به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

مدل‌های مولد عمیق

تمرین شماره ۲

نام و نام خانوادگی	عرفان باقری سولا
شماره دانشجویی	۸۱۰۱۹۸۳۶۱
تاریخ ارسال گزارش	۸ آذر ۱۴۰۲

فهرست گزارش سوالات

سوال ۱

الف)

$$q(z|x) = \mathcal{N}\left(z; \mu(x), diag\left(\sigma^2(x)\right)\right)$$

$$p(z) = \mathcal{N}(z; 0, I)$$

$$KL\left(q(z|x) \parallel p(z)\right) = E_q\left[\log \frac{q(z|x)}{p(z)}\right] = E_q[\log q(z|x) - \log p(z)]$$

$$= -\frac{1}{2} E_q\left[\left(z-\mu(x)\right)^T diag\left(\sigma^2(x)\right)^{-1}\left(z-\mu(x)\right) + \log\left(\det\left(diag\left(\sigma^2(x)\right)\right)\right) + k\log(2\pi) - z^T z - k\log(2\pi)\right]$$

$$= -\frac{1}{2} E_q\left[\left(z-\mu(x)\right)^T diag\left(\sigma^2(x)\right)^{-1}\left(z-\mu(x)\right) + \log\left(\det\left(diag\left(\sigma^2(x)\right)\right)\right) - z^T z\right]$$

$$= -\frac{1}{2} E_q\left[\sum_{i=1}^{k} \frac{\left(z_i-\mu_i(x)\right)^2}{\sigma_i^2(x)} + \log\left(\prod\nolimits_{i=1}^{k} \sigma_i^2(x)\right) - \sum_{i=1}^{k} z_i^2\right]$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{k} \frac{E_q\left[\left(z_i-\mu_i(x)\right)^2\right]}{\sigma_i^2(x)} + \log\left(\prod\nolimits_{i=1}^{k} \sigma_i^2(x)\right) - \sum_{i=1}^{k} E_q\left[z_i^2\right]\right)$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{k} \frac{\sigma_i^2(x)}{\sigma_i^2(x)} + \sum_{i=1}^{k} \log\left(\sigma_i^2(x)\right) - \sum_{i=1}^{k}\left(\sigma_i^2(x) + \mu_i^2(x)\right)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{k}\left(1 + \log\left(\sigma_i^2(x)\right) - \sigma_i^2(x) - \mu_i^2(x)\right)$$

$$= \frac{1}{2}\sum_{i=1}^{k}\left(\sigma_i^2(x) + \mu_i^2(x) - 1 - \log \sigma_i^2(x)\right)$$

ب)

$$\mathcal{L}_m = E_{q(z|x)}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m} \frac{p\left(x, z^{(i)}\right)}{q\left(z^{(i)}|x\right)}\right)\right]$$

Following Jensen’s inequality and the fact that log is a concave function:

$$E_{q(z|x)}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m} \frac{p\left(x, z^{(i)}\right)}{q\left(z^{(i)}|x\right)}\right)\right] \le \log\left(E_{q(z|x)}\left[\frac{1}{m}\sum_{i=1}^{m} \frac{p\left(x, z^{(i)}\right)}{q\left(z^{(i)}|x\right)}\right]\right)$$

$$= \log\left(\frac{1}{m}\sum_{i=1}^{m} E_{q(z|x)}\left[\frac{p\left(x, z^{(i)}\right)}{q\left(z^{(i)}|x\right)}\right]\right) = \log\left(\frac{1}{m}\sum_{i=1}^{m} \int q\left(z^{(i)}|x\right)\frac{p\left(x, z^{(i)}\right)}{q\left(z^{(i)}|x\right)} dz^{(i)}\right)$$

$$= \log\left(\frac{1}{m}\sum_{i=1}^{m} p(x)\right) = \log p(x)$$

$$\Longrightarrow \mathcal{L}_m \leq \log p(x)$$

Proof of Jensen’s inequality:

