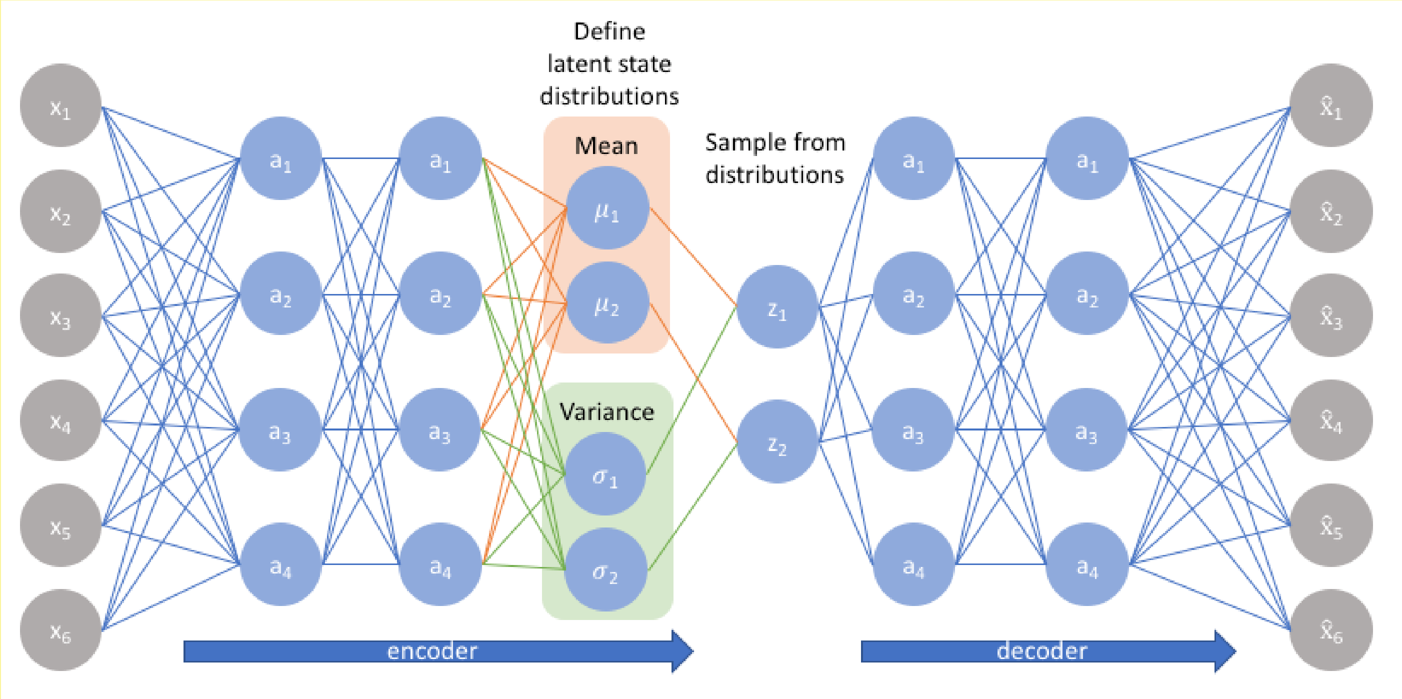
- Usual practice is to select a Normal distribution $N(0, I)$ for the prior.



Variational Autoencoder : Network Realization

- Instead of generating a fixed code for an input, Encoder now gives parameters of the distribution of the latent code.
- For a given input x, we need to generate mean vector $\mu(x)$ and diagonal covariance matrix, $\Sigma(x)$.
- We need to SAMPLE a code from that latent distribution and pass forward to the Decoder.

