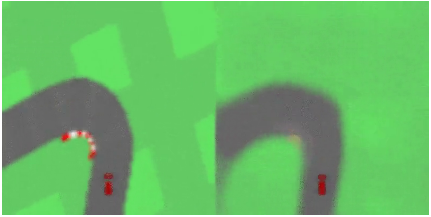
VAE

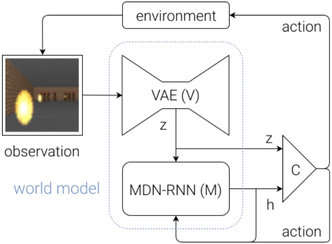
Titus Nicolae

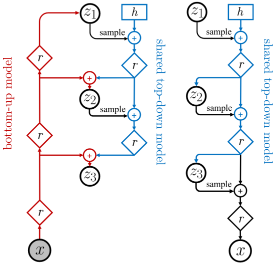
November 5, 2020

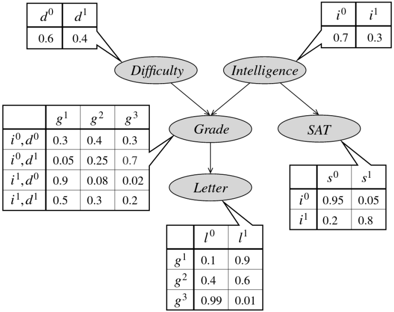
Motivation

World Models






bottom-up model
shared top-down model
shared top-down model
z_1
z_2
z_3
h
r
x
sample



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

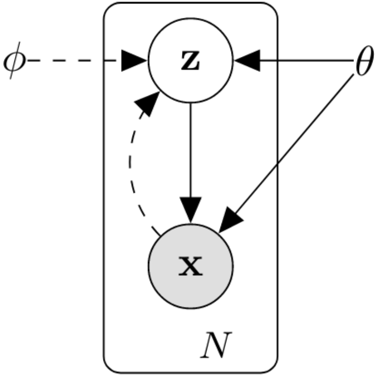
	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

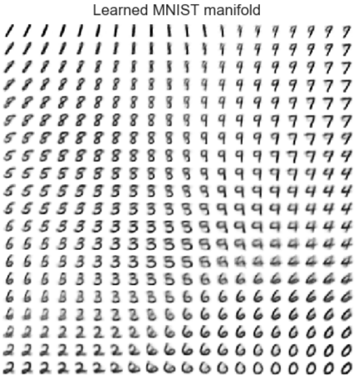
$$P(I, D, G, S, L) = P(I)P(D)P(G|I, D)P(S|I)P(L|G)$$

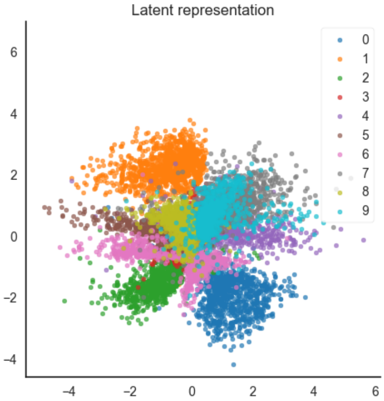
Graphical model for VAE



Latent space

VAE Mnist 2-dimenstions





Bayesian approach

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \tag{1}$$

- x - data
- z - latent code / unobserved variable
- $p(z)$ - prior - probability distribution of the latent code (let's only consider gaussian prior distribution)
- $p(x)$ - evidence - probability distribution of observed data independent of unobserved variable
- $p(z|x)$ - posterior - probability distribution of the parameter given the observed data
- $p(x|z)$ - likelihood - probability distribution of the data given the parameter value - how likely is the data for a particular latent code

Bayesian approach

$$p(z|x) = \frac{p(x,z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\int p(x|z)p(z)dz} \tag{2}$$

$\int p(x|z)p(z)dz$

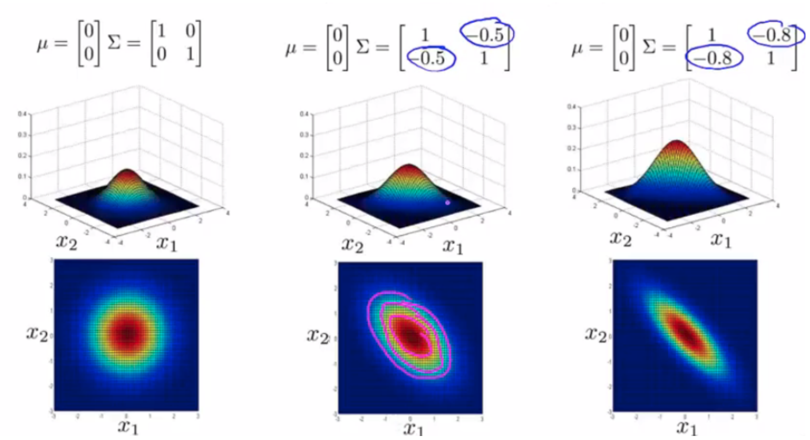
- can't be integrated directly
- for low dimensions could be solved with MCMC
- approximate $p(z|x)$ with a tractable distribtuion (Variational Inference)

KL divergence

$$KL(q||p) = \int q(x) log \frac{q(x)}{p(x)} dx = - \int q(x) log \frac{p(x)}{q(x)} dx \qquad (3)$$

- measure of closeness of probability distributions
- $KL(q||p) \geq 0$
- not symmetrical $KL(q||p) \neq KL(p||q)$
- for $KL(q||p)$ to be minimized we need p to have high probability wherever q has high probability

