Learning Phase Transtion in Ising Model by Confusion

Rahul M N - PH17B009

1 The Ising Model

The Ising Model is a simple model in Statistical Mechanincs of a solid that describes the phenomenon of ferromagnetism. It consists of discrete variables which represnts the direction of magnetic dipole moment of atomic spins, which can be represented as +1 or -1. The standard hamiltonian of the model considers only the nearest neighbour interaction. The simplified standard Hamiltonian used in this study is

$$H = -\sum_{<i,j>} J_{ij} s_i s_j - B \sum_i s_i, \tag{1}$$

where B is the magnetic field. The first summation is over nearest neighbors of each lattice point, while the second summation is over all the N^2 lattice points. J is the exchange energy for interaction between the spins .J and J' describe the interaction between the nearest neighbour. We shall consider no external magnetic field for the simulation ($B = 0$) . Hence the simplified standard Hamiltonian used in this study is

$$H = -\sum_{<i,j>} J_{ij} s_i s_j \tag{2}$$

However, more generalized version of the Hamiltonian may include long range interactions.

$J > 0$ implies that the energy of the system is lowered when the spins align in the same direction which is also known as the ferromagnetic behaviour. $J < 0$ implies that the energy is the lowest when the spins are anti-parallel, also known as anti ferromagnetism. The thermodynamic equilibrium is said to be achieved when it reaches a steady state minimum hence $J > 0$ and $J < 0$ represents ferromagnetism and Paramagnetism respectively. The total Magnetism of the the system can be given by

$$M = |\sum_{<i,j>} s_i| \tag{3}$$

There is scope for spontaneous Magnetisation where the spins are aligned even in the absence of an external magnetic field which leads us to stud the critical temperature ($\beta = 1/T_c$) at which the spontaneos magnetisation happens. The energy of any given state of the system can be obtained by

$$E = -\sum_i s_i s_j \tag{4}$$

2 Monte Carlo Simulations.

Monte Carlo algorithms are a class of Algorithms which uses samples of random numbers sampled according to a probability distribution which can accurately compute the observables of a dynamical system. Calculating $< O >$ for an observable O can be done analytically by

$$< O >= \frac{1}{Z} \sum_{\mu} O_{\mu} e^{-\beta E_{\mu}} \tag{5}$$

To compute this, a sample of N states μ_1,μ_2,μ_3....,μ_n are constructed which are in accordance with a probability distribution P_{μ}. Then an estimator(O_N) of $< O >$ is defined as

$$O_N = \frac{\sum_i^N O_{\mu_i} e^{-\beta E_{\mu_i}} P_{\mu_i}^{-1}}{e^{-\beta E_{\mu_i}} P_{\mu_i}^{-1}} \tag{6}$$

For large samples P_{μ} is the frequency of the system being in μ which leads us to

$$< O >= \lim_{N \to \infty} O_N \tag{7}$$

Here we wish to determine a value P_{μ} for convergence of the above equation through random sampling. In importance sampling , we choose a state μ and add it to the sample with probability P_{μ}. This ensures efficiency by choosing states which have higher contribution to calculate $< O >$. which gives us with the expression

$$O_N = \frac{1}{N} \sum_{i=1}^{N} O_{\mu_i} \tag{8}$$

2.1 Metropolis Algorithm

The metropolis algorithm is one of the most used Monte Carlo algorithm. We shall use it to compute the observables of the system like pecific heat, c, or the magnetization, m and others. We will see the dry run of the algorithm for a single Monte Carlo sweep.

1. Initialize a spin configuration. In our case the initial spin configuration will be random, corresponding to a high temperature initial configuration.

2. Compute the energy of the configuration using (2).

3. Pick a random site i on the lattice, and calculate the energy of the configuration if $s_i \to -s_i$. Then compute the change in energy of the whole configuration, ΔE.

4. If ΔE is negative, accept the change of the spin-configuration by letting $s_i \to -s_i$.

5. If ΔE is non-negative, generate a random number $r \in \langle 0, 1 \rangle$. If $e^{-\beta \Delta E} > r$, change the spin configuration by letting $s_i \to -s_i$.

6. Repeat the process N^2 times. This defines a single Monte Carlo sweep, t.

In short any move that reduces the energy will be accepted and any move that increases the energy will be rejected. The Drawback of this Algorithm is it's computational cost. To run this algorithm on huge lattices is hard.

