CS446 HW3

Shuzhen Zhang
Net ID: shuzhen2

1 Variational Auto-Encoders [Written]

(a)

$$p_\theta(x|z) = \Pi_{j=1}^{G}(\hat{y_j})^{x_j}(1-\hat{y_j})^{1-x_j} \tag{1}$$

(b) The output of the encoder will be a mean matrix and variance matrix, which will follow z's dimension that is R^2

(c)

- $$logp_\theta(x) = log\sum_z q_\phi(z|x)(\frac{p_\theta(x,z)}{q_\phi(z|x)}) = log\int_z q_\phi(z|x)\frac{p_\theta(x,z)}{q_\phi(z|x)}dz \tag{2}$$

$$= log\mathbf{E}_{q_\phi(z|x)}[q_\phi(z|x)\frac{p_\theta(x,z)}{q_\phi(z|x)}] \geq \mathbf{E}_{q_\phi(z|x)}[q_\phi(z|x)log(\frac{p_\theta(x,z)}{q_\phi(z|x)})] \tag{3}$$

$$= \mathbf{E}_{q_\phi(z|x)}[q_\phi(z|x) * (log(\frac{p(z)}{q_\phi(z|x)}) + log(p_\theta(x|z)))] \tag{4}$$

$$= -KL(q_\phi(z|x), p(z)) + \mathbf{E}_{q_\phi(z|x)}[log(p_\theta(x|z))] \tag{5}$$

(d)

- **Non-negativity:**the KL divergent always non-negative
- **Asymmetric:**The KL divergent is not symmetric

(e) The two equation is different and both are dealing with different cases

- EQ1 is preferred because EQ1 doesn't need to compute log directly, which means more stable in most circumstances, and we can calculate the KL divergent by easier way
- some times EQ2 is preferred if P(z) is complex and difficult to analyze

(f)NO

if we choose $q(z) = N(0,1)$, then we are assume the latent variable are iid. It will lead to a wrong approximation of posterior distribution.

(g) It will be 0 because KL divergent is looking for the difference between two probability distributions. Since the two distributions are the same:

- $$KL(q(z|x), q(z|x)) = \int q(z|x)log\frac{q(z|x)}{q(z|x)} = 0 \tag{6}$$

(h)

- since $\sigma = \sigma_q = \sigma_p$, I will use σ only in the following steps.
-

$$KL(q_\phi(z|x), p(z)) = \int_z q(z|x) log \frac{q(z|x)}{p(z)} dz = \mathbf{E}_{q_\phi(z|x)}[log(q_\phi(z|x)) - log(p(z))] \quad (7)$$

$$= \mathbf{E}_{q_\phi(z|x)}[log(\frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{1}{2\sigma^2}(z-\mu_\phi)^2}) - log(\frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{1}{2\sigma^2}(z-\mu_p)^2})] \quad (8)$$

$$= \mathbf{E}_{q_\phi(z|x)}[log(\frac{1}{\sqrt{2\pi\sigma^2}}) + (-\frac{1}{2\sigma^2}(z-\mu_\phi)^2)) - log(\frac{1}{\sqrt{2\pi\sigma^2}}) + (-\frac{1}{2\sigma^2}(z-\mu_p)^2))] \quad (9)$$

$$= -\frac{1}{2\sigma^2}\mathbf{E}_{q_\phi(z|x)}[(z-\mu_\phi)^2 - (z-\mu_p)^2] = -\frac{1}{2\sigma^2}\mathbf{E}_{q_\phi(z|x)}[(\mu_\phi-\mu_\phi)^2 - (\mu_\phi-\mu_p)^2] \quad (10)$$

$$= \frac{1}{2\sigma^2}(\mu_\phi - \mu_p)^2 \quad (11)$$

(i) Let's state the new equation with Lagrangian multiplier first:

-

$$L(q_\phi(z|x), \lambda) = \sum_z q_\phi(z|x) log(p_\theta(x|z)) - KL(q_\phi(z|x), p(z)) + \lambda(\sum_z q_\phi(z|x) - 1) \quad (12)$$

$$= \sum_z q_\phi(z|x) log(p_\theta(x|z)) - q_\phi(z|x)(log(q_\phi(z|x)) - log(p(z))) + \lambda(\sum_z q_\phi(z|x)) - \lambda \quad (13)$$

- Then let's taking the derivative

$$\frac{\partial L}{\partial q} = \sum_z log(p_\theta(x|z)) - log(\frac{q_\phi(z|x)}{p(z)}) - 1 + \lambda = 0 \quad (14)$$

$$= \sum_z log(\frac{p_\theta(x|z)p(z)}{q_\phi(z|x)}) + \lambda - 1 = 0 \Rightarrow q_\phi(z|x) = \frac{p_\theta(x|z)p(z)}{e^{1-\lambda}} \quad (15)$$

- by give the constrain of $\sum_z q_\phi(z|x) = 1$

$$\Rightarrow \frac{\sum_z p_\theta(x|z)p(z)}{e^{1-\lambda}} = 1 \Rightarrow \lambda = 1 - log \sum_z p_\theta(x|z)p(z) \quad (16)$$