$$\forall x, x_0: \ \ \log(x) \leq \log(x_0) + \left[\frac{d}{dx}\log(x)\right]_{x_0} (x - x_0)$$

$$\log(x) \leq \log(x_0) + b(x - x_0)$$

$$x_0 = E[x] \rightarrow \ \ \log(x) \leq \log(E[x]) + b(x - E[x])$$

$$E[\log(x)] \leq E[\log(E[x]) + b(x - E[x])]$$

$$E[\log(x)] \leq E[\log(E[x])] + E[b(x - E[x])]$$

$$E[\log(x)] \leq \log(E[x]) + b(E[x] - E[x])$$

$$E[\log(x)] \leq \log(E[x])$$

ج)

پدیده posterior collapse زمانی اتفاق می‌افتد که توزیع تابع variational به سمت توزیع prior فرو بریزد یا به زبان ساده‌تر اطلاعات ورودی x در z بسیار ضعیف باشد و به خوبی encode نشده باشد.

$$q_\phi(z|x) \approx p(z)$$

در نتیجه این اتفاق decoder شروع به نادیده گرفتن z می‌کند و خروجی آن به نوعی مستقل از z می‌شود.

$$p_\theta(x|z) \approx p_\theta(x)$$

یکی از رایج ترین روش‌ها برای مقابله با این پدیده استفاده از weight annealing روی عبارت KL در loss است. همچنین یکی از مقالات پیشنهاد کرده بود که هر موقع گرادیان عبارت KL در هر dimension کمتر از یک مقداری شد، آن را نادیده بگیریم و اعمال نکنیم. بعضی از مقالات هم استفاده از یک lower bound یا objective دیگر را پیشنهاد داده‌اند.

د)

$$KL\left(q_\phi(z|x) \parallel p(z|x)\right) = KL\left(q_\phi(z|x) \parallel p(z,x)\right) + \log p(x) \geq 0$$

$$-KL\left(q_\phi(z|x) \parallel p(z,x)\right) \leq \log p(x)$$

$$-E_{q_\phi(z|x)}\left[\log\frac{q_\phi(z|x)}{p_\theta(x|z)p(z)}\right] = -E_{q_\phi(z|x)}\left[\log\frac{q_\phi(z|x)}{p(z)} - \log p_\theta(x|z)\right] \leq \log p(x)$$

$$-KL\left(q_\phi(z|x) \parallel p(z)\right) + E_{q_\phi(z|x)}[\log p_\theta(x|z)] \leq \log p(x)$$

سوال ۲

الف)

کد مربوط به reparameterization trick به صورت زیر پیاده سازی شده است:

```
z = torch.normal(0, 1, size=v.shape).to(device)
z = m + torch.sqrt(v) * z
return z
```

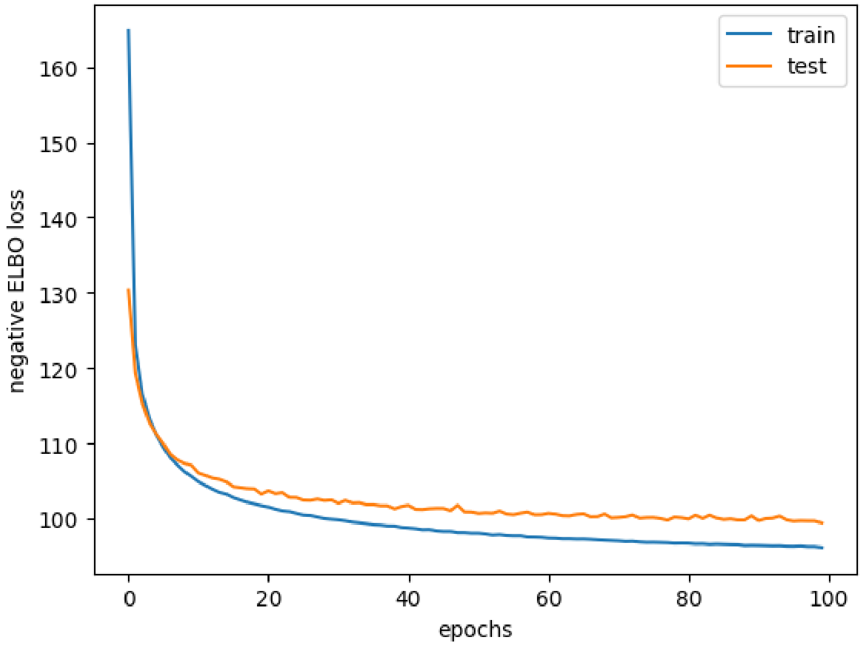
ب)

