University of Illinois Spring 2018

CS 446: Machine Learning
Homework

Due on Tuesday, April 10, 2018, 11:59 a.m. Central Time

1. **[2 points]** KL Divergence

   (a) [1 point] What is the expression of the KL divergence $D_{KL}(q(x)||p(x))$ given two continuous distributions $p(x)$ and $q(x)$ defined on the domain of $\mathbb{R}^1$?

   Your answer:

   (b) [1 point] Show that the KL divergence is non-negative. You can use Jensen's inequality here without proving it.

   Your answer:

2. **[3 points]** In the class, we derive the following equality:

$$\log p_\theta(x) = \int_z q_\phi(z|x) \log \frac{p_\theta(x,z)}{q_\phi(z|x)} dz + \int_z q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z|x)} dz$$

   Instead of maximizing the log likelihood $\log p_\theta(x)$ w.r.t. θ, we find a lower bound for $\log p_\theta(x)$ and maximize the lower bound.

   (a) [1 point] Use the above equation and your result in 1(b) to give a lower bound for $\log p_\theta(x)$.

   Your answer:

   (b) [1 point] What do people usually call the bound?

   Your answer:

   (c) [1 point] In what condition will the bound be tight?

   Your answer:

3. **[2 points]** Given $z \in \mathbb{R}^1$, $p(z) \sim \mathcal{N}(0,1)$ and $q(z|x) \sim \mathcal{N}(\mu_z, \sigma_z^2)$, write $D_{KL}(q(z|x)||p(z))$ in terms of σ_z and μ_z.

   Your answer:

4. **[1 points]** In VAEs, the encoder computes the mean μ_z and the variance σ_z^2 of $q_\phi(z|x)$ assuming $q_\phi(z|x)$ is Gaussian. Explain why we usually model σ_z^2 in log space, i.e., modeling $\log \sigma_z^2$ instead of σ_z^2 when implementing it using neural nets?

   Your answer:

5. **[1 points]** Why do we need the reparameterization trick when training VAEs instead of directly sampling from the latent distribution $\mathcal{N}(\mu_z, \sigma_z^2)$?

   Your answer: