**Deep Learning: Algorithms and Applications**
Peking University
Spring 2023
Prof. Yadong Mu

# Homework 2

**Due Date:** ***June 5th***

## 1. Variational Auto-Encoder

VAE is the typical generative model which has a lot of applications. The encoder portion of a VAE yields an approximate posterior distribution $q(z|x)$, and is parameterized on a neural network by weights collectively denoted $\theta$. Hence we more properly write the encoder as $q_\theta(z|x)$. Similarly, the decoder portion of the VAE yields a likelihood distribution $p(x|z)$, and is parameterized on a neural network by weights collectively denoted $\phi$. Hence we more properly denote the decoder portion of the VAE as $p_\phi(x|z)$. The output of the encoder are parameters of the latent distribution, which is sampled to yield the input into the decoder.

a) **Draw the model's framework of VAE**

b) **Derive the objective function e.g. the Evidence Lower Bound (ELBO)** using KL divergence.

[*HINT: 1) Please include the re-parametrisation trick implemented in VAE*
*HINT: 2) You may need to use Bayes theorem:* $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$]

c) **Derive the closed form VAE loss: Gaussian latents, based on question b)**. e.g. Say we choose:

$$p(z) \to \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{(x-\mu_p)^2}{2\sigma_p^2}\right)$$

and

$$q_\theta(z \mid x_i) \to \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(x-\mu_q)^2}{2\sigma_q^2}\right)$$

## 2. Policy based Reinforcement Learning

Please write down the detailed derivation of the policy gradient method in the episodic case, e.g. **Show that** $\nabla J(\boldsymbol{\theta}) \propto \mathbb{E}_\pi[\sum_a q_\pi(s,a)\nabla\pi(a|s)]$

Some definitions you may need:

$$J(\boldsymbol{\theta}) = v_{\pi_\theta}(s_0)$$

where $v_{\pi_\theta}$ is the true state value function for $\pi_\theta$, the policy determined by $\boldsymbol{\theta}$

state value function:

$$v_\pi = \mathbb{E}_\pi[G_t|S_t = s] = \sum_a \pi(a|s)q_\pi(s,a)$$

state-action value function:

$$q_\pi(s,a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a] = \sum_{s',r} p(s',r|s,a)(r + v_\pi(s'))$$

where we have the return defined as:

$$G_t = R_{t+1} + R_{t+2} + ...$$