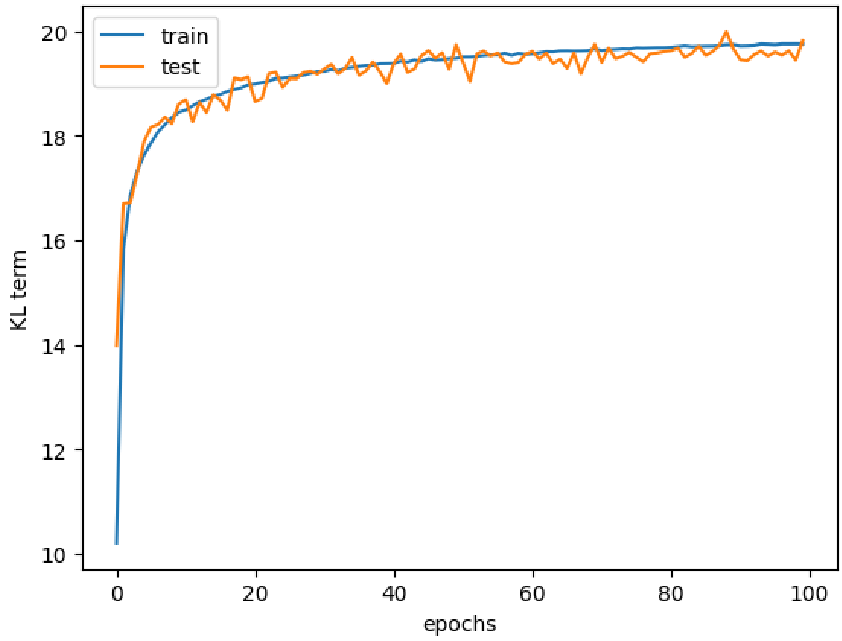
کد بخش مربوط به negative ELBO نیز به صورت زیر پیاده سازی شده است. همچنین برای محاسبه reconstruction loss از روش Monte Carlo sampling استفاده شده است.

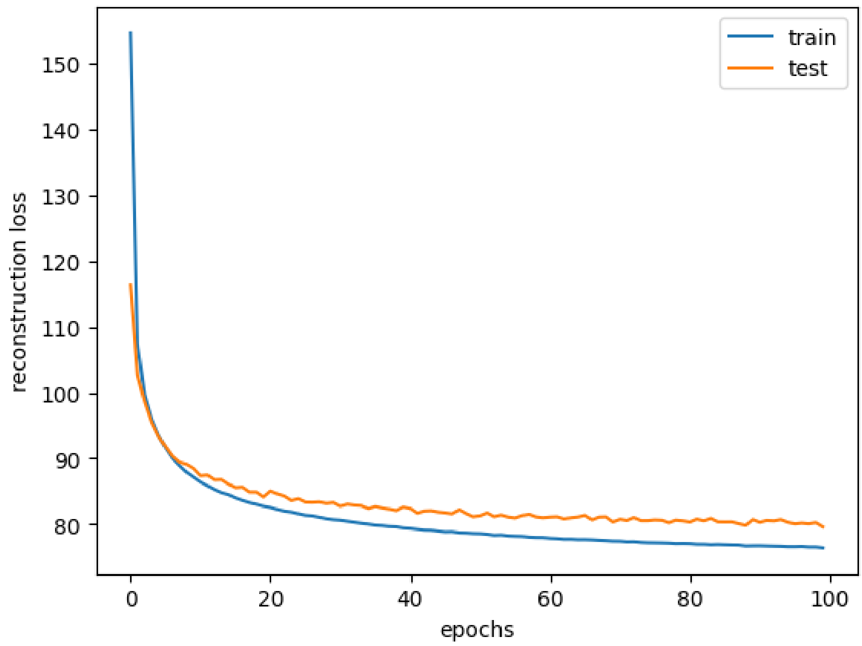
```
m, v = self.encoder(x)
z = sample_gaussian(m, v)
logits = self.decoder(z)
rec = -log_bernoulli_with_logits(x, logits).mean()
kl = kl_normal(m, v, self.z_prior_m, self.z_prior_v).mean()
nelbo = kl + rec
return nelbo, kl, rec
```

