Problem 1: Importance-weighted ELBO

In lecture, we briefly mentioned that instead of the ELBO

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

one can use a tighter bound on the likelihood in a VAE (or any latent variable model):

$$\log p_\theta(x) \geq \mathbb{E}_{z_1,\ldots,z_K \sim q_\phi(z|x)} \left[\log \frac{1}{K} \sum_{k=1}^{K} \frac{p_\theta(x \mid z_k) p(z_k)}{q_\phi(z_k \mid x)} \right].$$

1. Show that as $K \to \infty$, the new bound converges to $\log p_\theta(x)$, so the IW bound can be made arbitrarily tight (hint: what does the $\frac{1}{K} \sum_{k=1}^{K}$ turn into as $K \to \infty$?).

2. (Harder:) Show that the new bound is tighter than the ELBO, i.e., that for all $K \geq 1$,

$$\mathbb{E}_{z_1,\ldots,z_K \sim q_\phi(z|x)} \left[\log \frac{1}{K} \sum_{k=1}^{K} \frac{p_\theta(x \mid z_k) p(z_k)}{q_\phi(z_k \mid x)} \right] \geq \mathbb{E}_{z \sim q_\phi(z|x)} \left[\log \frac{p_\theta(x \mid z) p(z)}{q_\phi(z \mid x)} \right]$$

   (hint: this uses the fact that for any $c_1, \ldots, c_k$, we have $\log \left(\frac{1}{K} \sum_{k=1}^{K} c_k \right) \geq \frac{1}{K} \sum_{k=1}^{K} \log c_k$), which is a form of Gibbs' inequality).

Problem 2: Denoising autoencoder

This problem considers VAEs in which the data and latent spaces are of the same dimension d, which is connected to diffusion models (lectures 8-10).

1. (a) Suppose that the encoder $q(z \mid x)$ in a VAE is fixed to be Gaussian: $q(z \mid x) = \mathcal{N}(z; x, \sigma^2 I_d)$, where $\sigma > 0$ is a constant. What is the VAE objective for the parameters of the decoder $p_\theta(x \mid z)$ in this case? Assume the decoder is Gaussian with fixed variance: $p_\theta(x \mid z) = \mathcal{N}(x; \mu_\theta(z), \tau^2 I_d)$.

   (b) Suppose we optimise the VAE over $x \sim \pi_{\text{data}}$. Show that the optimal $\mu_\theta(z)$ equals the conditional expectation $\mathbb{E}[X \mid Z = z]$, where (X, Z) is distributed according to $\pi_{\text{data}}(x) q(z \mid x)$ (hint: use the fact that optimising $\mathbb{E}_{Y \sim p(Y)}[\|Y - c\|^2]$ over c gives $c^* = \mathbb{E}[Y]$).

2. Now consider instead a decoder $p(x \mid z)$ that is fixed to be Gaussian ($p(x \mid z) = \mathcal{N}(x; z, \tau^2 I_d)$). Show that the true posterior $p(z \mid x) \propto p(z) p_\theta(x \mid z)$ is also Gaussian and give its parameters (hint: the product of two Gaussian densities is proportional to a Gaussian density). What is the optimal encoder $q_\phi(z \mid x)$ in this case?