STA561: Probabilistic machine learning

Variational Inference (11/4/13)

Lecturer: Barbara Engelhardt *Scribes: Tracy Schifeling, Alireza Samany, and Matt Dickenson*

1 Introduction

2 Ising Model

Before proceeding with variational inference, it is helpful to review the Ising model. The main idea behind the Ising model is a lattice of unobserved variables $(x_1, ...x_n)$, each with its own (noisy) observation $(y_1, ..., y_n))$.

For example, suppose our goal is to reconstruct a denoised image given noisy observations of the pixels. We can think of the lattice as the pixels in a black and white image ($x_i \in \{-1, 1\}$), with a noisy grayscale observation of the pixels ($y_i \in R$). More generally, we wish to draw inferences about the unobserved lattice X from the observed values Y. Figure 1 illustrates an Ising model for $n = 9$, with the latent nodes colored white and the observed nodes shaded grey.

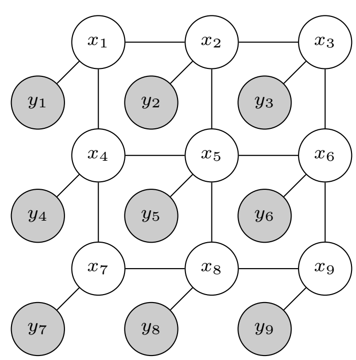

Figure 1: Schematic of an Ising Model

We now define the potential functions of the Ising model:

$$\begin{aligned} \psi_s(x_s) &= p(y_i|x_i) \equiv L_i(x_i) \\ \psi_{st}(x_s x_t) &= W_{st} x_s x_t \end{aligned}$$

Continuing with our image example above, we could set $W_{st} = 1$. In general, we set W to positive values if we want neighbors to agree, and negative values if we want them to differ.

Let $N(i)$ be a function that returns the first-degree neighbors of node i. For example, in Figure 1, calling $N(x_1)$ would return nodes x_2 and x_4.

Now we can specify functions for our prior:

$$
\begin{aligned}
p(x) &= \frac{1}{z_0} \exp\{-\sum_{i=1}^{n} \sum_{j \in N(i)} x_i x_j\} \\
p(y|x) &= \prod_{i=1}^{n} \exp\{-L_i(x_i)\}
\end{aligned}
$$

From this, we have the posterior:

$$
p(x|y) = \frac{1}{z} \exp\{-\sum_{i=1}^{n} \sum_{j \in N(i)} x_i x_j - \sum_{i=1}^{n} L_i(x_i)\}
$$

2.1 Mean Field Version of the Ising Model

Having seen an example of a basic Ising model, we now turn our attention to how we can analyze the mean field version of such a model. We do this by "breaking" the edges between the latent variables. We add a mean value (or variational parameter) μ to each x, such that $\mu_i = \mathbb{E}[x_i]$. The new structure is illustrated in Figure 2, with the same color coding as above.

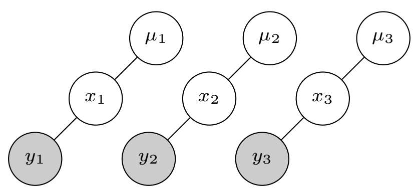

Figure 2: Mean Field Version of an Ising Model

$$
\begin{aligned}
q(x) &= \prod_{i=1}^{n} q_i(x_i) \\
\log(q_i(x_i)) &= \mathbb{E}_q[\log \tilde{p}(x)]
\end{aligned}
$$

We can maximize $\log(q_i(x_i))$ with a coordinate ascent method, as discussed previously in class.

Now, we rewrite the ELBO in terms of the Ising model:

$$
\begin{aligned}
\log(q_i(x_i)) &= \mathbb{E}_{q_i}[x_i \sum_{j \in N(i)} x_j + L_i(x_i) + c] \\
q_i(x_i) &\propto \exp\{x_i \sum_{j \in N(i)} \mu_j + L_i(x_i)\}
\end{aligned}
$$

Thus,

$$
q_i(x_i = +1) = \frac{\exp\{\sum_{j \in N(i)} \mu_j + L_i(+1)\}}{\sum_{x_i' \in \{+1,-1\}} \exp\{\sum_{j \in N(i)} \mu_j + L_i(x_i')\}}
$$

Note the resemblance of this function to a sigmoid function $\frac{1}{1+q_i(x_i=-1)}$. Effectively, this q_i is an approximation of the marginalized posterior, or basically a Gibbs step.

For the mean field, we now iteratively updateL

$$\begin{aligned} \mu_i &= +1(q_i(x_i = +1)) + -1(q_i(x_i = -1)) \\ z_i &\propto \exp[\mathbb{E}[\log p(z|x)]] \end{aligned}$$

In the model, we have three parameters of interest: μ_1, μ_2, and μ_{12}, where $\mu_{12} = \mathbb{E}[\psi_{12}(x_1 x_2)]$. We have the constraint $0 \leq \mu_{12} \leq \mu_1, \mu_2$. We limit acceptable values to those within the portion of the simplex that satisfies this constraint.

If we take a slice of this convex polytope, it is a triangle.

3 Loopy Belief Propogation