1

20170243 심재윤

$$\log p_\theta(x) = \log\left(\mathbb{E}_{z\sim q_\phi(z|x)}\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]\right)$$

(a) $KL(q_\phi(z|x), p_\theta(z))$

Traing VAE $\Leftrightarrow$ Maximization ELBO

$$\arg\max_{\theta,\phi} \sum_{i=1}^{N} \mathcal{L}(x^{(i)}, \theta, \phi)$$

$$\log p_\theta(x) \geq \mathcal{L}(x,\theta,\phi)$$

$$= \mathbb{E}_{z\sim q_\phi(z|x)}\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]$$

$$= \mathbb{E}_z\left[\log p_\theta(x|z)\right] - KL\left(q_\phi(z|x), p_\theta(z)\right)$$

(b) $q_\phi(z|x) = N(\mu_{z|x}, \sigma^2_{z|x})$, $p_\theta(z) = N(\mu, \sigma^2)$

$KL(q_\phi(z|x), p_\theta(z))$ has an analytic solution if both q and p follows the normal distribution. And all of integrals are about marginalization over z.

$$\begin{cases} q_\phi(z|x) = N(\mu_{z|x}, \sigma^2_{z|x}) = \frac{1}{\sqrt{2\pi\sigma^2_{z|x}}} \exp\left(-\frac{(z-\mu_{z|x})^2}{2\sigma^2_{z|x}}\right) \\ p_\theta(z) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) \end{cases}$$

$$\rightarrow KL(q_\phi(z|x), p_\theta(z)) = \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z)} dz$$

$$= \int q_\phi(z|x) \log q_\phi(z|x) dz - \int q_\theta(z|x) \log p_\theta(z) dz$$

$$= \underbrace{\int N(\mu_{z|x}, \sigma^2_{z|x}) \log N(\mu_{z|x}, \sigma^2_{z|x}) dz}_{①} \underbrace{- \int N(\mu_{z|x}, \sigma^2_{z|x}) \log N(\mu, \sigma^2) dz}_{②}$$

$$① = \int q_\phi(z|x) \log q_\phi(z|x) dz$$

$$= \int A_q \exp\left(-\frac{(z-\mu_{z|x})^2}{2\sigma^2_{z|x}}\right)\left(\log A_q + -\frac{(z-\mu_{z|x})^2}{2\sigma^2_{z|x}}\right) dz \cdots$$

$$= \int A_q \exp(-B_q^2)\left(\log A_q - B_q^2\right)\sqrt{2}\sigma_{z|x}\, dB_q \cdots$$

$$= \int A_q \exp(-B_q^2)\sqrt{2}\sigma_{z|x} \cdot \log A_q \, dB_q - \int A_q \exp(-B_q^2)\sqrt{2}\sigma_{z|x} B_q^2 \, dB_q$$

$$\frac{1}{\sqrt{2\pi\sigma^2_{z|x}}} = A_q$$

$$\frac{z-\mu_{z|x}}{\sqrt{2}\sigma_{z|x}} = B_q$$

$$\frac{1}{\sqrt{2}\sigma_{z|x}} dz = dB_q$$

$$= \int \frac{\sqrt{2}\,\sigma_{z|x}}{\sqrt{2\pi}\,\sigma_{z|x}} \exp(-\beta_\phi^2) \log\left(2\pi\sigma_{z|x}^2\right)^{-\frac{1}{2}} d\beta_\phi - \int \frac{\sqrt{2}\,\sigma_{z|x}}{\sqrt{2\pi}\,\sigma_{z|x}} \exp(-\beta_\phi^2)\,\beta_\phi^2\, d\beta_\phi$$

$$= -\frac{\log(2\pi\sigma_{z|x}^2)}{2\sqrt{\pi}} \int \exp(-\beta_\phi^2)\, d\beta_\phi - \frac{1}{\sqrt{\pi}} \int \beta_\phi^2 \exp(-\beta_\phi^2)\, d\beta_\phi$$

