ECE 544- Homework 5

Ahmadreza Eslaminia (Ae15)

Questions:

Problem 1:

Ahmadreza Eslaminia ECE 544 HW#5

AE15

1.) [VAE,] Jason

a) $\log P_\theta(x) = \log \sum_z q(z|x) \frac{P_\theta(x,z)}{q(z|x)} \geq \sum_z q(z|x) \log \frac{P_\theta(x,z)}{q(z|x)}$

$\Rightarrow \log P_\theta(x) \geq \sum_z q(z|x) \log \frac{P_\theta(x|z) P(z)}{q(z|x)} = \sum_z q(z|x) \log \frac{P(z)}{q(z|x)} + \sum_z q(z|x) \log P_\theta(x|z)$

$= -KL(q(z|x), P(z)) + \mathbb{E}_{q(z|x)}[\log P_\theta(x|z)]$

b) $KL(q(z|x), P(z)) \geq 0 \rightarrow$ non-negative, equal to zero when $P(z) = q(z|x)$

Also, $KL(q(z|x), P(z))$ is convex in pair of probability mass function $(q(z|x), P(z))$

meaning: $KL(\lambda q_1(z|x) + (1-\lambda) q_2(z|x), \lambda P_1(z) + (1-\lambda) P_2(z))$

$\leq \lambda KL(q_1(z|x), P_1(z)) + (1-\lambda) KL(q_2(z|x), P_2(z))$

c) $KL(q(z|x), q(z|x))$ if $q(z|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(z-\mu_z)^2\right)$

$\hookrightarrow = \sum_z q(z|x) \log \frac{q(z|x)}{q(z|x)} = \sum_z q(z|x) \log 1 = 0 \rightarrow KL = 0$

it make sense bc we know when two q and P are same the KL between them is zero cause its divergence between those probabilities.

P_1

1. d) $p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} (z - \mu_p)^2 \right) \rightarrow KL(q(z|x), p(z))$?

$$KL(q(z|x), p(z)) = \sum_z q(z|x) \log \frac{q(z|x)}{p(z)} = \sum_z q(z|x) \log \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(z-\mu_q)^2\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(z-\mu_p)^2\right)}$$

$$= \sum_z q(z|x) \left[\frac{-1}{2\sigma^2} \left[(z-\mu_q)^2 - (z-\mu_p)^2 \right] \right] =$$

$$= \frac{-1}{2\sigma^2} \left[\sum q(z|x)(z-\mu_q)^2 - \sum_z q(z|x)(z-\mu_p)^2 \right]$$

$$= \frac{-1}{2\sigma^2} \left[\sum_z q(z|x) z^2 - 2\mu_q \sum_z q(z|x) z + \mu_q^2 \sum_z q(z|x) - \sum q(z|x) z^2 \right.$$

$$\left. + 2\mu_p \sum_z q(z|x) z - \mu_p^2 \sum_z q(z|x) \right]$$

$$= \frac{-1}{2\sigma^2} \left[-\mu_p^2 + \mu_q^2 - 2\mu_q^2 + 2\mu_q\mu_p \right] = \frac{(\mu_p - \mu_q)^2}{2\sigma^2}$$

P3

1. e) $\sum_z q(z|x)\log p_\theta(x|z) - KL(q(z|x), p(z))$ s.t. $\sum_z q(z|x) = 1$

$$L(z,\lambda) = \sum_z q(z|x)\log p_\theta(x|z) + \sum_z q(z|x)\log\frac{p(z)}{q(z|x)} + \lambda\left(\sum_z q(z|x) - 1\right)$$

$$\Rightarrow \frac{\partial L}{\partial q(z|x)} = \log p_\theta(x|z) + \log\frac{p(z)}{q(z|x)} + 1 - \lambda = 0$$

$$\Rightarrow \log q(z|x) = \lambda - 1 + \log p_\theta(x|z)\,p(z) \Rightarrow q(z|x) = e^{(\lambda-1)} \quad \left(\sum_z p_\theta(x|z)\,p(z)\right) = 1$$

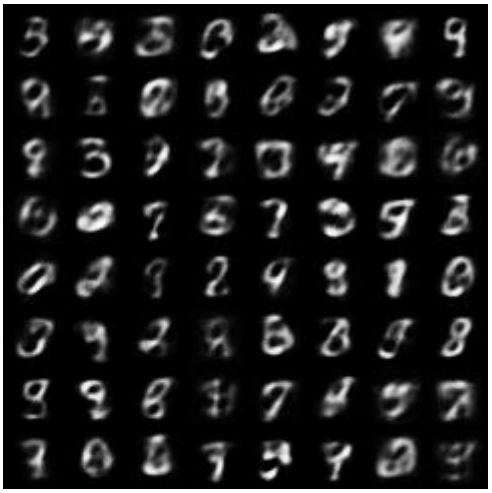
$$\Rightarrow e^{\lambda-1} = \frac{1}{\sum_z p_\theta(x|z)\,p(z)} \Rightarrow q(z|x) = \frac{p_\theta(x|z)\,p(z)}{\sum_z p_\theta(x|z)\,p(z)}$$

f) $$q(z|x) = \frac{p_\theta(x|z)\,p(z)}{\sum_z p_\theta(x|z)\,p(z)} = \frac{p_\theta(x|z)\,p(z)}{p_\theta(x)} = p_\theta(z|x)$$

