ADRL 2024 - Final Exam

27 November 2024

Instructions: (a) The exam is for **40 marks** for **180 minutes**. (b) Please refrain from writing much text but answer to the point with appropriate mathematical equations.

1. (a) Mathematically define a latent variable model (1).
   (b) Given a latent variable model, derive a bound on the log-likelihood (2) ELBO (EM Algo)
   (c) Give out an iterative algorithm for optimizing the previously derived lower bound and prove its correctness (1 + 3)

2. (a) Suppose the f-function for an f-divergence is defined as follows $f(t) = t\,ln\,t - ln\left(\frac{(1+t)}{2}\right)^{(t+1)}$. Give an expression for f-divergence between two density functions $p(x)$ and $q(x)$ (1)
   (b) Formulate an optimization problem to use the above f-divergence as a generative model (2)
   (c) With the above f-div, derive an expression for the optimal T function, that occurs in the optimization problem (3).

3. (a) Suppose $\mathbf{x}$ is a d dimensional continuous random variable. If $F_{\mathbf{x}}$ represents its distribution function, find the distribution of $F_{\mathbf{x}}$. Can the previous result be used for re-parametrization of $\mathbf{x}$, if so how? (3)
   (b) Consider a $d-$dimensional binary data. Derive an expression for the loss function of the VAE with the following form on the Decoder - The decoder models every data dimension independently as a Bernoulli RV (2)
   (c) Suppose the prior on the latent variable is also learnable in a VAE. Derive an expression for the optimal prior distribution that maximizes the ELBO in the VAE (3)

4. (a) Derive an expression to express the t^{th} latent variable in the forward process of a DDPM in terms of the data sample (2).
   (b) Suppose there are two multi-variate Gaussian distributions with different means. The Variances are identity matrices multiplied by two different scalars, α and β. Derive an expression for the KL divergence between the two. (1)
   (c) How is inference accomplished in a DDPM. Write down the algorithm (2).

Q5 Consider the following **discretized forward process:**

$$x_{k+1} = x_k + \frac{x_1 - x_k}{T - t_k}\Delta t + \sigma(t_k)\sqrt{\Delta t}\,\xi_k,$$

where:

- x_k is the state at time step $t_k = k\Delta t$,
- x_1 is the target terminal state at $t = T$,
- $\sigma(t_k)$ is a time-dependent noise coefficient,
- $\xi_k \sim \mathcal{N}(0,1)$ is independent Gaussian noise.

Answer the following questions:

1. Derive the **continuous-time SDE** that corresponds to the given discretized process in the limit $\Delta t \to 0$. (1)

2. Identify the drift and diffusion terms in the continuous SDE, and explain the role of each term in the context of the bridge process. (1)

3. Using Anderson's result, write the **reverse SDE** corresponding to the forward SDE derived in Part (a). (1)

4. If the log probability gradient $\nabla_x \log p_t(x)$ is given by $\nabla_x \log p_t(x) = -\frac{x-\mu(t)}{\sigma(t)^2}$, simplify the reverse SDE. (2)

5. Discretize the reverse SDE obtained in Part (b) into an iterative form:

$$x_{k-1} = x_k - [\text{Drift Term}]\,\Delta t + \sigma(t_k)\sqrt{\Delta t}\,\bar{\xi}_k,$$

where $\bar{\xi}_k \sim \mathcal{N}(0,1)$ represents reverse-time noise. (2)

6. Explicitly write the drift term, including the effects of the score function $\nabla_x \log p_t(x)$. (1)

7. Discuss how the time-dependent noise scaling $\sigma(t)$ influences the forward and reverse processes. (1)

8. If $\sigma(t)$ decreases linearly with t, explain the qualitative behavior of the reverse process as $t \to 0$. (1)

Consider a continuous-time state equation:

$$\dot{h}(t) = \mathbf{A}h(t) + \mathbf{B}x(t),$$

where $h(t)$ is the state vector, $x(t)$ is the input vector, $\mathbf{A}$ and $\mathbf{B}$ are system matrices. The input $x(t)$ is assumed to be constant over the sampling interval $[k\Delta, (k+1)\Delta]$ (called a zero-order hold) i.e. $x(t) = x[k], \quad t \in [k\Delta, (k+1)\Delta]$. The general solution to this state equation is:

$$h(t) = e^{\mathbf{A}(t-t_0)}h(t_0) + \int_{t_0}^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}x(\tau)\,d\tau,$$

where t_0 is some initial time at which $h(t_0)$ is known.

1. Write the solution to the state equation for the interval $[k\Delta, (k+1)\Delta]$. (1)

2. Consider the discrete version of the solution: (3)

$$h[k+1] = \bar{\mathbf{A}}h[k] + \bar{\mathbf{B}}x[k],$$

where $h[k] = h(k\Delta)$. Using the general solution provided, obtain the equivalent discrete-time relation above, and express the state matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ as functions of the matrix exponential $e^{\mathbf{A}}$. Are $\bar{A}$ and $\bar{B}$ time-invariant or time-variant?