Gaussian integral

$$\begin{cases} \int \exp(-x^2)\,dx = \sqrt{\pi} \\ \int x^n \exp(-ax^2)\,dx = \dfrac{2(n-1)!}{2^{\frac{(n+2)}{2}}\, a^{\frac{n}{2}}} \sqrt{\frac{\pi}{a}} \quad (n: \text{even number}) \end{cases}$$

$$= -\frac{\log(2\pi\sigma_{z|x}^2)}{2\sqrt{\pi}} \sqrt{\pi} - \frac{1}{\sqrt{\pi}} \frac{2}{4} \cdot \sqrt{\pi} \quad \cdots \boxed{n=2,\ a=1}$$

$$= -\frac{\log(2\pi\sigma_{z|x}^2)}{2} - \frac{1}{2}$$

$$= -\frac{1}{2}\left(\log(2\pi\sigma_{z|x}^2) + 1\right)$$

② $$= -\int q_\phi(z|x) \log p_\theta(z)\, dz$$

$$= -\int q_\phi(z|x) \log\left\{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)\right\} dz$$

$$= -\int q_\phi(z|x) \log \frac{1}{\sqrt{2\pi\sigma^2}}\, dz + \int q_\phi(z|x) \left(\frac{(z-\mu)^2}{2\sigma^2}\right) dz$$

$$= \frac{1}{2}\log(2\pi\sigma^2) \int q_\phi(z|x)\, dz + \frac{1}{2\sigma^2}\left\{\int q_\phi(z|x) z^2\, dz - \int q_\phi(z|x)\, 2\mu z\, dz\right.$$

$$\left. + \int \mu^2 q_\phi(z|x)\, dz\right\} \quad \cdots \boxed{\int q_\phi(z|x)\, dz = 1}$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(\mathbb{E}(z^2|x) - 2\mu\,\mathbb{E}(z|x) + \mu^2\right)$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + \mu_{z|x}^2 - 2\mu\mu_{z|x} + \mu^2\right)\cdots$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right)$$

$\sigma^2 = \mathbb{E}[x^2] - \mu^2$

$\mathbb{E}[x^2] = \sigma^2 + \mu^2$

$\therefore KL(q_\phi(z|x), p_\theta(z)) = ① + ②$

$$= \left(-\frac{1}{2}\log\left(2\pi\sigma_{z|x}^2\right) - \frac{1}{2}\right) + \left(\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right)\right)$$

$$= \frac{1}{2}\log\left(\frac{2\pi\sigma^2}{2\pi\sigma_{z|x}^2}\right) + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right) - \frac{1}{2}$$

$$= \frac{1}{2}\log\frac{\sigma^2}{\sigma_{z|x}^2} + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right) - \frac{1}{2}$$

$$= \frac{1}{2}\cdot 2\cdot\log\frac{\sigma}{\sigma_{z|x}} + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right) - \frac{1}{2}$$

$$= \log\frac{\sigma}{\sigma_{z|x}} + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right) - \frac{1}{2}$$

If $\mu = \mu_{z|x}$ & $\sigma = \sigma_{z|x}$, the KL divergence becomes zero.

(c) $P_\theta(x|z) = \mathcal{N}(\mu_{x|z}, \sigma^2_{x|z})$

$\nabla_\phi \mathbb{E}_{z\sim q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)}\right)\right]$ → intractable !

$\nabla_\phi \mathbb{E}_{z\sim q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)}\right)\right] \approx \sum_{j=1}^{J} \nabla_\phi \log\left(\frac{p_\theta(x,z^{(j)})}{q_\phi(z^{(j)}|x)}\right)$

with $z^{(j)} \sim q_\phi(z^{(j)}|x)$ → still intractable !

