Deep Learning Additional Homework - Variational Autoencoder

110550088 李杰穎

April 21, 2025

1 Closed-form function of KL term

Here, we need to derive the closed-form function of KL term, $\mathrm{KL}\left(q(\mathbf{Z}|\mathbf{X};\theta')\,||\,p(\mathbf{Z})\right)$. Suppose $Q \sim \mathcal{N}(\mu, \sigma^2)$ and the prior distribution $p(\mathbf{Z})$ is $\mathcal{N}(0,1)$, the KL term can be expanded as:

$$
\begin{aligned}
\mathrm{KL}\left(q(\mathbf{Z}|\mathbf{X};\theta')\,||\,p(\mathbf{Z})\right) &= \int_z q(z|\mathbf{X};\theta') \ln \frac{q(z|\mathbf{X};\theta')}{p(z)} && (1)\\
&= \int_{-\infty}^{\infty} \mathcal{N}(z;\mu,\sigma^2) \ln \frac{\mathcal{N}(z;\mu,\sigma^2)}{\mathcal{N}(z;0,1)}\, dz && (2)\\
&= \mathbb{E}\left[\ln \frac{\mathcal{N}(\mathbf{Z};\mu,\sigma^2)}{\mathcal{N}(\mathbf{Z};0,1)}\right] && (3)\\
&= \mathbb{E}\left[\ln \frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-1}{2}\left(\frac{\mathbf{Z}-\mu}{\sigma}\right)^2)}{\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}\mathbf{Z}^2\right)}\right] && (4)\\
&= \mathbb{E}\left[\frac{1}{2}\ln\frac{1}{\sigma^2} - \frac{1}{2}\left(\frac{\mathbf{Z}-\mu}{\sigma}\right)^2 + \frac{1}{2}\mathbf{Z}^2\right] && (5)\\
&= \frac{1}{2}\mathbb{E}\left[\ln\frac{1}{\sigma^2} - \left(\frac{\mathbf{Z}-\mu}{\sigma}\right)^2 + \mathbf{Z}^2\right] && (6)\\
&= \frac{1}{2}\left(\ln\frac{1}{\sigma^2} - \frac{\mathbb{E}\left[(\mathbf{Z}-\mu)^2\right]}{\sigma^2} + \mathbb{E}\left[\mathbf{Z}^2\right]\right). && (7)
\end{aligned}
$$

Because $z \sim \mathcal{N}(\mu,\sigma^2)$, by definition, we have $\mathbb{E}\left[(\mathbf{Z}-\mu)^2\right] = \sigma^2$. Also, $\mathrm{Var}\left[Q\right] = \sigma^2 = \mathbb{E}\left[\mathbf{Z}^2\right] - (\mathbb{E}\left[\mathbf{Z}\right])^2$. Thus, we have $\mathbb{E}\left[\mathbf{Z}^2\right] = \mu^2 + \sigma^2$. Finally, the closed-form function can be written as:

$$
\begin{aligned}
&= \frac{1}{2}\left(\ln\frac{1}{\sigma^2} - \frac{\mathbb{E}\left[(\mathbf{Z}-\mu)^2\right]}{\sigma^2} + \mathbb{E}\left[\mathbf{Z}^2\right]\right) && (8)\\
&= \frac{1}{2}\left(-\ln\sigma^2 - 1 + \mu^2 + \sigma^2\right). && (9)
\end{aligned}
$$

2 Gaussian Mixture Model as Prior

The Gaussian mixture model (GMM) combines multiple Gaussian distributions by calculating the weighted sum of each Gaussian distributions. Specifically, a GMM can be formally written as:

$$
p(\mathbf{Z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{Z}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{10}
$$

where π_k are mixture weights (with $\sum_{k=1}^{K}\pi_k = 1$), and $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ are the mean and covariance of each component. Compared with standard normal distribution, GMM can capture multi-modal data distributions more effectively, increasing the VAE's representational capacity. However, the issue arises when calculating the closed-form function of the KL term. The KL term doesn't have closed-form solution when using GMM as prior distribution, due to the summation in logarithm. The lack of closed-form solution prevents us not able to efficiently optimize the encoder network through back-propagation. To deal with this, we can use Monte Carlo sampling to estimate the KL divergence. Specifically, we can evaluate KLD as follow:

$$\mathrm{KL}\left(q(\mathbf{Z}|\mathbf{X};\theta')\,||\,p(\mathbf{Z})\right) \approx \frac{1}{N}\sum_{i=1}^{N} \ln \frac{q(z_i|\mathbf{X};\theta')}{p(z_i)}, \tag{11}$$

where z_i samples from $q(\mathbf{Z}|\mathbf{X};\theta')$, N is the number of samples and $p(\mathbf{Z})$ is the prior GMM distribution.