DGM HW1 Written part

ALT_PDG

2025 年 10 月 30 日

Problem 2a

Given data x (binary images for MNIST) and latent variable z, ELBO can be expressed as:

$$\log p(x) \geq E_{q(z|x)}\left[\frac{\log p(x,z)}{q(z|x)}\right] = \text{ELBO}$$

Give a detailed mathematical proof of the ELBO starting from the marginal log-likelihood log p(x). Show all steps clearly

Proof: The first step is to represent $logp(x)$ as an integral since we must deduce something involve mathematical expectation $E[\cdot]$, which is an integral of density. Recall that p(x) is the marginal distribution of joint distribution p(x,z), so

$$logp(x) = log \int_{R^d} p(x,z)\, dz$$

In order to have $q(z|x)$, we simply multiply and then devide $q(z|x)$:

$$logp(x) = log \int_{R^d} p(x,z)\, dz = log \int_{R^d} \frac{p(x,z)}{q(z|x)} q(z|x)\, dz$$

we have reached an critical point, first let $a = \int_{R^d} \frac{p(x,z)}{q(z|x)} q(z|x)\, dz$, then we find

$$logx \leq \frac{x-a}{a} + loga \qquad \forall x > 0$$

To see this, it sufficies to notice that the function $f(x) = logx - \frac{x-a}{a} + loga$ satisfies $f'(x) = \frac{1}{x} - \frac{1}{a}$, hence is increase on $(0,a)$, decrease on $(a, +\infty)$, hence $f(x) \leq f(a)$

Now let $x = \frac{p(x,z)}{q(z|x)}$, then multiy $q(z|x)$ on both sides yields

$$log(\frac{p(x,z)}{q(z|x)})q(z|x) \leq \frac{1}{a} q(z|x)\frac{p(x,z)}{q(z|x)} - q(z|x) + q(z|x)loga$$

The above formula seems little ugly but indeed we can easily modify it, first,take integral with respect to $z \in R^d$ on both sides, we have

$$\int_{R^d} log(\frac{p(x,z)}{q(z|x)})q(z|x)\, dz \leq \int_{R^d} \frac{1}{a} q(z|x)\frac{p(x,z)}{q(z|x)}\, dz - \int_{R^d} q(z|x)\, dz + \int_{R^d} q(z|x)loga\, dz$$

$$= \frac{1}{a}\int_{R^d} q(z|x)\frac{p(x,z)}{q(z|x)}\, dz - 1 + loga \int_{R^d} q(z|x)\, dz$$

where we used the fact that： $\int_{R^d} q(z|x)\, dz = \int_{R^d} \frac{q(z,x)}{q(x)}\, dz = \frac{q(x)}{q(x)} = 1$.($q(x)$ is the marginal distribution of joint distribution $q(z,x)$)

Recall that $a = \int_{R^d} q(z|x) \frac{p(x,z)}{q(z|x)}\, dz$, we have

$$\int_{R^d} log(\frac{p(x,z)}{q(z|x)}) q(z|x)\, dz \leq 1 - 1 + loga = log \int_{R^d} q(z|x) \frac{p(x,z)}{q(z|x)}\, dz$$

The left hand side $= E_{q(z|x)}[log\frac{p(x,z)}{q(z|x)}]$, the right hand side $= logp(x)$, thus we get the desired result.

Problem 3a

Assume that:

- The prior $p(z)$ is a standard Gaussian distribution: $p(z) = \mathcal{N}(z; 0, I)$.
- The approximate posterior $q(z|x)$ is a Gaussian distribution with mean $\mu(x)$ and diagonal covariance matrix $\text{diag}(\sigma^2(x))$: $q(z|x) = \mathcal{N}(z; \mu(x), \text{diag}(\sigma^2(x)))$.

The ELBO can then be expressed as:

$$\text{ELBO} = E_{q(z|x)}[\log p(x|z)] - D_{\text{KL}}(q(z|x) \parallel p(z))$$
$$= \frac{1}{2} \sum_d \left(1 + \log\left((\sigma_d)^2\right) - (\mu_d)^2 - (\sigma_d)^2\right) + \sum_{i,j} \log p(x_i|z_{i,j})$$

where d is the dimension of the latent space, and i, j are the indices of the data and the latent samples, respectively.

Give a detailed mathematical proof of the above ELBO decomposition. Show all steps clearly

Proof: By the previous problem, we know

$$ELBO = E_{q(z|x)}\left[\frac{\log p(x,z)}{q(z|x)}\right] = E_{q(z|x)}[log\frac{p(x|z)p(z)}{q(z|x)}] = E_{q(z|x)}[logp(x|z)] - E_{q(z|x)}[log\frac{q(z|x)}{p(z)}]$$
$$= E_{q(z|x)}[logp(x|z)] - D_{KL}(q(z|x)||p(z))$$

Now use our condition of $q(z|x), p(z)$,

$$D_{KL}(q(z|x)||p(z)) = E_{q(z|x)}[log\frac{q(z|x)}{p(z)}] = \int_{R^d} [logq(z|x) - logp(z)] q(z|x)\, dz$$