→ reparametrization trick

Represent random variable $z \sim \mathcal{N}(\mu_{z|x}, \sigma^2_{z|x})$ as a function of $\mu_{z|x}, \sigma^2_{z|x}$, and $\varepsilon \sim N(0,1)$:

$$z = \sigma_{z|x}\,\varepsilon + \mu_{z|x}$$

Sample of z using sample form $N(0,1)$:

$$z^{(j)} = \sigma_{z|x}\,\varepsilon^{(j)} + \mu_{z|x} \quad \forall j = 1, \ldots, J$$

Approximate $\nabla_\phi \mathbb{E}_{z\sim q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)}\right)\right]$ with sampling and reparametrization trick:

$$\nabla_\phi \mathbb{E}_{z\sim q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)}\right)\right]$$

$$= \nabla_\phi \mathbb{E}_z\left[\log\left(\frac{p_\theta(x|z)\,p_\theta(z)}{q_\phi(z|x)}\right)\right]$$

$$= \nabla_\phi \mathbb{E}_z\left[\log p_\theta(x|z) + \log\left(\frac{p_\theta(z)}{q_\phi(z|x)}\right)\right]$$

$$= \nabla_\phi \left(\mathbb{E}_z\left[\log p_\theta(x|z)\right] + \mathbb{E}_z\left[\log\left(\frac{p_\theta(z)}{q_\phi(z|x)}\right)\right]\right)$$

$$= \nabla_\phi \left(\mathbb{E}_z\left[\log p_\theta(x|z)\right] - \mathbb{E}_z\left[\log\left(\frac{q_\phi(z|x)}{p_\theta(z)}\right)\right]\right)$$

$$= \nabla_\phi \left(\mathbb{E}_z\left[\log p_\theta(x|z)\right] - KL\left(q_\phi(z|x), p_\theta(z)\right)\right)$$

$$\approx \underbrace{\frac{1}{J}\sum_{j=1}^{J} \nabla_\phi \log p_\theta(x|z^{(j)})}_{①} \underbrace{- \nabla_\phi KL(q_\phi(z|x), p_\theta(z))}_{②}$$

$$① = \frac{1}{J}\sum_{j=1}^{J} \nabla_\phi \log p_\theta(x|\sigma_{z|x}\varepsilon^{(j)} + \mu_{z|x}) \cdots$$ from reparametrization & sampling

$$② = -\nabla_\phi\left(\log\frac{\sigma}{\sigma_{z|x}} + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x}-\mu)^2\right) - \frac{1}{2}\right) \cdots$$ from (b) result

∴ approximation result = ①+②

$$= \frac{1}{J}\sum_{j=1}^{J} \nabla_{\phi} \log p_{\theta}(x \mid \sigma_{z|x}\epsilon^{(j)} + \mu_{z|x})$$

$$- \nabla_{\phi}\left(\log \frac{\sigma}{\sigma_{z|x}} + \frac{1}{2\sigma^2}\left(\sigma_{z|x}^2 + (\mu_{z|x} - \mu)^2\right) - \frac{1}{2}\right)$$

