STAT 542: Homework 7

Due: May 3 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

The purpose of this exercise is to understand why it helps to introduce ELBO in VAE. See p6 in `s17_VAE.pdf` for the definition of ELBO. Consider the following example: $p_\theta(x, z) = p(z)p_\theta(x|z)$, where $p(z = \cdot)$ is $\mathcal{N}(0, 1)$, and $p_\theta(x = \cdot|z)$ is $\mathcal{N}(z + \theta, 1)$. In other words, in the notation of p5 in `s17_VAE.pdf`, we have $\mu_\theta(z) = z + \theta$ and $\Sigma_\theta(z) = 1$. We chose this simplified example to streamline the calculations, but we can see that the intuition here (that taking the log reduces the variation of a function) extends to more general settings.

- Can you find an explicit expression for $p_\theta(z|x)$?
- Consider the variational approach, whereby we consider a variational family $q_\lambda = \mathcal{N}(\lambda_1, \lambda_2)$ of distributions, with $\lambda = (\lambda_1, \lambda_2) \in (0, \infty)^2$. In order to calculate $p_\theta(x, z)$, a naive approach is to consider

$$p_\theta(x, z) = \mathbb{E}_{q_\lambda}[p_\theta(x, z)/q_\lambda(z)] \tag{1}$$
$$\approx \frac{1}{n}\sum_{i=1}^{k} p_\theta(x, z^{(i)})/q_\lambda(z^{(i)}) \tag{2}$$

  where $z^{(i)}$, $i = 1, \ldots, k$ are samples drawn from q_λ. Here (2) is an unbiased estimator of (1) due to linearity of expectation. To make it a good estimator we also need to ensure its variance is small. Question: show that there exists a λ^* such that the variance of (2) is infinite whenever $\lambda_2 < \lambda^*$. What is the largest choice of λ^*?
- The ELBO method addresses the above issue: consider the method on p6 and p7 in `s17_VAE.pdf`. What is the range of λ for which the variance of $ELBO(x, \theta, \lambda)$ is finite?

Problem 2.

Suppose that ϕ is a smooth nonnegative function supported on $[-1, 1]$ satisfying $\int \phi = 1$, and $\phi(x) > 0$ for all $x \in [-1, 1]$. Let $p_1(x) = 0.5\phi(x + 2) + 0.5\phi(x - 2)$ and $p_2(x) = 0.1\phi(x + 2) + 0.9\phi(x - 2)$ be two density functions. An illustration can be found on p12 of `s18_diffusion.pdf`.

- Suppose that we use a "cold start" of the Langevin dynamics, i.e. set $X_0 = -2$ as initialization, and then follow the dynamics on p7 of `s18_diffusion.pdf` with $\nabla \log p \leftarrow \nabla \log p_2$ for a very large time t (i.e. achieving the stationary distribution). What is the distribution of X_t? Why?

- [1 bonus point] Suppose that we use a "warm start" of the Langevin dynamics, i.e. set $X_0 \sim p_1$ to be random. Then follow the Langevin dynamics with $\nabla \log p \leftarrow \nabla \log p_2$ for a very large time t (i.e. achieving the stationary distribution). What is the distribution of X_t? Why?

Problem 3.

Suppose that W_t is a standard Brownian motion process p6 of `s18_diffusion.pdf`, and $X \sim p$ is independent of the Brownian motion, where p is a distribution on $\mathbb{R}^d$. Let $p_t \sim p * \mathcal{N}(0, 2tI)$ be the distribution of $X + \sqrt{2}W_t$.

- If $p = \mathcal{N}(0, I)$, which of the following is/are correct? Why?

  1. For the stochastic differential equation (SDE) $dX_t = \sqrt{2}dW_t$ with initialization $X_0 \sim p$, we have that $X_t \sim p_t$.
  2. For the differential equation $dX_t = -\nabla \log p_t(X_t)dt$ with initialization $X_0 \sim p$, we have that $X_t \sim p_t$.
  3. For the SDE $dX_t = 9\nabla \log p_t(X_t)dt + 10\sqrt{2}dW_t$ with initialization $X_0 \sim p$, we have that $X_t \sim p_t$.
  4. For the SDE $dX_t = 99\nabla \log p_t(X_t)dt + 10\sqrt{2}dW_t$ with initialization $X_0 \sim p$, we have that $X_t \sim p_t$.
  5. The distribution of the whole sample paths $(X_t)_{t\geq 0}$ (not just the marginal at time t) in part 1) and 2) are the same.

- [1 bonus point for correct choice and 2 bonus point for correct derivations] If p is a general distribution, not necessarily Gaussian. Make the selection again, and please explain.
  Hint: we can use a similar integration by parts method on the Langevin process slide p7 of `s18_diffusion.pdf`*. Taylor expand to the second order terms and use $E[dW^2] = dt$ and $\mathbb{E}[dW_t|X_t]$=0.*