ج)

مدل ساخته شده را به مدت ۱۰۰ epoch آموزش می‌دهیم.







خوب است به این نکته توجه کنیم که مقدار KL divergence در حال افزایش است و این یعنی رفته توزیع $q_\phi(z|x)$ از prior که یک توزیع نرمال خالص است دور می‌شود تا بتواند با افزایش این مقدار، اطلاعات مفیدی را در فضای پنهان کد گذاری بکند و در عوض مقدار reconstruction loss را به میزان قابل توجهی کاهش داده و در نهایت باعث کم شدن loss کل شود.

د)

برای پیاده سازی کد مربوط به کران IWAE به صورت زیر عمل می‌کنیم:

$$\mathcal{L}_m = E_{q(z|x)}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m}\frac{p(x,z^{(i)})}{q(z^{(i)}|x)}\right)\right] = E_{q(z|x)}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m}\frac{p(x|z^{(i)})p(z^{(i)})}{q(z^{(i)}|x)}\right)\right]$$

$$\mathcal{L}_m = E_{q(z|x)}\left[\log\left(\sum_{i=1}^{m}\frac{p(x|z^{(i)})p(z^{(i)})}{q(z^{(i)}|x)}\right)\right] - \log(m)$$

$$\mathcal{L}_m = E_{q(z|x)}\left[\log\left(\sum_{i=1}^{m}\exp\left(\log p(x|z^{(i)}) + \log p(z^{(i)}) - \log q(z^{(i)}|x)\right)\right)\right] - \log(m)$$

```
m, v = self.encoder(x)
m_up = m.unsqueeze(1).repeat(1, iw, 1)
v_up = v.unsqueeze(1).repeat(1, iw, 1)
x_up = x.unsqueeze(1).repeat(1, iw, 1)
z = sample_gaussian(m_up, v_up)
logits = self.decoder(z)

log_p_x_g_z = log_bernoulli_with_logits(x_up, logits)
log_prior_z = log_normal(z, torch.tensor(0), torch.tensor(1))
log_q_z_g_x = log_normal(z, m_up, v_up)
log_w = log_p_x_g_z + log_prior_z - log_q_z_g_x
log_marginal_likelihood = log_mean_exp(log_w, 1).mean()

niwae = -log_marginal_likelihood
kl = kl_normal(m, v, torch.tensor(0), torch.tensor(1)).mean()
rec = -log_p_x_g_z.mean()
return niwae, kl, rec
```

ه)

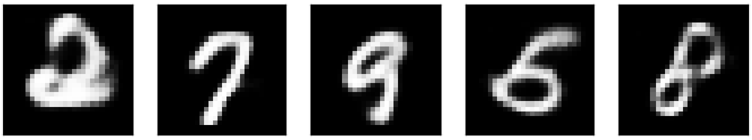
به منظور ارزیابی کران IWAE ، میانگین مقدار negative iwae bound روی داده های تست را به ازای m های گفته شده به دست می‌آوریم. مشاهده می‌شود که همه آنها کران دقیق تری نسبت به ELBO هستند.

```
1- negative ELBO: 99.42

2- negative IWAE bound:

   k = 5    :  97.75
   k = 50   :  96.59
   k = 150  :  96.30
```