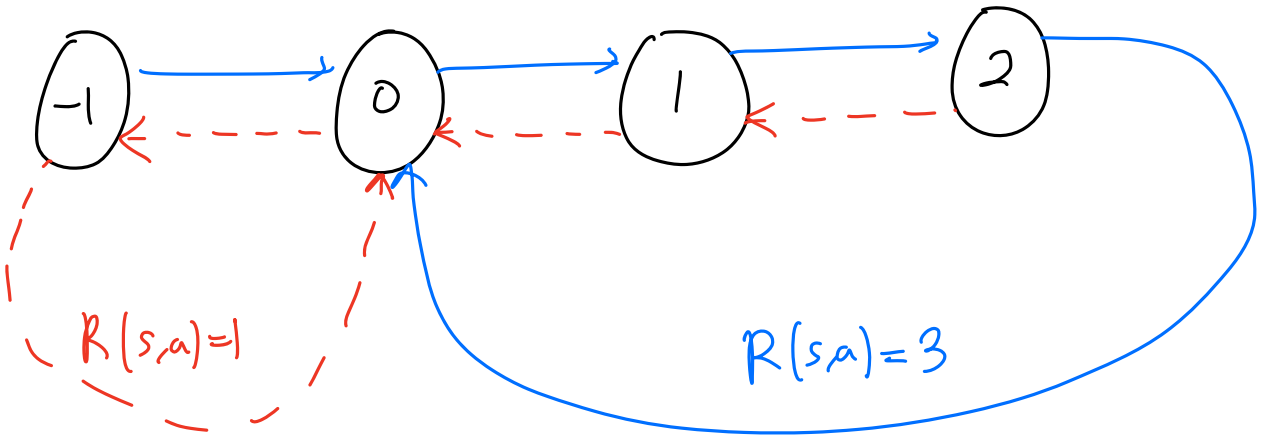
(d)

```python
def reparameterize(self, mu, logvar):
    ################################
    ## implement the reparameterize function
    std = torch.exp(logvar / 2)
    eps = torch.randn_like(std)
    ################################
    return eps.mul(std).add_(mu)
```

