Neural Variational Learning in Undirected Graphical Models

Volodymyr Kuleshov   Stefano Ermon

Stanford University

December 2016

Undirected graphical models

Markov Random Fields (MRFs)

An MRF is a probability distribution of the form

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z(\theta)}, \quad \text{where } Z(\theta) = \int_x \tilde{p}_\theta(x) dx,$$

where $\tilde{p}_\theta(x) = \exp(\theta \cdot x)$ is an unnormalized probability and $Z(\theta)$ is the partition function.

Undirected graphical models

Markov Random Fields (MRFs)

An MRF is a probability distribution of the form

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z(\theta)}, \quad \text{where } Z(\theta) = \int_x \tilde{p}_\theta(x) dx,$$

where $\tilde{p}_\theta(x) = \exp(\theta \cdot x)$ is an unnormalized probability and $Z(\theta)$ is the partition function.

This work proposes variational lower bounds on MRF log-likelihood:

$$\begin{aligned} \log p_\theta(\mathcal{D}) &= \frac{1}{n} \sum_{i=1}^{n} \theta \cdot x_i - \log Z(\theta) \\ &\geq \mathcal{L}(p_\theta, q_\phi) \end{aligned}$$

An Upper Bound on Partition Function

Consider an importance sampling estimate of $Z(\theta)$ with proposal q:

$$Z(\theta) = \int_x \tilde{p}_\theta(x) dx = \int_x \frac{\tilde{p}_\theta(x)}{q(x)} q(x) dx = \mathbb{E}_{q(x)} \frac{\tilde{p}_\theta(x)}{q(x)} dx.$$

An Upper Bound on Partition Function

Consider an importance sampling estimate of $Z(\theta)$ with proposal q:

$$Z(\theta) = \int_x \tilde{p}_\theta(x) dx = \int_x \frac{\tilde{p}_\theta(x)}{q(x)} q(x) dx = \mathbb{E}_{q(x)} \frac{\tilde{p}_\theta(x)}{q(x)} dx.$$

Upper bound on partition function

The importance sampling variance is a natural upper bound on $Z(\theta)$

$$\underbrace{\mathbb{E}_{q(x)} \left[\left(\frac{\tilde{p}(x)}{q(x)} - Z(\theta) \right)^2 \right]}_{\text{Importance sampling variance}} = \underbrace{\mathbb{E}_{q(x)} \left[\frac{\tilde{p}(x)^2}{q(x)^2} \right]}_{\text{upper bound on partition function}} - Z(\theta)^2 \geq 0$$

Choice of Approximating Distribution q

We use a flexible family for q that includes auxiliary variables a.

Auxiliary-variable models

Let $\tilde{p}(z, a) = \tilde{p}(z)p(a|z)$ and $q(z, a) = q(z|a)q(a)$. Then

$$\mathbb{E}_{q(a,z)}\left[\frac{p(a|z)^2\tilde{p}(z)^2}{q(z|a)^2q(a)^2}\right] \geq \mathbb{E}_{q(a,z)}\left[\frac{\tilde{p}(z)^2}{q(z)^2}\right] \geq Z(\theta)^2.$$

Choice of Approximating Distribution q

We use a flexible family for q that includes auxiliary variables a.

Auxiliary-variable models

Let $\tilde{p}(z, a) = \tilde{p}(z)p(a|z)$ and $q(z, a) = q(z|a)q(a)$. Then

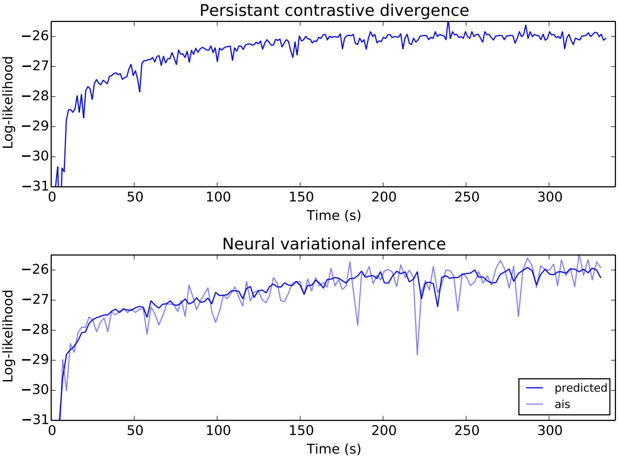
$$\mathbb{E}_{q(a,z)}\left[\frac{p(a|z)^2\tilde{p}(z)^2}{q(z|a)^2q(a)^2}\right] \geq \mathbb{E}_{q(a,z)}\left[\frac{\tilde{p}(z)^2}{q(z)^2}\right] \geq Z(\theta)^2.$$

Different instantiations of $q(z|a)$ lead to variants of:

- Non-parametric variational inference (Gershman et al., 2012)
- Auxiliary deep generative models (Maaloe et al, 2016)
- Markov chain variational Inference (Salimans et al., 2015)

Results

Training an RBM with 100 hidden units on sklearn digit dataset



The end

Thank you!

For more details come see our poster (#29)