Homework 1

Haopeng Zhang - 2023010770

2025 6 - 7

Lecture on AR and VAE

Code and results available at:

deepul/hw1.ipynb

deepul/hw2.ipynb

Question 1:

Explain results in deepul/hw2/Q1.

Answer:

Latents are used only in covariance Gaussian because in diagonal Gaussian, the decoder can reconstruct features independently without using latents.

Question 2:

Prove that $\mathrm{KL}(p||q) \geq 0$,and the equality holds iff $q = p$.

Answer:

$$
\begin{aligned}
\mathrm{KL}(p||q) &= \sum_x p(x) \log \frac{p(x)}{q(x)} \\
&= \sum_x p(x) \left[-\log \frac{q(x)}{p(x)} \right] \\
&\geq -\log \left[\sum_x p(x) \frac{q(x)}{p(x)} \right] \quad \text{(Jensen inquality)} \\
&= -\log \sum_x q(x) \\
&= -\log(1) \\
&= 0
\end{aligned}
$$

Equality holds iff $\forall x$, $\frac{p(x)}{q(x)} = \text{const}$. Since both p and q are probability density functions, $\sum_x p(x) = \sum_x q(x) = 1$, $\frac{p(x)}{q(x)} = \text{const} \iff p(x) = q(x)$.

Question 3:

Explain why minimizing KL divergence is equivalent to Maximum Likelihood Estimation (MLE).

Answer:

Let $p_\theta(x)$ be the parameterized probability density function, and $p(x)$ be the true distribution of the data. MLE seeks to find $\theta = \arg\max_\theta \prod_x p_\theta(x)$.

$$
\begin{aligned}
\theta &= \arg\max_\theta \prod_x p_\theta(x) \\
&= \arg\max_\theta \log \prod_x p_\theta(x) \\
&= \arg\max_\theta \sum_x \log p_\theta(x) \\
&= \arg\max_\theta \frac{1}{\#\{x\}} \sum_x \log p_\theta(x) \\
&= \arg\max_\theta \mathbb{E}\left[\log p_\theta(x)\right] \\
&= \arg\min_\theta -\mathbb{E}\left[\log p_\theta(x)\right] \\
&= \arg\min_\theta \mathbb{E}\left[\log p(x)\right] - \mathbb{E}\left[\log p_\theta(x)\right] \\
&= \arg\min_\theta \mathbb{E}\left[\log \frac{p(x)}{p_\theta(x)}\right] \\
&= \arg\min_\theta \mathrm{KL}(p||p_\theta)
\end{aligned}
$$