همچنین ویژگی زیر که یکی از ویژگی های ذکر شده در مقاله مربوطه بود نیز برقرار است:

$$\mathcal{L}_5 \le \mathcal{L}_{50} \le \mathcal{L}_{150} \le \log p(x)$$

در ادامه چند نمونه از تصاویر تولید شده توسط VAE را نیز مشاهده می‌کنیم:



سوال ۳

الف)

$$x = f(z) = 1.5x + 3 \rightarrow z = f^{-1}(x) = g(x) = \frac{x-3}{1.5}$$

$$p_x(x) = p_z(f^{-1}(x))\left|\frac{\partial g}{\partial x}\right| = \frac{p_z\left(\frac{x-3}{1.5}\right)}{1.5} = \frac{2}{3}\mathcal{N}(\mu = 3, \sigma = 1.5)$$

ب)

مقاله مطالعه گردید.

ج)

$$x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}, y = \begin{bmatrix} y_a \\ y_b \end{bmatrix}$$

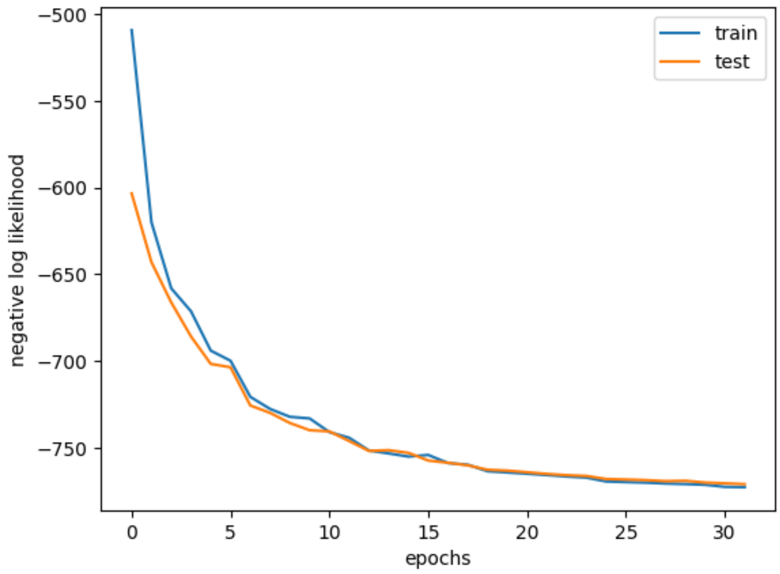
$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_a}{\partial x_a} & \frac{\partial y_a}{\partial x_b} \\ \frac{\partial y_b}{\partial x_a} & \frac{\partial y_b}{\partial x_b} \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 \\ \frac{\partial y_b}{\partial x_a} & diag\left(\exp\left(s(x_a)\right)\right) \end{bmatrix}$$

$$\det\left(\frac{\partial y}{\partial x}\right) = \prod_i \exp(s(x_a)_i) = \exp\left(\sum_i s(x_a)_i\right)$$

د)

در این قسمت تبدیل کوپلینگ را در یک ماژول pytorch پیاده سازی می‌کنیم. در این ماژول به جای تقسیم تنسور ورودی به دو قسمت و سپس concatenate کردن آن، از یک mask باینری استفاده می‌کنیم که اگر ورودی در این mask ضرب شود یک بخش از ورودی و اگر در معکوس آن ضرب شود بخش دیگر ورودی را می‌سازد.

مدل را با هدف maximize کردن log likelihood داده ها در ۳۲ epoch آموزش می‌دهیم.



سپس به صورت تصادفی ۱۶ تصویر را با استفاده از مدل تولید می‌کنیم که در ادامه قابل مشاهده است.