EQ2415 – Machine Learning and Data Science
HT22

Tutorial 9 A. Honoré, A. Ghosh

1 Approximate inference for graphical models

Suppose that $\mathbf{x} \in \mathbb{R}^d$ is an observable random variable and that $\mathbf{z} \in \mathbb{R}^n$ is a latent variable. We model the relation of the these two variables with a graphical model which allow us to calculate the joint distribution $p(\mathbf{x}, \mathbf{z})$. We call the evidence of the data, the log likelihood of the data $\ln p(\mathbf{x})$.

1.1 Variational inference

1.1.1 Evidence Lower Bound

Question 1 Using standard laws of propability, propose two ways to calculate the evidence and explain why they are not feasible in practice.

Question 2 Suppose that we have access to an approximate distribution $q_\phi(\mathbf{z}|\mathbf{x})$ of the true posterior $p(\mathbf{z}|\mathbf{x})$ where ϕ is a set of parameters. Show that $E_{q_\phi(\mathbf{z}|\mathbf{x})}[\ln \frac{p(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}]$ is a lower bound for $\ln p(\mathbf{x})$.

Recall:

- The KL-divergence between two probability distribution p and q:

$$D_{KL}(q||p) = E_q\left[\ln \frac{q}{p}\right] \geq 0 \tag{1}$$

Question 3 Let us now introduce variational auto-encoders. Let us introduce parameters for the conditional distribution $p_\theta(\mathbf{x}|\mathbf{z})$. Show that maximizing the ELBO consists in maximizing a "reconstruction" cost: $E_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\ln p_\theta(\mathbf{x}|\mathbf{z})\right]$ and minimizing a "prior matching" term: $D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$.

Question 4 In the context of an auto-encoder (See figure 1), what quantity can be interpreted as the encoder for a vector $\mathbf{x}$ and what quantity can be interpreted as a decoder of a vector $\mathbf{z}$?

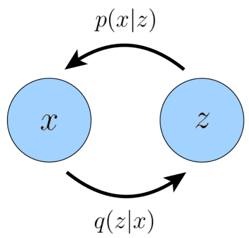

Figure 1: Graphical model for variational autoencoders.[?]

Question 5 Let us consider a generalization of VAEs: Markovian Hierarchical VAEs (Markovian HVAE, Figure 2).

Question 5a Factorize the joint distribution

$$p_\theta(\mathbf{x}, \mathbf{z}_1, \dots, \mathbf{z}_T), \tag{2}$$

in terms of the quantities on the edges of the graph in figure 2?

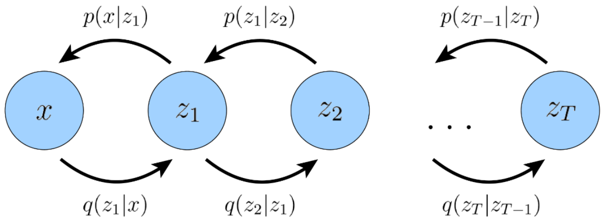

Figure 2: Markovian HVAE [?]

Question 5b Similarly, factorize the posterior of the Markovian HVAE:

$$q_\phi(\mathbf{z}_1, \dots, \mathbf{z}_T|\mathbf{x}). \tag{3}$$

1.1.2 Variational diffusion models (VDM)

The easiest way to think of a Variational Diffusion Model (VDM) is simply as a restricted Markovian Hierarchical Variational Autoencoder. The architecture, depicted on figure 3, is behind stable diffusion models.

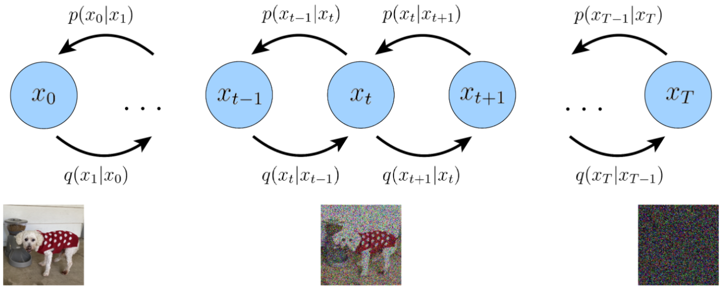

Figure 3: A visual representation of a Variational Diffusion Model; x_0 represents true data observations such as natural images, x_T represents pure Gaussian noise, and x_t is an intermediate noisy version of x_0. Each $q(x_t|x_{t-1})$ is modeled as a Gaussian distribution that uses the output of the previous state as its mean. [?]

2 Variational distributions

2.1 Factorized approximation

Consider a factorized variational distribution $q(\mathbf{z})$ of the form:

$$q(\mathbf{z}) = \prod_{i=1}^{M} q_i(\mathbf{z}_i) \tag{4}$$

Question 1: By using the technique of Lagrange multipliers, verify that the minimization of the Kullback-Leibler divergence $KL(p||q)$ with respect to one of the factors $q_j(\mathbf{z}_j)$, keeping all other factors fixed, leads to the solution:

$$q_j^\star(\mathbf{z}_j) = \int p(\mathbf{z}) \prod_{i \neq j} d\mathbf{z}_i = p(\mathbf{z}_j) \tag{5}$$