208891 Probabilistic Graphical Models **Spring 2019**

Lecture 17: March 6

Lecturer: Donlapark Pornnopparath

17.1 Ising model

Motivated from statistical physics, the Ising model is a simple mathematical model of a square lattice of "spins", where each spin can be in one of two states (+1 or −1). Each of these spins only interacts with its neighbors, giving out low energy if their states agree and high if they are opposite. The overall structure can change overtime due to external factors such as heat.

We will try to denoise an image using mean field variational inference on the Ising model. Here, the nodes in the model represent the pixels values $\boldsymbol{y}$ of a noisy image, each of which is connected to $x_i \in \{-1, 1\}$, a hidden pixel value of the clean image. The joint probability of $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$$

where

$$p(\boldsymbol{x}) = \frac{1}{Z_0}\exp(-E_0(x)).$$

Here, E_0 is the total energy given by

$$E_0(x) = -\sum_{i=1}^{n}\sum_{j\in N(i)} W_{ij}x_i x_j.$$

We assume that the noise is Gaussian with a fixed variance σ^2 , yielding the marginal likelihood

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_i p(\boldsymbol{y}_i|\boldsymbol{x}_i) \propto \exp\Big(\sum_i f_i(x_i)\Big)$$

$$f_i(x_i) = -\frac{(y_i - x_i)^2}{2\sigma^2}.$$

Therefore, the posterior is proportional to

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z}\exp\left(E_0(\boldsymbol{x}) - \sum_i f_i(\boldsymbol{x})\right).$$

From here, the mean field approximation is a product of pdf of Poisson variables x_i parametrized by μ_i

$$q(x) = \prod_i q(x_i, \mu_i).$$

To derive an approximate solution, we compute

$$\begin{aligned}\mathbb{E}_{q_{-i}} \log\tilde{p}(\boldsymbol{x}) &= \mathbb{E}_{q_{-i}}\left[x_i \sum_{j\in N(i)} W_{ij}x_j + f_i(x_i) + \text{const}\right] \\ &= x_i \sum_{j\in N(i)} W_{ij}\mu_j + f_i(x_i) + \text{const.}\end{aligned}$$

which leads to

$$\begin{aligned} q_i(x_i) &\propto \exp\left(\mathbb{E}\log\tilde{p}\right) \\ &\propto \exp\left(x_i \sum_{j\in N(i)} W_{ij}\mu_j + f_i(x_i)\right). \end{aligned}$$

We simply the expression by denoting $m_i = \sum_{j\in N(i)} W_{ij\mu_j}$, $f_i^+ = f_i(+1)$ and $f_i^- = f_i(-1)$. Thus the marginal posterior is given by

$$q_i(x_i = 1) = \frac{e^{m_i+f_i^+}}{e^{m_i+f_i^+} + e^{-m_i-f_i^+}} = \frac{e^{m_i+f_i^+}}{e^{m_i+f_i^+} + e^{-m_i-f_i^+}} = S(2a_i)$$

where S is the sigmoid function and

$$a_i = m_i + 0.5(f_i^+ - f_i^-).$$

Similarly, we have $q_i(x_i = -1) = S(-2a_i)$. Finally, we compute the mean of x_i.

$$\begin{aligned} \mu_i &= q_i(x_i = 1) - q_i(x_i = -1) \\ &= \frac{1}{1+e^{-2a_i}} - \frac{1}{1+e^{2a_i}} \\ &= \frac{e^{a_i}}{e^{a_i}+e^{-a_i}} - \frac{e^{-a_i}}{e^{a_i}+e^{-a_i}} \\ &= \tanh(a_i). \end{aligned}$$

This leads to the coordinate descent algorithm on the KL divergence given by

$$\mu_i^t = \tanh\left(\sum_{j\in N(i)} W_{ij}\mu_j^{t-1} + 0.5(f_i^+ - f_i^-)\right).$$

Sometimes we want to "smooth out" the updates by partially weighting on the previous value. This is called *damped updates.*

$$\mu_i^t = (1-\lambda)\mu_i^{t-1} + \lambda\tanh\left(\sum_{j\in N(i)} W_{ij}\mu_j^{t-1} + 0.5(f_i^+ - f_i^-)\right),$$

where $0 < \lambda < 1$ is the smoothing parameter. Without damping, we might get some "checkerboard" artifacts.