Training Deep Generative Models: Variations on a Theme

Philip Bachman and Doina Precup
McGill University, School of Computer Science

- This work is primarily a didactic exercise — to help tidy up our mental workspace and make room for new ideas.
- We want to see the relationships among existing methods more clearly, and better understand their underlying axes of variation.
- We modestly extend the variational free energy, by comparing the full distributional behaviours of a pair of sample generating systems.
- Many methods for training deep generative models are easily described in terms of the resulting LL bound.

The General Likelihood Bound

- We assume a pair of distributions p and q, where:

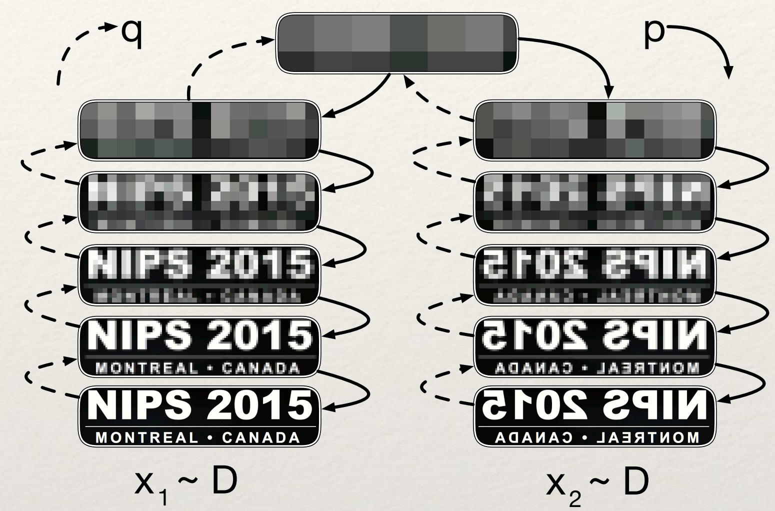
$$q(x, z_0 ..., z_n) \quad \text{and} \quad p(x, z_0, ..., z_n)$$

- It's then easy to show that:

$$\mathrm{KL}(q(x, z_0, ..., z_n) \,||\, p(x, z_0, ..., z_n)) - \mathop{\mathbb{E}}_{q(x)} [\log q(x)] \geq \mathop{\mathbb{E}}_{q(x)} [-\log p(x)]$$

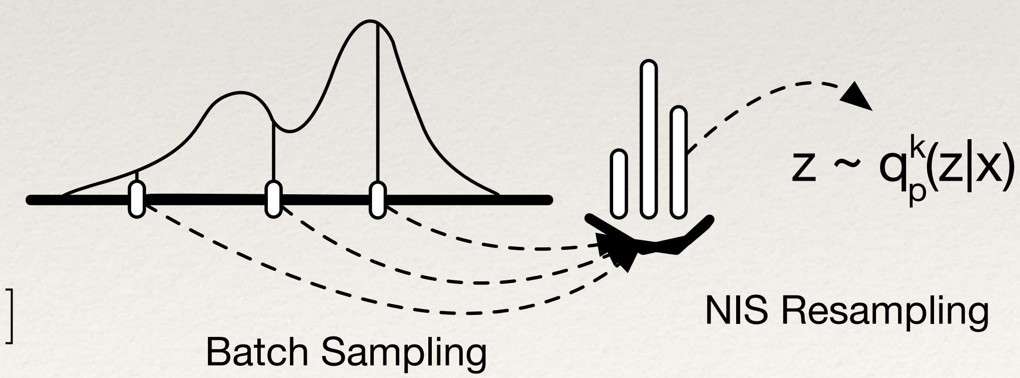
- In many cases, this bound is convenient to optimize.
- It gives maximum likelihood if we define q(x) = D(x).

Deep Unsupervised Learning using Non-Equilibrium Thermodynamics



Basic idea: define q that gradually destroys structure in D, and learn to invert the process.

Importance Weighted Autoencoders



Basic idea: define a *meta* $q_p^k(z\,|\,x)$ using a NIS correction towards the *true posterior* $p(z\,|\,x)$.