Variational Autoencoder : Network Realization



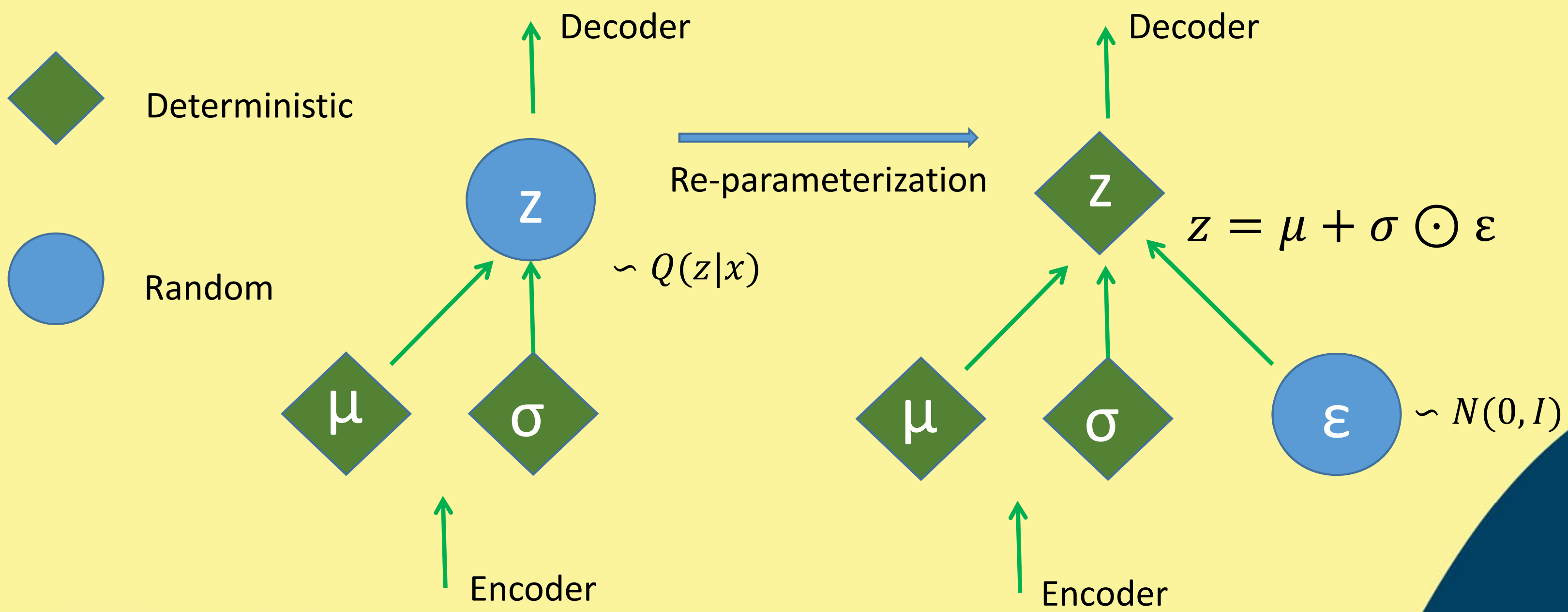
https://www.jeremyjordan.me/variational-autoencoders/

Variational Autoencoder : Network Realization

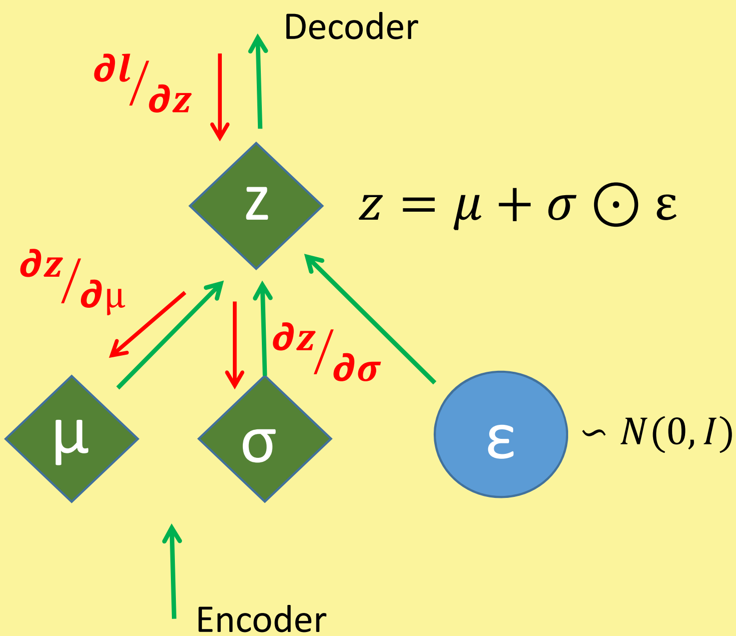
Sampling breaks computational graph
and
hinders Gradient Descent based
optimization

Variational Autoencoder : Reparameterization Trick

- We randomly sample ε from a unit Gaussian, and then shift the randomly sampled ε by the latent distribution's mean μ and scale it by the latent distribution's variance σ.



Variational Autoencoder : Reparameterization Trick



Re-parameterization enables

- ❑ Optimization of the parameters of the distribution.
- ❑ Still maintaining the ability to randomly sample from that distribution.

Variational Autoencoder : Coding the Cost Functions

$$E_{Q(z|x)} \log P(x|z) - KL(Q(z|x) \,||\, P(z))$$

Maximize

Minimize

Variational Autoencoder : Coding the Cost Functions

- Maximizing $E_{Q(z|x)} \log P(x|z)$ is a maximum likelihood estimation. It is observed all the time in discriminative supervised model, for example Logistic Regression, SVM, or Linear Regression.
- In the other words, given an input z and an output x, we want to maximize the conditional distribution $P(x|z)$ under some model parameters.
- So we could implement it by using any classifier with input z and output x, then optimize the objective function by using for example log loss or regression loss.

Variational Autoencoder : Coding the Cost Functions

- We want to minimize the second component of the loss, $KL((Q(z|x) \mid\mid P(z))$
- We assumed that $P(z)$ follows $N(0, I)$, so we have to push $Q(z|x)$ towards $N(0, I)$

Assuming $P(z)$ to be $N(0, I)$ has 2 advantages:

- Easy to sample latent vectors from $N(0, I)$ when we want to generate samples.
- Assuming $Q(z|x)$ to be a Gaussian distribution with parameters, $\mu(x)$ and $\Sigma(x)$ allows $KL(Q(z|x) \mid\mid P(z))$ to be in a closed form and easy for optimization.

NPTEL ONLINE CERTIFICATION COURSES

Thank you