מודלים גנרטיביים עמוקים

תרגיל #2

(1) Show that:

$$\log(p(x)) = ELBO(p,q,x) + KL(q(z)\|p(z|x))$$

$$\mathbb{E}_{z\sim q}\left[\log(p(x|z))\right] - KL(q(z)\|p(z))$$

הוכחה:

$$KL(q(z)\|p(z)) \triangleq \mathbb{E}_{z\sim q}\left[\log\left(\frac{q(z)}{p(z)}\right)\right]$$

$$KL(q(z)\|p(z|x)) \triangleq \mathbb{E}_{z\sim q}\left[\log\left(\frac{q(z)}{p(z|x)}\right)\right]$$

נקבל:

$$ELBO(p,q,x) + KL(q(z)\|p(z|x)) =$$

$$= \mathbb{E}_{z\sim q}\left[\log(p(x|z))\right] - \mathbb{E}_{z\sim q}\left[\log\left(\frac{q(z)}{p(z)}\right)\right] + \mathbb{E}_{z\sim q}\left[\log\left(\frac{q(z)}{p(z|x)}\right)\right]$$

$$= \mathbb{E}_{z\sim q}\left[\log(p(x|z)) - \cancel{\log(q(z))} + \log(p(z)) + \cancel{\log(q(z))} - \log(p(z|x))\right]$$

$$= \mathbb{E}_{z\sim q}\left[\log(p(x|z)) + \log(p(z)) - \log(p(z|x))\right]$$

עפ"י נוסחת בייס:

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

נקבל:

$$= \mathbb{E}_{z\sim q}\left[\log(p(x|z)) + \log(p(z)) - \log\left(\frac{p(x|z)\,p(z)}{p(x)}\right)\right]$$

$$= \mathbb{E}_{z\sim q}\left[\log\left(\frac{\cancel{p(x|z)}\cdot\cancel{p(z)}\,p(x)}{\cancel{p(x|z)}\cdot\cancel{p(z)}}\right)\right] =$$

$$= \mathbb{E}_{z\sim q}\left[\log(p(x))\right] = \log(p(x))$$

$\log(p(x))$ אינו תלוי ב-z

② Show that:

$$\mathbb{E}_p\left[\log\left(q(x)\right)\right] \leq \mathbb{E}_p\left[\log\left(Q(x)\right)\right]$$

$$\mathbb{E}_p\left[\log\left(q(x)\right)\right] \triangleq \int \log\left(q(x)\right) p(x)\, dx$$

עפ"י אי-שוויון ינסן :

$$\int \log\left(q(x)\right) p(x)\, dx \underset{①}{\leq} \log\left(\int_x q(x) p(x)\, dx\right)$$

באותו אופן :

$$\mathbb{E}_p\left[\log\left(Q(x)\right)\right] = \int_x \log\left(Q(x)\right) p(x)\, dx \underset{②}{\leq} \log\left(\int_x Q(x) p(x)\, dx\right)$$

נדרש להראות כי מתקיים :

$$\mathbb{E}_p\left[\log\left(q(x)\right)\right] \leq \mathbb{E}_p\left[\log\left(Q(x)\right)\right]$$

$$0 \leq \mathbb{E}_p\left[\log\left(Q(x)\right)\right] - \mathbb{E}_p\left[\log\left(q(x)\right)\right]$$

$$\log \int_x Q(x) p(x)\, dx - \log \int_x q(x) p(x)\, dx$$

נציב את $Q(x)$:

$$Q(x) = \int_0^1 q(x+u)\, du$$

$$\log \int_x \int_0^1 q(x+u) p(x)\, du\, dx - \log \int_x q(x) p(x)\, dx =$$

$$= \log \int_x \left[\int_0^1 q(x+u)\, du - q(x)\right] p(x)\, dx$$

עפ"י ההגדרה, u הינה תוספת יוניפורמית <u>חיובית</u> בתחום $[0,1]$

ולכן הפרמטר באינטגרל מקבל ≥ 0 :

$$= \log \int_x \underbrace{\left[\int_0^1 q(x+u)\, du - q(x)\right]}_{\geq 0} p(x)\, dx$$

$\Leftarrow$ עבור $q(x)>0$ מתקבל ערך גדול מ-0

$\Leftarrow$ עבור $q(x)=0$ מתקבל ערך השווה ל-0.

ולכן מתקיים:

$$\log \int_x \left[\int_0^1 g(x+u)\,du - g(x) \right] p(x)\,dx \geq 0$$

③

ⓐ עפ"י הגדרת ה-Score function Estimator:

$$\nabla_\mu \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)}\left[f(z)\nabla_\mu \log(q_\phi(z))\right]$$

נתון: $z = g(\varepsilon, \phi) = \mu_\phi + \varepsilon \sigma_\phi^2 \sim N(\mu_\phi, \sigma_\phi^2)$

$$\mathbb{E}_{q_\phi(z)}\left[f(z)\nabla_\mu \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu_\phi)^2}{2\sigma_\phi^2}}\right)\right] =$$

$$= \mathbb{E}_{q_\phi(z)}\left[f(z)\nabla_\mu \left[\frac{-(z-\mu_\phi)^2}{2\sigma_\phi^2} - \underbrace{\log\left(\sqrt{2\pi}\sigma_\phi^2\right)}_{\text{גורם שאינו תלוי ב-}\mu}\right]\right]$$

