Phase Transition of Ising Model with Belief Propegation

Leon Ma [1]

Abstract

Ising models have traditionally been solved with Markov chain Monte Carlo methods, the most popular of which is the Gibbs' sampler. However, MCMC suffers from correlated samples and long mixing times. In recent years, a new class of algorithms based on variational inference has come to light. These algorithms promise faster convergence abit at the cost of convergence. Belief propagation is an inference algorithm that uses message passing in order to approximate the probability distribution over a factor graph. In this paper, we apply belief propagation to solve the Ising model and show that during belief propagation, the 2D Ising model exhibits a phase transition similar to simulations using MCMC. We compare metrics of phase transitions to MCMC approaches and known theoretical results and conclude that belief propagation can solve the Ising model to a high degree of accuracy and retain many interesting properties that the Ising model is known for.

1. Introduction

The Ising model is a mathematical model of ferromagnestim in statistical mechanics. Invented first by Wilhelm Lenz in 1920 and the equations of the one-dimensional Ising model were solved in by Lenz's student Ernst Ising in 1924. In 1944, Lars Onsager solved the equations for the two-dimensional case. As of today, no analytic solution has been found in dimensions three or higher. Over the last 100 years, many other problems such as percolation, min cut max flow, error correction, neural networks, and neurodegenerative diseases were shown to be very closely mathematically related to the Ising model. The Ising model's relatively simple formulation yet rich and general mathematical properties makes a *drosophila* of physics, mathematics, statistics, and computer science.

Since Onsager's solution, it is known that the Ising model exhibits a phase transition just like magnets in the real world. Above the a special temperature, the critical temperature (or more generally the critical point), the system under goes a rapid change, changing from an ordered magnetic state to a disordered non-magnetic state. This rapid change of order is fundementally related to the relationship between energy, entropy and free energy. When the system is at the critical point, it is said to be at criticality and many interesting properties such as scale invariance and long range correlations are present during the phase transition.

However, due to the difficulty of solving the Ising models in higher dimensions, many approxiomation methods have been devolved over the years. Statistical physicists developed mean field theory and the cavity method to find approximate analytic solutions. In more recent times, the rise of computational power allowed computer algorithms to approximate solutions to the Ising model. The most well known class of algorithms are Markov Chain Monte Carlo (MCMC) algorithms such as Metropolis-Hasting algorithm and Gibbs' sampling. Although these algorithms are powerful and still widely used, they suffer long convergence times and autocorrelated samples.

In this paper, we explore an alternative algorithm to approximately solve the Ising model called belief propagation. Belief propagation works by sending messages from each node of a factor graph to iteratively find the probability distrubition. We show that this method can calculate solutions comparable to Metropolis-Hastings and Gibbs sampling and show that solutions given by belief propegation exhibit a phase translation like predicted theoretically.

2. Background

2.1. Ising Model

The Ising model models ferromagnetic materials as a two-dimensional lattice where each lattice site represents an electron that is either spin up $\sigma_i = +1$ or spin down $\sigma_i = -1$. Each lattice point only interacts with it's adjacent neighbors. In this paper, we impose periodic boundary conditions. This is equivalent to the lattice wrapped around the surface of a torus. This simulates an infinitely large lattice and midigates boundary effects. Thus, the total energy (also known as the

[1]Department of Statistics, UC Berkeley. Correspondence to: Leon Ma <lma00@berkeley.edu>.

Hamiltonian H in physics literature) is

$$E(\sigma) = -J \sum_{(i,j)\in\mathcal{N}} \sigma_i\sigma_j - B\sum_j \sigma_j \tag{1}$$

where $(i,j)\in\mathcal{N}$ is understood to be the sum over all pairs of (i,j) that are adjacent neighbors, J is the physical constant knows as the coupling strength and B is the strength of the external magnetic field. We assume $J = 1$ for simplicity and no external magnetic field $B = 0$. Then our total energy is

$$E(\sigma) = -\sum_{(i,j)\in\mathcal{N}} \sigma_i\sigma_j \tag{2}$$

We notice that the energy is lower when σ_i and σ_j are the same sign and greater when σ_i and σ_j are different signs. The energy is minimized when the spins are alligned in the same direction, corresponding to a magnetized state. However, the entropy is maximized when the spins are misalligned corresponding to a demagnetized state.

