Auto-Encoding Variational Bayes (Variational Autoencoders)

Faisal Mohamed

- Core idea.
- Implementation.
- Some Experimental results and comparisons.

- Core idea.
- Implementation.
- Some Experimental results and comparisons.

Scalable Variational Inference

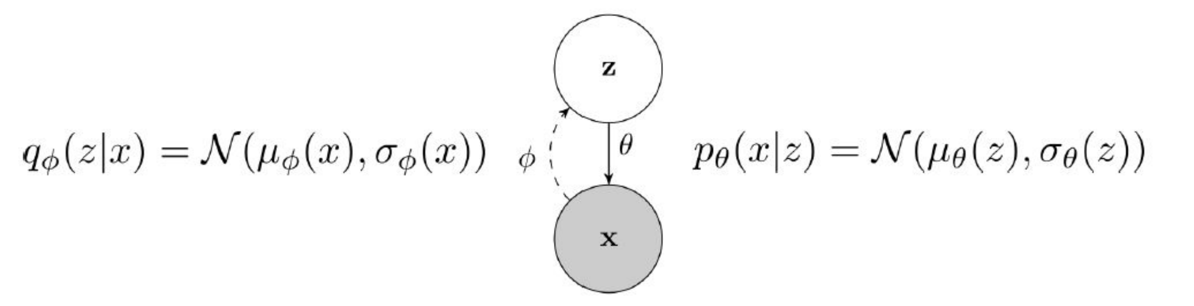
- Works on the intractable case.
- Can be optimized with SGD.

Problem setup

- **We have a dataset X of size N.**
- **We want the learn the data distribution P(X)**
- **We assume that there is some hidden lower dimensional variable called z that can represent the data efficiently.**
- **The problem now is** $P(X) = \int P(X|Z)P(Z)\, dz$
- **This integral is intractable for vectors.**

Varitional auto conoders

- **Solve this problem by learning both P(Z) and P(X|Z).**
- **We sample z from P(Z) network and pass it to P(X|Z) to return a data sample.**

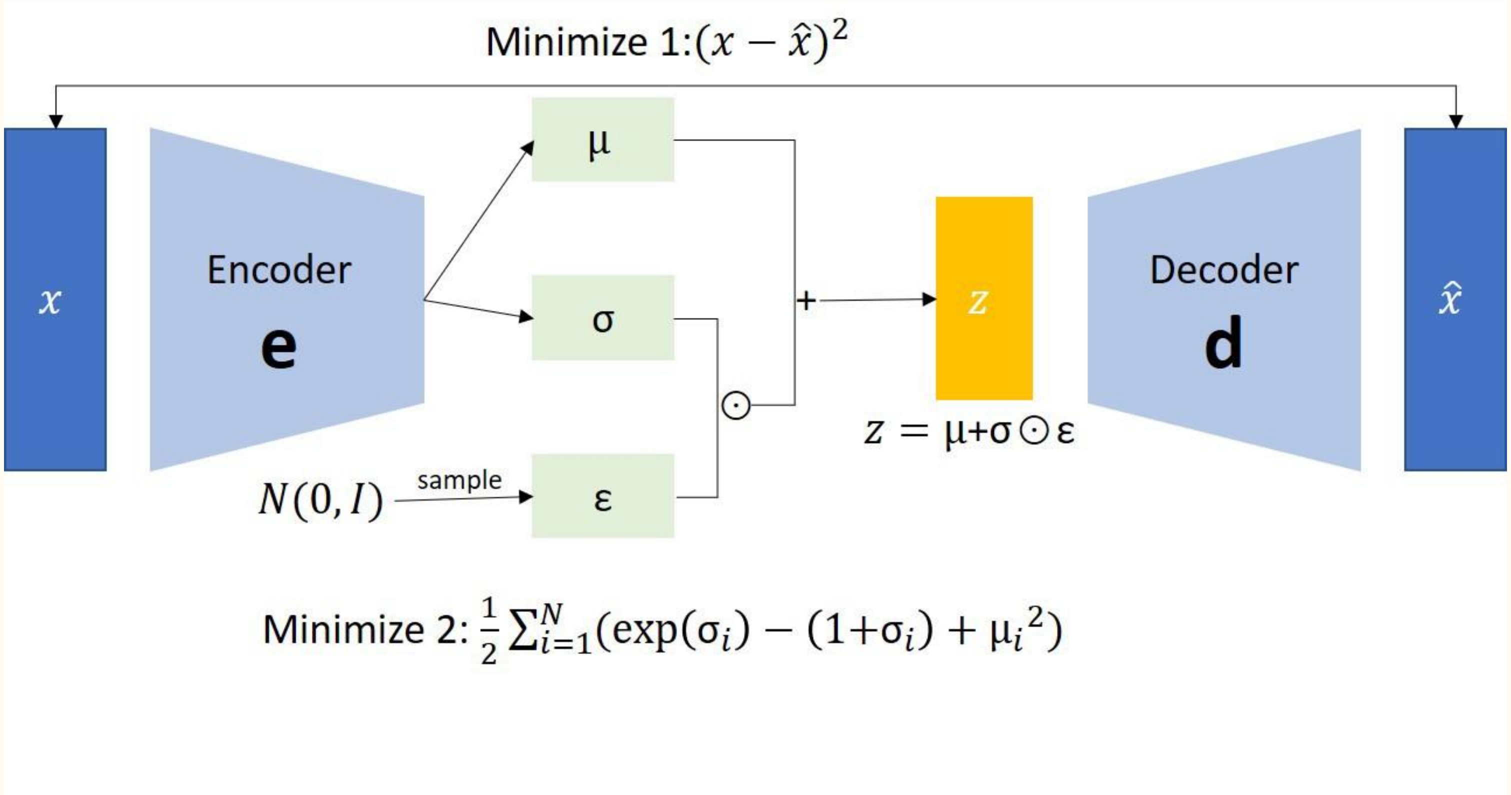


- Core idea.
- Implementation.
- Some Experimental results and comparisons.

