ASSIGNMENT 4: THEORY OF GENERATIVE MODELS [IFT6135]

JOSEPH D. VIVIANO

1. Reparameterization Trick of Variational Autoencoders

1.1. **Transformation of Gaussian Noise.**

1.2. **Encoders vs. Mean Fields.**

2. Importance Weighted Autoencoder

2.1. **IWLB as a Lower Bound on log Likelihood.**

2.2. **IWLB with k=2 is tighter than ELBO with k=1.**

3. Maximum Likelihood for Generative Adversarial Networks

The original GAN objective can be writen as:

$$\max_D \mathbf{E}_{pD}(x)\big[\log D(x)\big] + \mathbf{E}_{p_G}\big[\log(1 - D(G(z)))\big]; \quad \max_G \mathbf{E}_{p_G}\big[\log D(G(z))\big]$$

Note that the definition of an optimial discriminator using this notation is:

$$D^*(\mathbf{x}) = \frac{p_D(\mathbf{x})}{p_G(\mathbf{x}) + p_D(\mathbf{x})} \tag{3.1}$$

The goal here is to find the maximum likelihood objective (cost function) instead of the negative log liklihood objective to apply to samples coming from the generator $\mathbf{E}_{p_G}\big[f(D(G(z)))\big]$, where $f(D(G(z)))$ must be found.

As a reminder, the maximum likelihood estimate in this case would be:

$$\hat{\theta} \in \{\arg\max_{\theta \in \Theta} \{(\theta\,; x)\}$$

Date: April 2018.

Each step of learning in a GAN consists of reducing the expectation $f(x)$ run on a bunch of samples pulled from generator G, which we express as

$$\mathbf{E}_{x\ p_G} f(x) \tag{3.2}$$

here, p_G represents a sample pulled from the probability distribution generated by G. First, let's take partial derivative with respect to the weights θ on a sample $x\ p_G$ from the generator and represent it as an integral. Then we applied Leibniz's rule, and sub in the identity $\frac{\partial}{\partial\theta} p_G(x) = p_G(x)\frac{\partial}{\partial\theta} log p_G(x)$:

$$\begin{aligned} \frac{\partial}{\partial\theta}\mathbf{E}_{x\ p_G} f(x) = \int f(x)\frac{\partial}{\partial\theta} p_G(x) \\ = \int f(x)\frac{\partial}{\partial\theta} p_G(X)dx \\ = \int f(x)p_G(x)\frac{\partial}{\partial\theta} log p_G(x) \end{aligned} \tag{3.3}$$

While tells us that we can express the aformentioned expectation (3.2) as:

$$\mathbf{E}_{x\ p_G} f(x)\frac{\partial}{\partial\theta} log p_G(x) \tag{3.4}$$

This tells us the maximum likelihood can be found given:

$$f(x) = -\frac{p_D(x)}{p_G x} \tag{3.5}$$

Now if we assume that our optimal discriminator $D^*(\mathbf{x})$ from (3.1) is the logistic sigmoid $\sigma(a(x))$ then

$$\sigma(a(x)) = \frac{p_D(\mathbf{x})}{p_G(\mathbf{x}) + p_D(\mathbf{x})} \tag{3.6}$$

And it follows that

$$f(x) = -\exp(a(x)) \tag{3.7}$$

which is our function f such that the objective corresponds to maximum liklihood.

Université de Montréal
E-mail address: joseph@viviano.ca