The PMF of the system is given by the Boltzmann distribution

$$P(\sigma) = \frac{e^{-\beta E(\sigma)}}{Z} \tag{3}$$

where the inverse temperature $\beta = 1/(kT)$, k is Boltzmann's constant, and the normalization constant $Z = \sum e^{-\beta H(\sigma)}$. Given an initial square lattice of σ and inverse temperature β the goal is to find the equalibrium distrubition over σ.

2.2. Phase Transition

In physics a phase transition refers to a change of a system from one state to another. A classical example is ice melting into water. However, more generally it refers to the rapid change of a systems order parameter as a function of one or more of the systems parameters. This more general defintion can be applied to other fields such as rate distortion theory. The Ising model predicts a phase transition of a ferromagnetic system like an iron magnet from a demagnetized state to a magnetized state. It is known emperically that heating a magnet above a certain temperature known as the Curie temperature (770 °C for iron) causes rapid demagnetization. The characteristic feature of the phase transition is that the magnetization strength changes very rapidly over a short range of temperature similar to how ice melts into water over a narrow temperature of range of 0°C. This transition from order to disorder is due to the changing strength of energetic and entropic forces.

In nature, systems tend to minimize a thermodynamic quanity known as free energy. The free energy is given by

$$F = E - TS \tag{4}$$

where S is the thermodynamic entropy. At high temperatures the entropy term is dominant and at low temperature the energetic term is dominant. This explains why the spins are misalligned at high temperature. The thermal energy from the envirnoment is sufficent to flip the lattice site spins. This thermal function acts like a corrupting force constantly distrupting the spins of the system. It is also convient to define the free energy as

$$F = kT\ln Z \tag{5}$$

The normalization constant Z is called the partion function and it turns out it is very important. These two formulas are equivalent

$$\begin{aligned}
S &= -k\sum_i p_i \ln p_i && (6)\\
&= -k\sum_i \frac{e^{-\beta E_i}}{Z}\ln\left(\frac{e^{-\beta E_i}}{Z}\right) && (7)\\
&= -k\sum_i \frac{e^{-\beta E_i}}{Z}(-\beta E_i - \ln Z) && (8)\\
&= k\beta\sum_i \frac{e^{-\beta E_i}}{Z}E_i + k\sum_i \frac{e^{-\beta E_i}}{Z}\ln Z && (9)\\
&= k\beta\langle E\rangle + k\ln Z && (10)\\
&= k\beta E + k\ln Z && (11)
\end{aligned}$$

since the energy E is simply the energy averaged all over microstates $\langle E\rangle$. Plugging this into (4) yields

$$\begin{aligned}
F &= E - T(k\beta E + k\ln Z) && (12)\\
&= E - T(k\frac{1}{kT}E + k\ln Z) && (13)\\
&= E - E + kT\ln Z && (14)\\
&= kT\ln Z && (15)
\end{aligned}$$

The energy can also be defined in terms of the partition function as

$$E = -\frac{\partial \ln Z}{\partial \beta} \tag{16}$$
$$= -\frac{1}{Z}\frac{Z}{\beta} \tag{17}$$
$$= -\frac{1}{Z}\frac{\partial}{\partial \beta}\sum_i e^{-\beta E_i} \tag{18}$$
$$= -\frac{1}{Z}\sum_i -E_i e^{-\beta E_i} \tag{19}$$
$$= \sum_i E_i \frac{e^{-\beta E_i}}{Z} \tag{20}$$
$$= \langle E \rangle \tag{21}$$
$$= E \tag{22}$$

These equations will be relevant when we discuss the connection between belief propagation and minimization of the Bethe free energy.

By analyizing at what temperature the energy and entropy terms balance, Onsager solved the 2D Ising model analytitically in the case of no external magnetic field and showed that the critical temperature is

$$T_c = \frac{2J}{k \ln\left(1+\sqrt{2}\right)} \tag{23}$$

or in terms of $\beta_c = \frac{\ln\left(1+\sqrt{2}\right)}{2J}$. This derivation is very technical and will not be covered here.