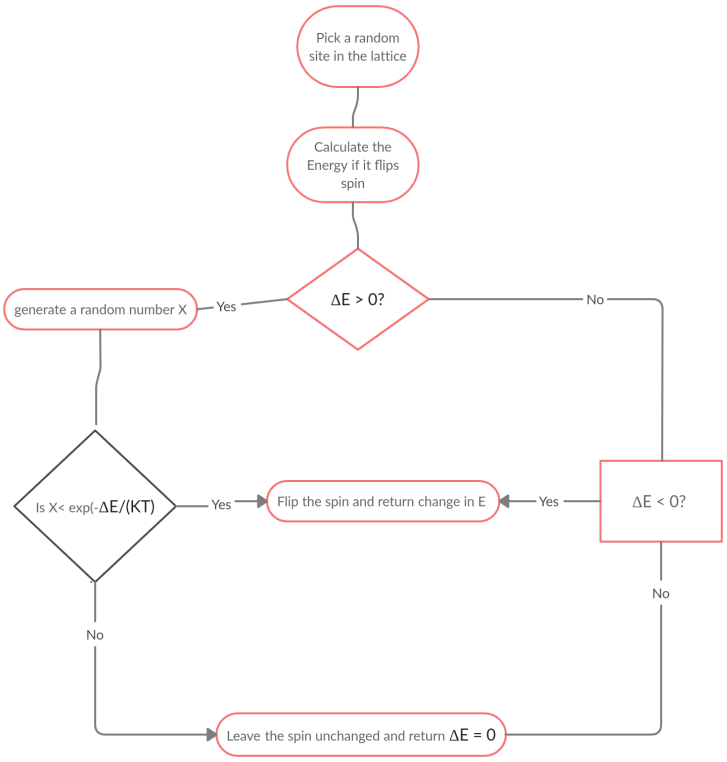


Figure 1: Flowchart of Metropolis Algorithm

2.2 Data Sampling Parameters

The parameters used to sample the data were as follows:

Size of the lattice (N) : 16

Critical temperature (T_c) : 2.269

Equillibrium steps : 2000

Temperature points : 1000

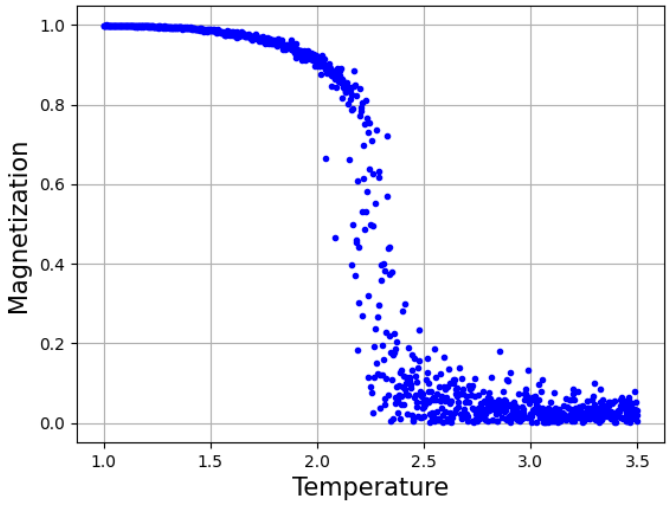


Figure 2: Magnetization visualized in the data generated.

3 Learning by Confusion

To understand why the confusion scheme was required in the first place we shall try to train a simple neural network to classify the two phases in the data and analyze the results.

3.1 Training a Simple NN

A neural network with an input layer of 100 nodes, and a single hidden layer and an output layer of two neurons to classify high and low temperature phases was built and trained on the generated dataset. This helps to comprehend how the network is able to learn the order parameter of the phase transition with high accuracy.

On comparing the magnetization curve with the plot obtained from the trained neural network the high accuracy of learning is apparent. The cross over corresponds to the Critical Temperature T_c

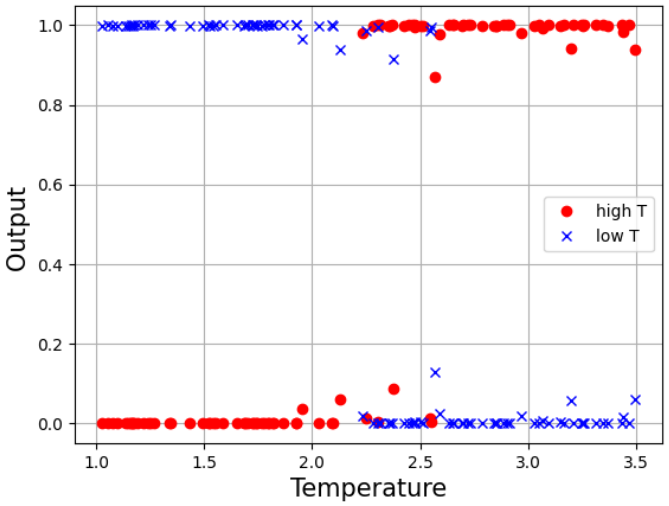


Figure 3: Phase classification as done by a simple neural network without confusion

3.2 Confusion

Confusion can be thought of as a combination of both supervised and unsupervised learning methods. It is a method of training where the training data is deliberately mislabelled in order to confuse the network which enables it to extract features and learn rather than depending the user defined labels to train.