ולכן מתקבל:

$$= \mathbb{E}_{q_\phi(z)}\left[f(z)\cdot\left[-\frac{\cancel{2}(z-\mu_\phi)}{\cancel{2}\sigma_\phi^2}\cdot\cancel{(-1)}\right]\right]$$

$$= \mathbb{E}_{q_\phi(z)}\left[f(z)\frac{(z-\mu_\phi)}{\sigma_\phi^2}\right]$$

עבור דגימה בודדת נקבל:

$$f(z)\frac{(z-\mu_\phi)}{\sigma_\phi^2}$$

ⓑ עפ"י הגדרת reparametrization trick:

$$\nabla_\mu \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)}\left[\nabla_\mu f(g(\varepsilon,\phi))\right]$$

$$= \mathbb{E}_{q_\phi(z)}\left[\nabla_\mu f(\underbrace{\mu_\phi + \varepsilon\sigma_\phi^2}_{=z})\right] = \mathbb{E}_{q_\phi(z)}\left[f'(z)\right]$$

$$Var(x) \triangleq E(x^2) - E^2(x)$$ ⓒ

$$f(x) = ax$$

נציב את $f(x)$ ונקבל:

$$Var\left(f(z)\frac{(z-\mu)}{\sigma^2}\right) = E\left[\left(f(z)\frac{(z-\mu)}{\sigma^2}\right)^2\right] - E^2\left[f(z)\frac{(z-\mu)}{\sigma^2}\right]$$

$$= E\left[\left(\frac{az(z-\mu)}{\sigma^2}\right)^2\right] - E^2\left[\frac{az(z-\mu)}{\sigma^2}\right]$$

$$= \underbrace{E\left[\frac{a^2z^2(z-\mu)^2}{\sigma^4}\right]}_{1} - \underbrace{E^2\left[\frac{az^2 - az\mu}{\sigma^2}\right]}_{2}$$

נפתח את 1:

$$1 = E\left[\frac{a^2z^4}{\sigma^4} - \frac{2a^2z^3\mu}{\sigma^4} + \frac{a^2z^2\mu^2}{\sigma^4}\right]$$

$$= \frac{a^2}{\sigma^4}E[z^4] - \frac{2a^2\mu}{\sigma^4}E[z^3] + \frac{a^2\mu^2}{\sigma^4}E[z^2]$$

$$= \frac{a^2}{\sigma^4}(\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4) - \frac{2a^2\mu}{\sigma^4}(\mu^3 + 3\mu\sigma^2) + \frac{a^2\mu^2}{\sigma^4}(\mu^2 + \sigma^2)$$

$$= \frac{a^2\mu^4}{\sigma^4} + \cancel{\frac{6a^2\mu^2\sigma^2}{\sigma^4}} + \frac{3a^2\sigma^4}{\sigma^4} - \frac{2a^2\mu^4}{\sigma^4} - \cancel{\frac{6a^2\mu^2\sigma^2}{\sigma^4}}$$

$$+ \frac{a^2\mu^4}{\sigma^4} + \frac{a^2\mu^2\sigma^2}{\sigma^4} =$$

$$= \cancel{\frac{2a^2\mu^4}{\sigma^4}} + 3a^2 - \cancel{\frac{2a^2\mu^4}{\sigma^4}} + \frac{a^2\mu^2}{\sigma^2} = 3a^2 + \frac{a^2\mu^2}{\sigma^2}$$

נפתח את 2:

$$2 = E\left[\frac{az^2}{\sigma^2}\right] - E\left[\frac{az\mu}{\sigma^2}\right] = \frac{a}{\sigma^2}E[z^2] - \frac{a\mu}{\sigma^2}E[z]$$

$$= \frac{a}{\sigma^2}(\mu^2 + \sigma^2) - \frac{a\mu^2}{\sigma^2} = \cancel{\frac{a\mu^2}{\sigma^2}} + \frac{a\sigma^2}{\sigma^2} - \cancel{\frac{a\mu^2}{\sigma^2}} = a$$

נציב את הפיתוחים של 1 ו-2 ונקבל:

$$Var\left(f(z)\frac{(z-\mu)}{\sigma^2}\right) = 3a^2 + \frac{a^2\mu^2}{\sigma^2} - a^2 = a^2\left(2 + \frac{\mu^2}{\sigma^2}\right)$$

עבור תוצאות ה- reparametrized trick קטנים

בסעיף 4 (5) ט המשערך הינו: $f'(z)$

ולכן בהצבה $f(x)=a*x$ נקבל:

$$f'(z)=\frac{df}{dz}(z)=a$$

הווריאנס המתקבל הינו 0:

$$Var\left(f'(z)\right)=E\left[\left(f'(z)\right)^2\right]-E^2\left[f'(z)\right]$$

$$=E\left[a^2\right]-E^2\left[a\right]=a^2-a^2=0$$