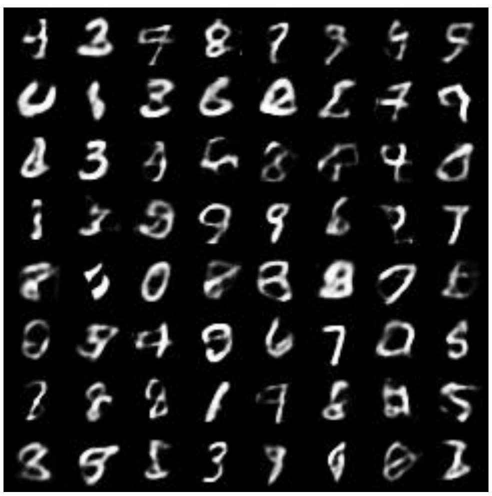
g) →
```
std = torch.exp(0.5 * logvar)
eps = torch.randn_like(std)
return mu + eps * std
```

loss: 102.9940

Fig: epoch_0



Epoch 29:



Problem 2:

2. [VAEs]

a) ELBO is: $L(\theta, \varphi, x_i) = -D_{KL}(q_\varphi(z|x_i) \| p(z)) + E_{q(\varphi)}[\log p_\theta(x_i|z)]$

so we need to minimize following program for VAEs for dataset $X = \{x_i\}_{i=1}^{N}$

max L
↓
$X(-)$ $\Rightarrow \min \sum_{i=1}^{N} \left(E_{q_\varphi(z,x_i)}[-\log p_\theta(x_i|z)] + D_{KL}(q_\varphi(z|x_i) \| p(z)) \right)$

reconstruction error b) $\rightarrow \sum_{i=1}^{N} E_{q(\varphi)(z|x_i)}[-\log p_\theta(x_i|z)]$

" c) drawing M sample from $q(\varphi)(z|x_i)$

$E_q(z|x_i) \longrightarrow \frac{1}{M} \sum_{m=1}^{M} (-\log p_\theta(x_i|z_i^m))$

d) assuming $\rightarrow$ $p_\theta(x_i|z) \sim N(f(z), \sigma^2 I)$. σ cons $I \rightarrow D \times D$

$\Rightarrow \sum_{i=1}^{N} \frac{1}{M} \sum_{m=1}^{M} (-\log p_\theta(x_i|z_{i,m})) = \sum_i \frac{1}{M} \sum_m \left(-\log \left(\frac{1}{\sqrt{2\pi\sigma^D}} \left(\exp\frac{\|x_i - f(z_{i,m})\|^2}{2\sigma^2} \right) \right) \right)$

$= \sum_{i=1}^{N} \frac{1}{M} \sum_{m=1}^{M} -\log \frac{1}{\sqrt{2\pi}\,\sigma D} + \frac{\|x_i - f(z_{i,m})\|_2^2}{2\sigma^2}$

σ, D constant can be ignored $\Rightarrow$ empirical $= \boxed{\frac{1}{M} \sum \frac{1}{2\sigma^2} \| x_i - f(z_{i,m}) \|_2^2}$

2.e) if all data points are binary $\forall i \quad x_i \in \{0,1\}^D$ $p_\theta(x_i|z)$?

$g = f(z) \in [0,1]^D$

$$\log p_\theta(x_i|z) = \sum_{d=1}^{D} x_i^{(d)} \log g^{(d)} + (1-x_i^{(d)}) \log(1-g^{(d)})$$

$$\Rightarrow p_\theta(x_i, z) = \exp\left(\sum_{d=1}^{D} x_i^{(d)} \log g^{(d)} + (1-x_i^{(d)}) \log(1-g^{(d)})\right)$$

Name: Multivariate Bernoulli dist.

