MLE/VAE for nonlinear ICA

Bryon Aragam

August 2, 2022

Suppose $X \in \mathbb{R}^n$ and $Z \in \mathbb{R}^m$ with $X = f(Z) + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$. Assuming Z is normally distributed with independent marginals, this is equivalent to the following latent variable model:

$$\begin{aligned} Z &\sim N(0, I) \\ X|Z &\sim N(f(Z), \sigma^2 I). \end{aligned} \tag{1}$$

This is a special case of the well-known *nonlinear ICA* model.

Let $\varphi(u; \mu, \Sigma)$ denote the density of a $N(\mu, \Sigma)$ random variable and $p(x, z)$ denote the joint density under the model (1). It is easy to see that

$$p(x, z) = p(x|z)p(z) = \varphi(x; f(z), \sigma^2 I)\varphi(z; 0, I) \tag{2}$$

$$\implies p(x) = \int \varphi(x; f(z), \sigma^2 I)\varphi(z; 0, I)\,\mathrm{d}z. \tag{3}$$

Now suppose we let g_θ denote a family of deep neural network distributions parametrized by θ (i.e. weights, biases, etc.). To approximate the marginal density $p(x)$, we replace f with g_θ and try to find the choice of θ that maximizes the observed data likelihood.

Given n observations $x^{(i)} \overset{\text{iid}}{\sim} p(x)$, we wish to solve the following maximum likelihood problem:

$$\max_{\theta, \sigma} \underbrace{\sum_{i=1}^{n} \int \varphi(x^{(i)}; g_\theta(z), \sigma^2 I)\varphi(z; 0, I)\,\mathrm{d}z}_{:=\ell(\theta, \sigma)}. \tag{4}$$

There are two approaches to this:

1. *Direct MLE.* Directly solve the MLE problem (4) by computing gradients of $\ell(\theta, \sigma)$ wrt θ and σ. For *arbitrary* models, this is intractable, but worst-case thinking does not apply to special cases like (1).

2. *Variational inference and VAEs.* Use the ELBO to obtain a lower bound on the likelihood $\ell(\theta, \sigma)$ and optimize the ELBO using SGD.

Exercise 1. Simulate the model (1). Try different settings: $n = m = 1$, $n > m > 1$, f polynomial or sigmoid, f is a neural network, etc. Visualize the latent space and generate data from the model. Use this to simulate data for the rest of the exercises.

Exercise 2. Can you compute the gradients of $\ell(\theta, \sigma)$ directly? If so, can you run SGD directly on the MLE? If yes to all of the above, try to implement it.

Exercise 3. Design a VAE architecture for this problem, compute the corresponding ELBO, and implement it.