The Ising model at the critical temperature exhibits many interesting properties. The first is the rapid chance of the order parameter, in the case of the Ising model, the magnetization given the average value of the spins.

$$\langle M \rangle = \frac{1}{N}\sum_{i=1}^{N} \sigma_i \tag{24}$$

In order to not give preference to major spin down or spin up configurations, it is convient to define the mangetization squared $\langle M^2 \rangle = \frac{1}{N}\sum_{i=1^N} \sigma_i^2$.

At the critical temperature, the system reaches a critical state where order and disorder are perfectly matched. Here, the model exhibits long range correlations not found at other temperature reigmes. The Ising model only takes into account interactions between neigboring lattice sites. Therefore, we expect the correlation between lattice sites to be very short and lattice sites long distances apart to act inpendendently. This is in general true except when the system is at criticality. When order and disorder are perfectly balanced, local interactions have a ripple effect, meaning that local interactions lead to long range communication. The correlation function is given by

$$C_{i,j} = \langle (\sigma_i - \langle \sigma_i \rangle)(\sigma_j - \langle \sigma_j \rangle) \rangle \tag{25}$$

where $\langle \cdot \rangle$ denotes the time average. Then we can define the correlation length function,

$$C(d) = \mathbb{E}\left[C_{i,j} \mid D(i,j) = d\right] \tag{26}$$

which is the correlation strength as a function of the distance between lattice sites.

Another interesting phenomona observed at criticality is scale invariance. Suppose for each 3 by 3 block of spins, we replace that block with the spin corresponding to the majority spin value out of those 9 smaller spins. Apply this procedure to the whole lattice and we have a new lattice with a third of the distance scale. This procedure is called a course graining. We can continue this procedure indefinitely on an infinitely large lattice. At the critical point, all these lattices of varying length scales look the same, in other words there is no characteristic scale. To show this idea graphically, we plot the distribution of the cluster sizes where a cluster is a group of lattice points with the same spin. If the distrubition of the cluster sizes is scale invariant, they should follow a power law

$$f(x) = Ax^{-\gamma} \tag{27}$$

where $f(x)$ is the probability of observing a cluster size with size x. We can visualize this as a straight line on a log-log plot where the slope is γ.

3. Methods

3.1. Metropolis-Hasting

In the previous sections we have described the mathematical background behind the Ising model and theory of phase transitions. However, solving the Ising model analytitically is extremely difficult and in practice computational methods are used. The most common of these methods are Monte Carlo Markov Chain methods. The simpliest example of MCMC is the Metropolis-Hasting algorithm. Metropolis-Hasting works by building an implict Markov Chain model of the system and sampling from the stationary distrubtion of the chain to generate samples of the system's true probability distrubtion. This works because when the Ising model is at thermodynamic equalibrium, it's Markov chain satifies the detailed balance equations

$$\pi_\mu P(\mu \to \nu) = \pi_\nu P(\nu \to \mu) \tag{28}$$

where μ and ν are states, p_μ is the probability that the system is in state μ and $P(\mu \to \nu)$ is the probability that the system will transition from state μ to state ν. Then we have that

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{\pi_\nu}{\pi_\mu} \tag{29}$$
$$= \frac{e^{-\beta E_\nu}/Z}{e^{-\beta E_\mu}/Z} \tag{30}$$
$$= e^{-\beta(E_\nu - E\mu)} \tag{31}$$
$$= e^{-\beta \Delta E} \tag{32}$$

where the change in energy $\Delta E = E_\nu - E\mu$. Suppose that we start in state μ. At each step of the algorithm, we flip a random spin $\sigma_j \mid j \sim U\{1, 2, \ldots, n\}$ to create a new state ν. If $E_\nu < E_\mu$ then $\Delta E < 0$ and the system would enter the lower energy state so $P(\nu \to \mu) = 1$. This gives us $P(\mu \to \nu) = e^{-\beta\Delta E}$. Note that the probability that the system moves to a higher energy state is non zero due to random fluctuations due to the thermal energy. Furthermore as T increases, β decreases causing $e^{-\beta\Delta E}$ to increase meaning that spin flips are more likely at higher temperature. If $E_\mu < E_\nu$ we can similarly define $P(\mu \to \nu) = 1$ and $P(\nu \to \mu) = e^{-\beta(-\Delta E)}$. This gives us the acceptance probability for $\mu \to \nu$ to be

$$\alpha = \min(1, e^{-\beta\Delta E}) \tag{33}$$

Thus, we only need to calculate the difference in energy $\Delta E = E_\mu - E_\nu$ when determining wether to accept or reject a canidate state. However, as we'll see, this naive method of randomly flipping a spin can lead to pathelogical a Markov chain.