Scanned with CamScanner

P_5

Problem 3:

3. [GANs and Duality]

a) 1. If discriminator is too good, training Gn fail due to vanishing gradients since optimal discriminator doesn't provid enough info to Generator to train

2. GAN convergence is hard to ~~[illegible]~~ Idntify.

b) $p_w(y=1|x) = \frac{1}{1+\exp(w^T x)}$

$$\rightarrow \max_\theta \min_w -\sum_{x\in D} \log\left(\frac{1}{1+e^{w^T x}}\right) - \sum \log\left(1-\frac{1}{1+e^{w^T G_\theta(z)}}\right) + \frac{c}{2}\|w\|_2^2$$

$$\Rightarrow \max_\theta \min_w \sum_{x\in D} \log\left(1+\exp(w^T x)\right) + \sum_{z\in Z} \log\left(\frac{1+\exp(w^T G_\theta(z))}{\exp(w^T G_\theta(z))}\right) + \frac{c}{2}\|w\|_2^2$$

$$= \max_\theta \min_w \sum_{x\in D} \log\left(1+\exp(w^T x)\right) + \sum_{z\in Z} \log\left(1+\exp(w^T G_\theta(z))\right) - \sum_{z\in Z} w^T G_\theta(z) + \frac{c}{2}\|w\|_2^2$$

c) when $\frac{c}{2}\|a\|_2^2 - a^T b$ convex in a? why!

$$f(a) = \frac{c}{2}\|a\|_2^2 - a^T b \rightarrow \frac{\partial f}{\partial a} = c a^T - b^T \rightarrow H = \frac{\partial^2 f}{\partial a^2} = cI \rightarrow c \geq 0 \rightarrow \forall_x \; x^T H x = c\|x\|_2^2 \geq 0$$

$\Rightarrow f$ convex when $c \geq 0$

Scanned with CamScanner

P6

3.d) $\log(1+\exp^{a^Tb})$ convex? why?

$f(a) = \log\{1+\exp(a^Tb)\} \rightarrow \frac{\partial f(a)}{\partial a} = \frac{\exp(a^Tb)}{1+\exp(a^Tb)} b^T \qquad H = \frac{\partial^2 f}{\partial a^2}$

$\Rightarrow H = \frac{\exp^{a^Tb}(1+\exp^{a^Tb}) - (\exp^{a^Tb})^2}{(1+\exp^{a^Tb})^2} bb^T = \frac{\exp^{a^Tb}}{(1+\exp^{a^Tb})^2} bb^T$

$x^T H x = \frac{\exp^{a^Tb}}{(1+\exp^{a^Tb})^2} x^T bb^T x = \frac{\exp^{a^Tb}}{(1+\exp^{a^Tb})^2} \|b^T x\|_2^2 \geq 0 \quad \forall x$

$\Rightarrow$ H is semi positive $\Rightarrow$ f always convex.

e) so in the domain that $c \geq 0$ we have ~~convex~~

$\sum \log(1+\exp^{w^Tx})$ convex according to d

$\sum \log(1+\exp^{w^T G_\theta(z)})$ convex " to d

$\sum_{z\in Z} w^T G_\theta(z) + \frac{c}{2}\|w\|_2^2$ convex when $c \geq 0$ according to c

Function is convex in C

f) $J_x = w^T x \qquad J_z = w^T G_\theta(z)$

lagrangian

$\Rightarrow L = \sum_{x\in D} \log(1+\exp J_x) + \sum_{z\in Z} \log(1+\exp J_z) - \sum_{z\in Z} w^T G_\theta(z) + \frac{c}{2}\|w\|_2^2$

$+ \sum_{x\in D} \lambda_x (J_x - w^T x) + \sum_{z\in Z} \lambda_z (J_z - w^T G_\theta(z))$

P7

g) $f = \min \frac{c}{2}\|w\|_2^2 - w^T b$ in c, b?

$w^* \to \frac{\partial f}{\partial w} = c w^T - b^T = 0 \Rightarrow w^* = \frac{1}{c} b$

$\min \frac{c}{2}\|w\|_2^2 - w^T b = \frac{c}{2}\|w^*\|_2^2 - w^{*T} b = \frac{c}{2}\left(\frac{1}{c}\right)^2 \|b\|_2^2 - \frac{1}{c}\|b\|_2^2$

$= \frac{-1}{2c}\|b\|_2^2$

h) $\min \lambda y + \log^{1+\exp y}$ in λ? domain?

$\frac{\partial}{\partial y} = 0 \Rightarrow \frac{\partial}{\partial y} = \lambda + \frac{\exp y}{1+\exp y} = 0 \Rightarrow \exp y^* = \frac{-\lambda}{\lambda+1} \Rightarrow y^* = \log \frac{-\lambda}{\lambda+1}$

$\boxed{\Rightarrow -1 < \lambda < 0}$

$\min \lambda y + \log^{1+\exp y} = \lambda \log\left(\frac{-\lambda}{1+\lambda}\right) + \log\left(1 - \frac{\lambda}{1+\lambda}\right) = \lambda\left(\log^{-\lambda} - \log^{\lambda+1}\right) - \log^{1+\lambda}$