Algorithm

$\boldsymbol{\theta}, \boldsymbol{\phi} \leftarrow$ Initialize parameters
repeat
  $\mathbf{X}^M \leftarrow$ Random minibatch of M datapoints (drawn from full dataset)
  $\boldsymbol{\epsilon} \leftarrow$ Random samples from noise distribution $p(\boldsymbol{\epsilon})$
  $\mathbf{g} \leftarrow \nabla_{\boldsymbol{\theta},\boldsymbol{\phi}} \widetilde{\mathcal{L}}^M(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}^M, \boldsymbol{\epsilon})$ (Gradients of minibatch estimator (8))
  $\boldsymbol{\theta}, \boldsymbol{\phi} \leftarrow$ Update parameters using gradients $\mathbf{g}$ (e.g. SGD or Adagrad [DHS10])
until convergence of parameters $(\boldsymbol{\theta}, \boldsymbol{\phi})$
return $\boldsymbol{\theta}, \boldsymbol{\phi}$

Algorithm

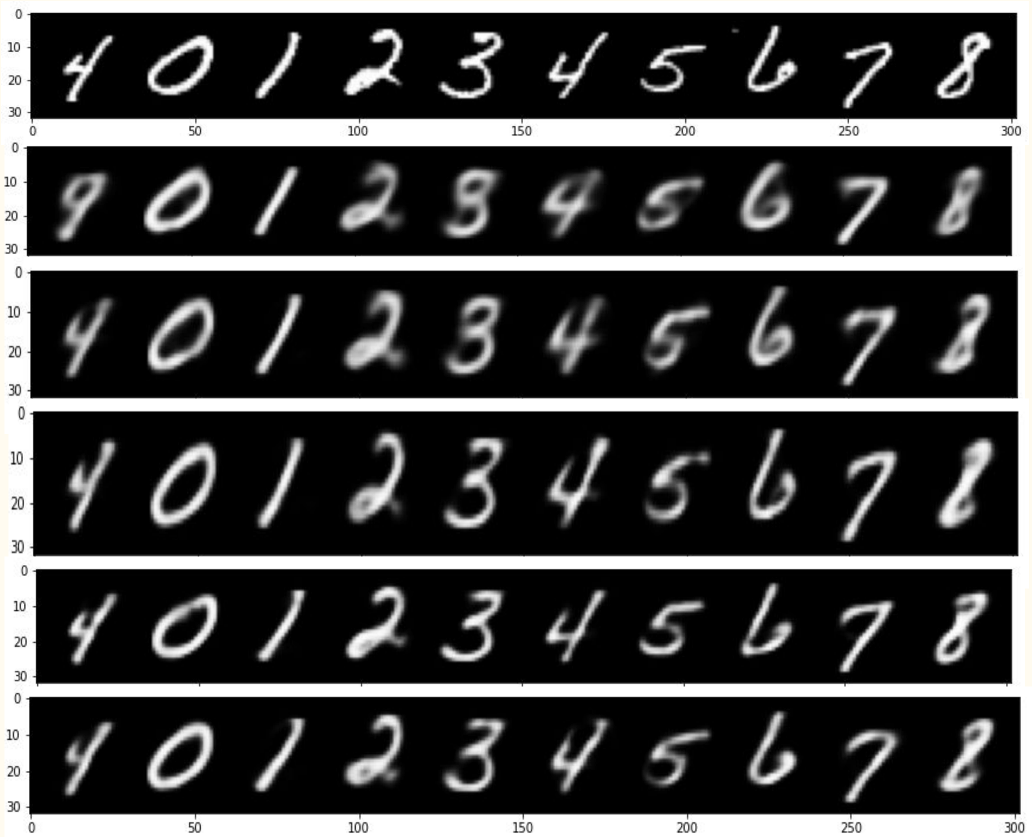


Minimize 1: $(x - \hat{x})^2$
x
Encoder
e
μ
σ
ε
$N(0, I)$
sample
+
⊙
z
$z = \mu + \sigma \odot \varepsilon$
Decoder
d
$\hat{x}$
Minimize 2: $\frac{1}{2}\sum_{i=1}^{N}(\exp(\sigma_i) - (1+\sigma_i) + {\mu_i}^2)$

- Core idea.
- Implementation.
- Experimental results.

Experiment

- Dataset: MNIST
- Model: Single layer MLP with 500 units.



Original Image

Size of z = 3

Size of z = 5

Size of z = 10

Size of z = 20

Size of z = 200