- let back to

$$\sum_z log(\frac{p_\theta(x|z)p(z)}{q_\phi(z|x)}) + \lambda - 1 = 0$$

$$\Rightarrow \sum_z (log(\frac{p_\theta(x|z)p(z)}{q_\phi(z|x)}) - log \sum_z p_\theta(x|z)p(z)) = 0 \tag{17}$$

$$\Rightarrow \sum_z log(\frac{p_\theta(x|z)p(z)}{q_\phi(z|x)\sum_z p_\theta(x|z)p(z)}) = 0 \Rightarrow \frac{p_\theta(x|z)p(z)}{q_\phi(z|x)\sum_z p_\theta(x|z)p(z)} = 1 \tag{18}$$

- Then we finally get

$$q_\phi(z|x) = \frac{p_\theta(x|z)p(z)}{\sum_z p_\theta(x|z)p(z)}$$

(j)

- $$\sum_z p_\theta(x|z)p(z) = \int_z p_\theta(x|z)p(z)dz = p_\theta(x)$$

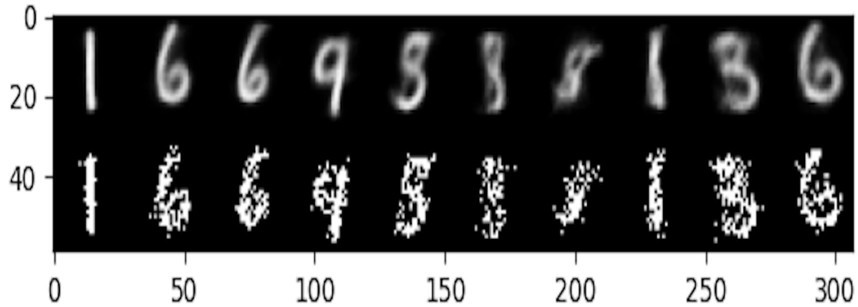
- $$p_\theta(x|z)p(z) = p_\theta(x, z)$$

- Therefore

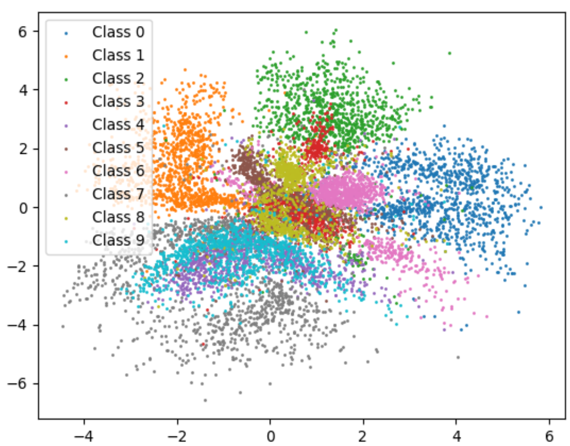
$$q_\phi(z|x) = \frac{p_\theta(x, z)}{p_\theta(x)} = p_\theta(z|x)$$

2 Variational Auto-Encoders [Coding]

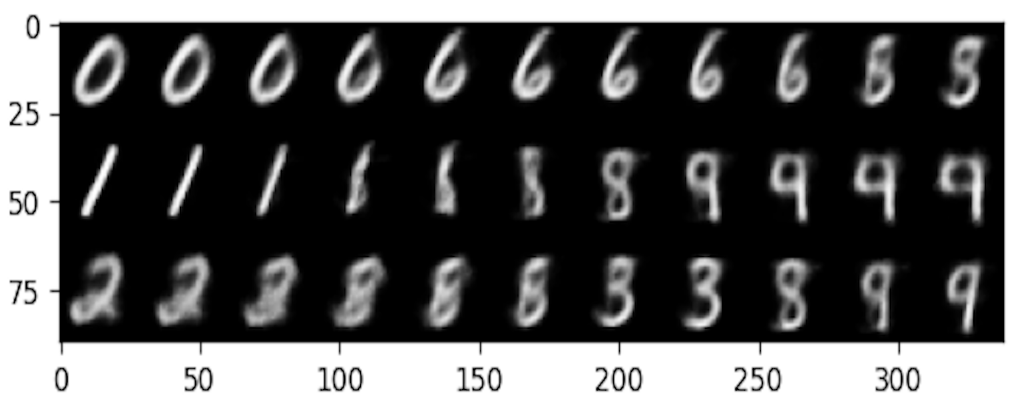
(e)



- (i)



- (ii)



- (iii)

3 Generative Adversarial Networks [Written]

(a) the classical cost function will be:

- $$\mathbb{E}_{x \sim p_{data}(x)}[\log D_\omega(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D_\omega(G_\theta(z)))] \tag{19}$$

(b) The new equation will be:

- $$\mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x \sim p_G(x)}[\log(1 - D(x))] \tag{20}$$

(c) since we are looking for the optimal discrimination, then the equation will be:

- $$-\int_x p_{data}(x) \log D(x) dx - \int_z p(z) \log(1 - D(x)) dz \tag{21}$$
- $$= -\int_x p_{data}(x) \log D(x) + p(z) \log(1 - D(x)) dx \tag{22}$$
- and we get: $$L(x, D, D^{\cdot}) = p_{data}(x) \log D(x) + p(z) \log(1 - D(x)) \tag{23}$$
- and, $S(D) = \int_x L(x, D, D^{\cdot}) dx$
- we need to find the stationary point, which means:
- Stationary D from: $$\frac{\partial L(x, D, D^{\cdot})}{\partial D} - \frac{d}{dx} \frac{\partial L(x, D, D^{\cdot})}{\partial D^{\cdot}} = 0 \tag{24}$$
- and we know $\frac{d}{dx} \frac{\partial L(x,D,D^{\cdot})}{\partial D^{\cdot}} = 0$, then:

$$\frac{\partial L(x, D, D^{\cdot})}{\partial D} = -\frac{p_{data}}{D} + \frac{P_G}{1 - D} = 0 \Rightarrow D^*(x) = \frac{p_{data}(x)}{p_{(data)}(x) + p_G(x)} \tag{25}$$

(d) The equation for an optimal generator will be:

- $$-\int_x p_{data}(x)\log D^*(x)dx + p_G(x)\log(1-D^*(x))dx \tag{26}$$

$$= -\int_x p_{data}(x)\log(\frac{p_{data}(x)}{p_{(data)}(x)+p_G(x)}) + p_G(x)\log(\frac{p_G(x)}{p_{(data)}(x)+p_G(x)})dx \tag{27}$$

$$= -2JSD(p_{data}, P_G) + log(4) \tag{28}$$

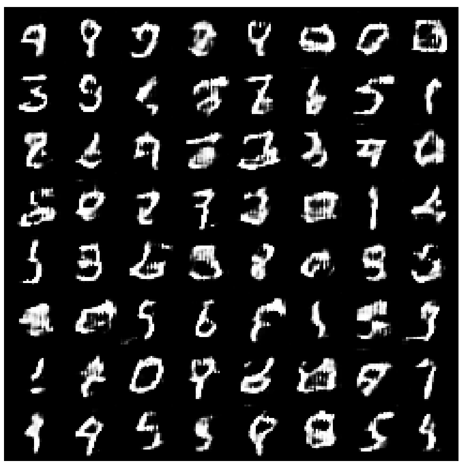
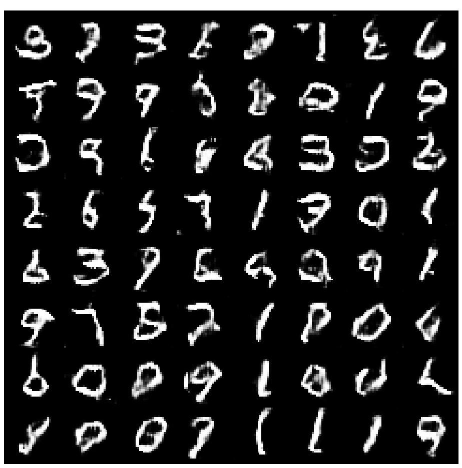
To max this equation, we need to make $JSD(p_{data}, P_G) = 0$

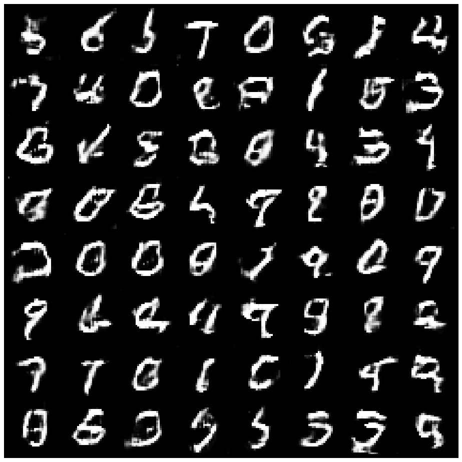
- since
$$JSD(p_{data}, P_G) = \frac{1}{2}(KL(p_{data}, M)) + \frac{1}{2}(KL(p_G, M)) \tag{29}$$
- since $M = \frac{1}{2}(p_{data} + p_G)$, then $JSD(p_{data}, P_G) = 0$, if $p_{data} = p_G$
- Therefore, $p_G^* = p_{data}$

(e)

- When $P_2 = 0$
$$KL(P_1, P_2) = \int P_1 log\frac{P_1}{P_2} = \inf$$
- When $P_3 = 0$
$$KL(P_1, P_3) = \int P_1 log\frac{P_1}{P_3} = \inf$$
- $$W(P_1, P_2) = \int_0^{0.5} xdx + \int_{0.5}^1 x-(x-0.5)dx + \int_1^{1.5}(1-(x-0.5))dx = 0.5 \tag{30}$$
- $$W(P_1, P_3) = \int_0^1 xdx + \int_1^2 (1-(x-1))dx = 1 \tag{31}$$

4 Generative Adversarial Networks [Coding]

- 10 epochs 
- 30 epochs 
- 50 epochs 

- 70 epochs 
- 90 epochs