Multivariate Gaussian distribution



Variational Inference

$$p(z|x) = \frac{p(x,z)}{p(x)} \tag{4}$$

$$KL(q||p) = -\int q(x)log(\frac{p(x)}{q(x)})dx \tag{5}$$

- We'll approximate $p(z|x)$ with a diagonally gaussian distribution $q(z|x)$

$$KL(q(z|x)||p(z|x)) = -\int_z q(z|x)log\frac{p(z|x)}{q(z|x)} \tag{6}$$

$$= -\int_z q(z|x)log(\frac{p(x,z)}{q(z|x)}\frac{1}{p(x)}) \tag{7}$$

$$= -\int_z q(z)\left[log(\frac{p(x,z)}{q(z|x)}) - log(p(x))\right] \tag{8}$$

$$= -\int_z q(z)log\frac{p(x,z)}{q(z|x)} + \int_z q(z|x)log(p(x)) \tag{9}$$

Variational Inference

$$KL(q(z|x)||p(z|x)) = -\int_z q(z|x) log \frac{p(x,z)}{q(z|x)} + \int_z q(z|x) log(p(x)) \tag{10}$$

$$= -\int_z q(z|x) log \frac{p(x,z)}{q(z|x)} + log(p(x)) \int_z q(z|x)) \tag{11}$$

$$= -\int_z q(z|x) log \frac{p(x,z)}{q(z|x)} + log(p(x)) \tag{12}$$

$$log(p(x)) = KL(q(z|x)||p(z|x)) + \int_z q(z) log \frac{p(x,z)}{q(z|x)} \tag{13}$$

- For a given x we have p(x) constant, thus

$$ct. = KL(q(z|x)||p(z|x)) \Big\downarrow + \int_z q(z|x) log \frac{p(x,z)}{q(z|x)} \Big\uparrow \tag{14}$$

Variational Inference

$$log(p(x)) = KL(q(z|x)||p(z|x))\Big\downarrow + \int_z q(z|x) log \frac{p(x,z)}{q(z|x)}\Big\uparrow \quad (15)$$

- The term $\int_z q(z|x) log \frac{p(x,z)}{q(z)}$ is called variational lower bound also known as evidence lower bound (ELBO)

Variational Inference

- Evidence lower bound (ELBO)

$$\int_z q(z|x) log \frac{p(x,z)}{q(z|x)} = \int_z q(z|x) log \frac{p(x|z)p(z)}{q(z|x)} \tag{16}$$

$$= \int_z q(z|x) log(p(x|z)) + \int q(z|x) log \frac{p(z)}{q(z|x)} \tag{17}$$

$$= \mathbb{E}_{q(z)}\, log(p(x|z)) - KL(q(z|x)||p(z)) \tag{18}$$

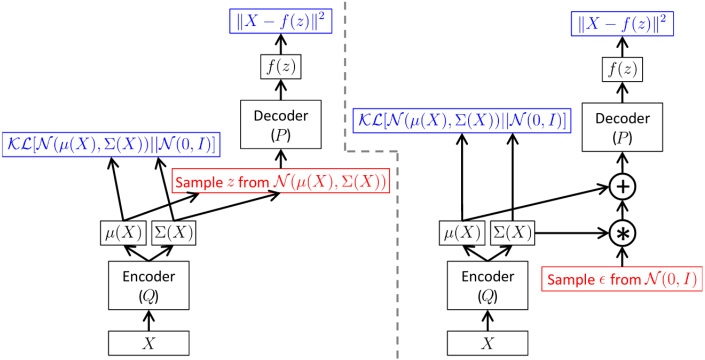
- We want to maximize likelihood $p(x|z)$ and make $q(z|x)$ and $p(z)$ more similar

Variational Inference

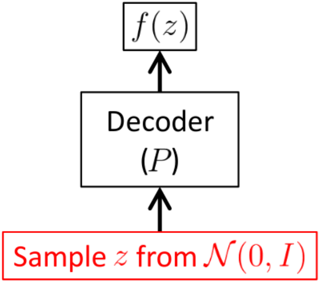
- $E_{q(z)} log(p(x|z))$ will effectively be a binary cross entropy loss if we consider that $p(x|z)$ is a MLP neural network with sigmoid activations
- $E_{q(z)} log(p(x|z))$ will effectively be a MSE if we consider that $p(x|z)$ is a gaussian distributed
- $KL(q(z|x)||p(z))$ can be computed analytically if we consider $p(z)$ prior as normal distribution

Reparametrization trick

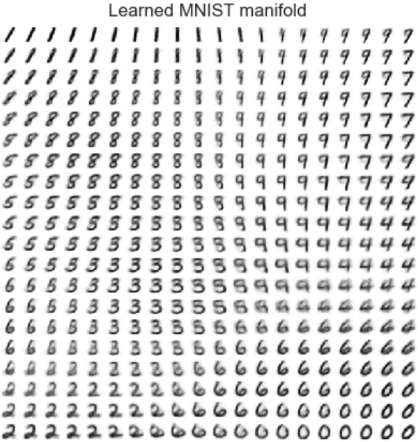
- We sample z from $q(z|x)$, since this is a stochastic process we can't backpropagate through it



Inference



$f(z)$
Decoder
(P)
Sample z from $\mathcal{N}(0, I)$



Learned MNIST manifold

References

- Kingma, Diederik P and Welling, Max. Auto-Encoding Variational Bayes.
- https://blog.evjang.com/2016/08/variational-bayes.html
- http://gregorygundersen.com/blog/2018/04/29/reparameterization/
- https://www.youtube.com/watch?v=uaaqyVS9-rM
- https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.html