Recall the knowledge of Multivariate normal distribution, we know

$$logq(z|x) = log\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{(z-\mu(x))^T\Sigma^{-1}(z-\mu(x))}{2}} \qquad \Sigma = diag(\sigma^2(x))$$

where $|A|$ means the determinate of A if A is a matrix.

$$= -\frac{1}{2}\sum_{i=1}^{d} \frac{(z_i - \mu(x)_i)^2}{\sigma_i^2} - \frac{d}{2}log2\pi - log\sqrt{\prod_{i=1}^{d}\sigma_i^2}$$

Since $p(z)$ is standard multivariate normal distribution, similar as $q(z|x)$, with $\sigma_i = 1, \mu_i(x) = 0$, we have

$$logp(z) = -\frac{1}{2}\sum_{i=1}^{d} z_i^2 - \frac{d}{2}log2\pi - log1 = -\frac{1}{2}\sum_{i=1}^{d} z_i^2 - \frac{d}{2}log2\pi$$

Hence we have

$$D_{KL}(q(z|x)||p(z)) = \int_{R^d} \left[\frac{1}{2}\sum_{i=1}^{d} z_i^2 - \frac{1}{2}\sum_{i=1}^{d}\frac{(z_i - \mu(x)_i)^2}{\sigma_i^2} - \frac{1}{2}\sum_{i=1}^{d} log\sigma_i^2\right] q(z|x)\, dz$$

$$= \int_{R^d} \left[\frac{1}{2} \sum_{i=1}^{d} z_i^2 - \frac{1}{2} \sum_{i=1}^{d} \frac{(z_i - \mu(x)_i)^2}{\sigma_i^2} \right] q(z|x)\, dz - \frac{1}{2} \sum_{i=1}^{d} log\sigma_i^2$$

Next we calculate the first term in the above integral:

$$\int_{R^d} z_i^2 q(z|x)\, dz = \int_{R^d} z_i^2 \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(z-\mu(x))^T \Sigma^{-1} (z-\mu(x))}{2}}\, dz$$

$$\overset{y_j = z_j - \mu(x)_j}{\Longrightarrow\!=} \int_{R^d} (y_i + \mu(x)_i)^2 \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{j=1}^{d} \sigma_j} e^{-\frac{\Sigma_{j=1}^{d} \frac{y_j^2}{2\sigma_j^2}}{2}}\, dy$$

$$= \int_{-\infty}^{+\infty} (y_i + \mu(x)_i)^2 \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}}\, dy_i \prod_{j \geq 1 \&\& j \neq i}^{d} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{y_j^2}{2\sigma_j^2}}\, dy_j$$

$$= \int_{-\infty}^{+\infty} (y_i + \mu(x)_i)^2 \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}}\, dy_i = \sigma_i^2 + 2\mu_i(x) \cdot 0 + \mu(x)_i^2 \cdot 1 = \sigma_i^2 + \mu(x)_i^2$$

Do this for all $1 \leq i \leq d$, we have

$$\int_{R^d} \frac{1}{2} \sum_{i=1}^{d} z_i^2 q(z|x)\, dz = \frac{1}{2} \sum_{i=1}^{d} [\sigma_i^2 + \mu(x)_i^2]$$

For the second term:

$$\int_{R^d} \frac{(z_i - \mu(x)_i)^2}{\sigma_i^2} q(z|x)\, dz = \int_{R^d} \frac{(z_i - \mu(x)_i)^2}{\sigma_i^2} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(z-\mu(x))^T \Sigma^{-1} (z-\mu(x))}{2}}\, dz$$

$$\overset{y_j = z_j - \mu(x)_j}{\Longrightarrow\!=} \int_{R^d} \frac{y_i^2}{\sigma_i^2} \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{j=1}^{d} \sigma_j} e^{-\frac{\Sigma_{j=1}^{d} \frac{y_j^2}{2\sigma_j^2}}{2}}\, dy$$

$$= \int_{-\infty}^{+\infty} \frac{y_i^2}{\sigma_i^2} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}}\, dy_i \prod_{j \geq 1 \&\& j \neq i}^{d} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{y_j^2}{2\sigma_j^2}}\, dy_j = \int_{-\infty}^{+\infty} \frac{y_i^2}{\sigma_i^2} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}}\, dy_i = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

Combine those calculation yields

$$D_{KL}(q(z|x)||p(z)) = \frac{1}{2} \sum_{i=1}^{d} [\sigma_i^2 + \mu(x)_i^2] - \frac{1}{2} \sum_{i=1}^{d} 1 - \frac{1}{2} \sum_{i=1}^{d} log\sigma_i^2 = -\frac{1}{2} \sum_{d} [1 + log\sigma^2 - \mu(x)^2 - \sigma^2]$$

For $E_{q(z|x)}[logp(x|z)]$, we directly use the Monte-Carlo sampling so that $E_{q(z|x)}[logp(x|z)]$ can be approximated as

$$\sum_{i,j} logp(x_i|z_{ij})$$

Hence

$$ELBO = \sum_{i,j} logp(x_i|z_{ij}) + \frac{1}{2} \sum_{d} [1 + log\sigma^2 - \mu(x)^2 - \sigma^2]$$