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL

Train Epoch: 26 [25600/60000 (43%)]     Loss: 101.455467
Train Epoch: 26 [38400/60000 (64%)]     Loss: 99.498177
Train Epoch: 26 [51200/60000 (85%)]     Loss: 108.227180
===== Epoch: 26 Average loss: 103.9785 =======
Train Epoch: 27 [0/60000 (0%)]  Loss: 100.385132
Train Epoch: 27 [12800/60000 (21%)]     Loss: 101.097992
Train Epoch: 27 [25600/60000 (43%)]     Loss: 103.895073
Train Epoch: 27 [38400/60000 (64%)]     Loss: 101.896179
Train Epoch: 27 [51200/60000 (85%)]     Loss: 103.818451
===== Epoch: 27 Average loss: 103.8807 =======
Train Epoch: 28 [0/60000 (0%)]  Loss: 104.533897
Train Epoch: 28 [12800/60000 (21%)]     Loss: 103.687950
Train Epoch: 28 [25600/60000 (43%)]     Loss: 103.391647
Train Epoch: 28 [38400/60000 (64%)]     Loss: 104.651192
Train Epoch: 28 [51200/60000 (85%)]     Loss: 105.794754
===== Epoch: 28 Average loss: 103.7341 =======
Train Epoch: 29 [0/60000 (0%)]  Loss: 104.814400
Train Epoch: 29 [12800/60000 (21%)]     Loss: 104.002884
Train Epoch: 29 [25600/60000 (43%)]     Loss: 104.355698
Train Epoch: 29 [38400/60000 (64%)]     Loss: 103.353668
Train Epoch: 29 [51200/60000 (85%)]     Loss: 101.666145
===== Epoch: 29 Average loss: 103.6788 =======
(assn) simjaeyoon@simjaeyoonui-MacBookPro AIGS515_ASSN5 %
```

2



$(s,a) \in \{(-1,-), (2,+)\}$ → non-zero

$r(-1,-)=1$

$r(2,+)=3$

(a) $\gamma = 0.10$

optimal policy $\pi^* = (\bigcirc_{s:-1}, \bigcirc_{0}, \bigcirc_{1}, \bigcirc_{2})$

optimal value function $V^*(s,\gamma) = (\bigcirc_{s:-1}, \bigcirc_{0}, \bigcirc_{1}, \bigcirc_{2})$

It is possible to generator 16 ($=2^4$) combination of policies. We can use some iteration method to compute value function and find optimal policy. To compute the value function, we can use recursive function form. Since discount factor γ is near to 0, it would converge much faster. So it need shorter iteration to find optimal solution. We need large reward as soon as possible.

(-1) $(-, -, +/-, +/-)$

$$V_\pi(-1) = 1 + 0 + \lambda^2 \cdot 1 + 0 + \lambda^4 \cdot 1 + \cdots = \frac{1}{1-\lambda^2} = 1.0\dot{1}$$

(0) $(-, -, +/-, +/-)$

$$V_\pi(0) = 0 + \lambda \cdot 1 + 0 + \lambda^3 \cdot 1 + 0 + \lambda^5 \cdot 1 + \cdots = \frac{\lambda}{1-\lambda^2} = 0.\dot{1}0$$

(1) $(-, -, +, +)$

$$V_\pi(1) = 0 + \gamma^1 \cdot 3 + 0 + \gamma^3 \cdot 1 + 0 + \gamma^5 \cdot 1 + 0 + \gamma^7 \cdot 1 = 3\gamma + \frac{\gamma^3}{1-\gamma^2} = 0.3\dot{0}\dot{1}$$

(2) $(-, -, +/-, +)$

$$V_\pi(2) = 3 + 0 + \gamma^2 \cdot 1 + 0 + \gamma^4 \cdot 1 + \cdots = 3 + \frac{\gamma^2}{1-\gamma^2} = 3.0\dot{1}$$

$\pi^*_{short} = (-, -, +, +)$, $V^*(s, \gamma) = (1.0\dot{1}, 0.1\dot{0}, 0.30\dot{1}, 3.0\dot{1})$

(b) $\gamma = 0.99$.

Method is similar to (a), but the discount factor is close to 1, so it is okay to get reward slower.

(-1) $(-, +, +, +)$

$$V_\pi(-1) = 1 + 0 + 0 + \gamma^3 \cdot 3 + 0 + 0 + \gamma^6 \cdot 3 + \cdots = 1 + \frac{3 \cdot \gamma^3}{1 - \gamma^3} \doteq 99.00\eta$$

(0) $(+/-, +, +, +)$

$$V_\pi(0) = 0 + 0 + \gamma^2 \cdot 3 + 0 + 0 + \gamma^5 \cdot 3 + \cdots = \frac{3\gamma^2}{1 - \gamma^3} \doteq 98.99\eta$$

(1) $(+/-, +, +, +)$

$$V_\pi(1) = 0 + \gamma \cdot 3 + 0 + 0 + \gamma^4 \cdot 3 + 0 + 0 + \cdots = \frac{3\gamma}{1 - \gamma^3} \doteq 99.99\eta$$

(2) $(+/-, +, +, +)$

$$V_\pi(2) = 3 + 0 + 0 + \gamma^3 \cdot 3 + 0 + 0 + \gamma^6 \cdot 3 = 3 + \frac{3\gamma^3}{1 - \gamma^3} \doteq 101.00\eta$$

$\pi^*_{long} = (-, +, +, +)$, $V^*(S, \gamma) = (\underbrace{99.00\eta, 98.99\eta, 99.99\eta, 101.00\eta}_{\text{approximation.}})$

(c) $\gamma = 0.99$ with problem 2a.

$V^{\pi^*_{short}}(s, \gamma = 0.99) = (50.25, 49.75, 51.73, 52.25)$

$V^{*}(s, \gamma = 0.10) = (1.\dot{0}\dot{1}, 0.\dot{1}\dot{0}, 0.3\dot{0}\dot{1}, 3.\dot{0}\dot{1})$

$\gamma = 0.10$ → accumulate ⇓ step → converge fast
→ low optimal value function.

$\gamma = 0.99$ → accumulate ⇑ step → converge slow
→ high optimal value function.

γ controls the accumulation of value function from longer timestep. If γ is close to 1, it can consider far future, however if γ is close to 0, it can consider near future.
When discount factor becomes larger (0.10 → 0.99), then optimal value function results become larger comparing with π^*_{short}. Comparing the result between π^*_{long} and π^*_{short} with $\gamma = 0.99$, π^*_{long} is optimal policy when γ is large.