3.2. Gibbs Sampler

The Gibbs sampler begins with $\mathbf{X} = (x_1, \ldots, x_2)$ with joint distribution $p(x_1, \ldots, x_n)$. Denote the ith sample $\mathbf{X}^{(i)} = (x_1^1, \ldots, x_n^{(i)})$. Then the Gibbs sampler samples

$$x_j^{(i+1)} \sim P(x_j^{(i+1)} \mid x_1^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_{j+1}^{(i+1)}, \ldots, x_n^{(i+1)}) \tag{34}$$

In our specific case we choose the index $j \sim U\{1, 2, \ldots, n\}$. $x_j^{(i+1)}$ is given by 34 and $x_{\setminus j}^{(i+1)} = x_{\setminus j}^{(i)}$. The advantage of Gibbs sampling is that in some cases is it simpler ot sample from the conditional distrubtion then it is to marginalize the joint distrubtion.

In the case of our Ising model, at each time step we sample a lattice point j uniformly. Then according to our update rule, the probability of σ_j being spin up is given by

$$P(\sigma_j = +1) = \frac{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots))}{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots)) + \exp(-\beta E(\ldots, \sigma_j = -1, \ldots))} \tag{35}$$

where . . . denotes the spins of the other lattice points at the previous time step. $P(\sigma_j = -1)$ is given similarly. Notice that this update rule does not depend on the value of σ_j at the previous time step. This update rule is referred to as the heat bath update rule because the orientation of each spin depends only on the energy transferred from the thermal energy of the heat bath.

Now lets try to reformulate this update rule to better match our Metropolis-Hasting update rule. Suppose that initially $\sigma_j = -1$. Let the event that $\sigma_j = +1$ at the next time step be denoted $\mu \to \nu$ as before. Then

$$P(\mu \to \nu) = P(\sigma_j = +1) = \tag{36}$$
$$\frac{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots))}{\exp(-\beta E(\ldots, \sigma_j = +1, \ldots)) + \exp(-\beta E(\ldots, \sigma_j = -1, \ldots))} \tag{37}$$
$$= \frac{\exp(-\beta E_\nu)}{\exp(-\beta E_\nu) + \exp(-\beta E_\mu)} \tag{38}$$
$$= \frac{\exp(-\beta(E_\mu + (E_\nu - E\mu)))}{\exp(-\beta(E_\mu + (E_\nu - E\mu))) + \exp(-\beta E_\mu)} \tag{39}$$
$$= \frac{\exp(-\beta(E_\mu + \Delta E))}{\exp(-\beta(E_\mu + \Delta E)) + \exp(-\beta E_\mu)} \tag{40}$$
$$= \frac{\exp(-\beta E_\mu)\exp(-\beta\Delta E)}{\exp(-\beta E_\mu)\exp(-\beta\Delta E) + \exp(-\beta E_\mu)} \tag{41}$$
$$= \frac{\exp(-\beta\Delta E)}{\exp(-\beta\Delta E) + 1} \tag{42}$$
$$= \frac{1}{1 + \frac{1}{\exp(-\beta\Delta E)}} \tag{43}$$

Following the same derivation for the probability that σ_j is initially at spin up then is flipped spin down gives the same expression. Thus, the probability of the spin flip under the heat bath update rule is

$$\alpha = \left(1 + \frac{1}{\exp(-\beta\Delta E)}\right)^{-1} \tag{44}$$

We can just replace our Metropolis-Hasting algorithm with this new acceptance probability showing that the Gibbs sampler is special case of the Metropolis-Hasting algorithm. This class of update rules that depend soley on the Boltzmann distrubtion is known as Glauber dynamics. We shall see in practice this update rule generally results in faster convergence.

3.3. Belief Propegation

4. Experiments

5. Discussion

6. Conclusion