$= -(1+\lambda)\log^{1+\lambda} + \lambda \log^{-\lambda}$

P8

$$3.i)\ L(w, y_x, y_z, \lambda_x, \lambda_y) = \sum \log^{(1+\exp)y_x} + \sum_{z\in Z} \log^{(1+\exp)y_z} - \sum_{z\in Z} w^T G_\theta(z)$$

$$+ \frac{c}{2}\|w\|_2^2 + \sum_{x\in D} \lambda_x (y_x - w^T x) + \sum_{z\in Z} \lambda_z (y_z - w^T G_\theta(z))$$

$$= \sum_{x\in D} (\lambda_x y_x + \log^{(1+\exp y_x)}) + \sum_{z\in Z} (\lambda_z y_z + \log^{1+\exp y_z}) + \frac{c}{2}\|w\|_2^2$$

$$- w^T \Big(\sum_{z\in Z} G_\theta(z) + \sum_{x\in D} \lambda_x x + \sum_{z\in Z} \lambda_z G_\theta(z) \Big)$$

From $g \rightarrow w_z^* = \frac{1}{c} \Big(\sum_{z\in Z} (\lambda_z + 1) G_\theta(z) + \sum_{x\in D} \lambda_x x \Big)$

From $h \rightarrow \frac{\partial}{\partial y_x} = 0 \rightarrow y_x^* = \log \left(\frac{-\lambda_x}{1+\lambda_x} \right) \ \forall x \in D$, $y_z^* = \log \frac{-\lambda_z}{1+\lambda_z} \ \forall z \in Z$

Dual $\Rightarrow g(\lambda_x, \lambda_z) = \frac{-1}{2c} \Big\| \sum_{z\in Z} (\lambda_z + 1) G_\theta(z) + \sum_{x\in D} \lambda_x x \Big\|_2^2 + \sum_{z\in Z} \Big(\lambda_z \log^{-\lambda_z} - (\lambda_z + 1) \log^{\lambda_z + 1}$

$$+ \sum_{x\in D} \lambda_x \log^{-\lambda_x} - (\lambda_x + 1) \log^{\lambda_x + 1}$$

Dual Program: $\max g(\lambda_x, \lambda_z)$ s.t. $-1 \leq \lambda_x < 0 \quad \forall x \in D$

$-1 \leq \lambda_z < 0 \quad \forall z \in Z$

here we maximizing $fun(\theta, \lambda_x, \lambda_y)$ so we solved the synchronization problem between Discriminator and Generator $\rightarrow$ also we max a function which means we solved the vanishing gradient problem

P_3

Problem 4:

4. [GAN]

a) VAEs are probabilistic models that learn a latent space and uses reconstruc loss and divergence. under. than smoother output + more stable training process.
Gan → game theorical approach → two competing network → more realistic output while harder to train.

b) $\max_{\theta} \min_{\omega} - \sum_{x} \log D_\omega(x) - \sum_{z} \log(1 - D_\omega(G_\theta(z)))$

c) data domain induced by Gen → $P_{G(x)}$

⇒ equivalent Prob: $- \sum_{x} \log D_\omega(x) - \sum \log(1 - D_\omega(G_\theta(z)))$

⇒ $\max_{P_G} \min - \int_x P_{data}(x) \log D(x) dx - \int_x P_G(x) \log(1 - D(x)) dx$ (I)

d) $\frac{\partial L(x, D, \dot{D})}{\partial D} - \frac{d}{dx} \frac{\partial L(x, D, \dot{D})}{\partial \dot{D}} = 0$ where $\dot{D} = \frac{\partial D}{\partial x}$

give station point → $S(D) = \int_x L(x, D, \dot{D}) dx$ P_D, P_G

(I) ⇒ $\frac{\partial L(x, D, \dot{D})}{\partial D} = -\frac{P_{data}}{D} - \frac{P_G}{D - 1} = 0$ ⇒

$$\boxed{\Rightarrow D^*(x) = \frac{P_{data}}{P_G + P_{data}}}$$ (II)

P12

4.e) show $G^*(x) \Rightarrow P_G^* = P_{data}$.

$$JSD(P_{data}, P_G) = \frac{1}{2} D_{KL}(P_{data}, M) + \frac{1}{2} D_{KL}(P_G, M) \quad M = \frac{1}{2}(P_G + P_{data})$$

I $\rightarrow -\int_x P_{data}(x) \log D^*(x) + P_G(x) \log (1 - D^*(x)) \, dx$

II $\rightarrow -\int_x P_{data} \log \frac{P_{data}(x)}{P_{data}(x) + P_G(x)} + P_G(x) \log \frac{P_G(x)}{P_{data}(x) + P_G(x)} \, dx$

$= \log 4 - 2\, JSD(P_{data}, P_G)$

$JSD > 0 \xrightarrow{min} P_{data} = P_G$ ✓ III

f) according to III and II from previous sections:

$$D^*(x) = \frac{P_{data}}{P_{data} + P(G)} = \frac{1}{2} = 0.5 \checkmark$$