Mean field theory approximation for the Ising model

Ye Li

These notes give a short description of Mean Field Theory and focus on applying it to the Ising model for image. These notes assume you're familiar with basic probability and graphical models.

1 Mean Field Theory

The idea of the mean field theory approximation method is to use an approximating distribution $Q(.)$ to approximate the target distribution when the target distribution is intractable. So that we can obtain estimates such as marginal distributions on certain random variable(s) in the network or MAP estimator on the states of the random variables in the network. The approximating distribution usually is a distribution over a set of independent nodes but there could be structures/dependencies among the nodes. The later is so-called structured mean field and the dependencies should be defined at the beginning. Often, when the target full joint distribution is intractable, we can use the MFT approximation method to find an approximating distribution that's close to the target distribution and compute estimates on it. For the Ising model, we are looking for an approximating distribution $Q(\underline{\mathbf{S}})$ that's close to $P(\underline{\mathbf{S}}|\underline{\mathbf{I}})$

1.1 Free energy and approximating distribution

$Q(\underline{\mathbf{S}})$ is the approximating distribution. Free energy is defined as follows:

$$F_{MFT}(b) = \sum_{x} \sum_{S(x)\in\{0,1\}} (S(x) - I(x))^2 q_x(S(x)) +$$
$$\lambda \sum_{x} \sum_{y\in Nhb(x)} \sum_{S(x)\in\{0,1\},S(y)\in\{0,1\}} (S(x)-S(y))^2 q_x(S(x)) q_y(S(y)) + \sum_{x} \sum_{S(x)\in\{0,1\}} q_x(S(x)) log q_x(S(x)) \tag{1}$$

1.2 Kullback-Leibler divergence

To explain how the K-L divergence can measure the similarity between $P(.)$ and $Q(.)$, we need to write down the general form of K-L divergence:

$$KL(Q, P) = \sum_{x} q(x) ln \frac{q(x)}{p(x)} + \sum_{i} \lambda_i c_i(q)$$

where $c_i(q) = \sum_{x_i} q_i(x_i) - 1 = 0$ As we can see from the natural log term above that the K-L divergence measures the difference between the two distributions (only when q(x)=p(x) the natural log term =0 otherwise it will contribute to the total sum) and hence we need to minimize it to find the optimal distribution. In the Ising model, we have the K-L divergence written as follows:

$$KL(Q, P) = F_{MFT}(q) + logZ$$

1.3 Update equations for the Ising model

The update equation is for the steepest descent algorithm $q_x^{t+1} = q_x^t - \delta \frac{\partial KL(Q,P)}{\partial q_x} q_x^t$ and in order to derive the gradient term, we'd better clean up the free energy term first by letting $q_x(S(x) = 1) = q_x$ and $q_x(S(x) = 0) = 1 - q_x/$ Then, the free energy term becomes:

$$\begin{aligned} F_{MFT}(b) = &\sum_x \{(1 - I(x))^2 q_x + I(x)^2(1 - q_x)\} \\ &+ \lambda \sum_x \sum_{y \in Nbh(x)} \{q_x(1 - q_y) + (1 - q_x)q_y\} + \sum_x \{q_x log q_x + (1 - q_x) log(1 - q_x)\} \end{aligned} \quad (2)$$

And,

$$\frac{\partial KL(Q, P)}{\partial q_x} = \sum_x \{(1 - I(x))^2 - I(x)^2\} + \lambda \sum_x \sum_{y \in Nbh(x)} \{(1 - q_y) - q_y\} + \sum_x \{log q_x - log(1 - q_x)\}$$

Once we have the gradient we can then plug it in the update equation above to obtain q_x^{t+1}. We will need to do this for all values that the node X can take, namely, $q_x^{t+1}(S(x) = 1)$ and $q_x^{t+1}(S(x) = 0)$. The choice of δ is very important for this algorithm.