To understand how exactly the data is mislabelled, the following sequence of events can be followed:
1. Consider the range (a,b) within which the T_c lies. Let c lie between a and b such that the data can be classified into two classes or phases (c may or may not be the actual critical point).

2. Propose c' to be the critical point. Everything below c' will be labelled 0 and the ones above c' as 1 and the network is trained.

3. Data is assumed to have two different structures(features) above and below c.

Now let us consider the following cases:

1. c' = a, In this case both the features are labelled as 1(higher temperature phase) hence the NN predicts with maximum accuracy.

2. When c=c' , data on either sides of c is labelled correctly and hence NN predicts accordingly without any confusion.

3. Consider $a < c' < c.$, NN sees the same data between (a,c') and (c',c) but labels them differently (hence the Confusion).

4. Similarly when $c < c' < b$ same data has different label between (c,c') and (c',b). In both 3 and 4 it will choose to learn the majority label. the model performance $P(c')$ is

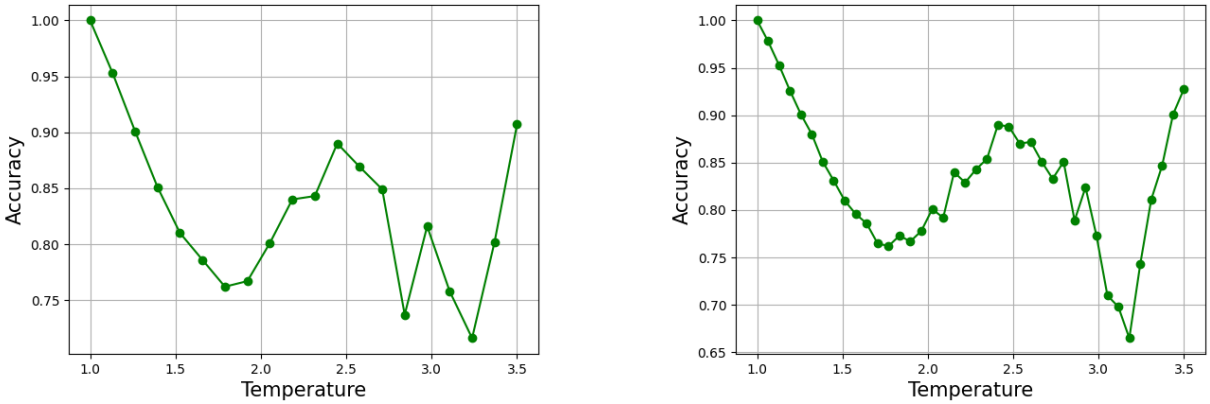
$$P(c') = 1 - \frac{min(c - c', c' - a)}{b - a} \tag{9}$$

The reason why this scheme was used in the first place was to increase the model performance at the crossover or T_c. Classifying phases at either extreme of the temperature scale is an easy task, however at the crossover model performance maybe low beacuse the data on either side of the t_c is similar.

3.3 Method

A feed forward neural network with a single hidden layer was trained.A lattice size of $N = 16$ was used to generate the spin configurations. The input layer had N^2 nodes and the output layer had 2 neurons corresponding to the two classes of phases involved. Sigmoid activation function was used in the network along with Adam Optimizer. Accuracy was used to measure the model performance. The model was trained over 2000 epochs.

4 Results and Conclusion



(a) Step size of 50

(b) Step size 25

Figure 4: Model performance for 2 different step sizes.

A W-shaped curve can be observed with a peak in between. The peak tells us that the model performance at the critical temperature is maximum which was our initial goal. The peak also gives us the Critical Temperature T_c for the phase

transition. The Network trained does not depend on order parameter or other specifics, therefore can be used to explore new phases and transitions. In conclusion the scheme of confusion increases the model performance at the crossover point or T_c.

References

[1] van Nieuwenburg, E., Liu, YH. Huber, S. Learning phase transitions by confusion. Nature Phys 13, 435–439 (2017). https://doi.org/10.1038/nphys4037