The Ising Model and The Statistical Mechanics of Learning

compiled by D.Gueorguiev 12/30/2024

# The Ising Model

## What the Ising Model is about

The model consists of discrete variables that represent magnetic dipole moments of atomic “spins” that can be in one of two states (+1 or -1). The spins are arranged in a graph, usually a lattice allowing each spin to interact with its neighbors. Neighboring spins that agree have a lower energy than those that disagree; obviously the system attempts to occupy the lowest possible energy state, but heat would disturb this equilibrium, thus creating the possibility of different structural phases. The model allows identification of phase transitions.

The Ising problem without an external field can be equivalently formulated as a graph maximum cut (Max-Cut) problem (ref. [2]) that can be solved via combinatorial optimization.

## More formal discussion

Consider a set of lattice sites, each with a set of adjacent sites (e.g. a graph) forming a -dimensional lattice. For each lattice site there is a discrete variable such that , representing the site’s spin. A *spin configuration*, is an assignment of spin value to each lattice site.

For any two adjacent sites there is an *interaction* . Also a site has an *external magnetic field* interacting with it. The *energy* of a configuration is given by the Hamiltonian function

where the first sum is over pairs of adjacent spins (every pair is counted once). The notation indicates that sites and are nearest neighbors. The magnetic moment is given by .

The configuration probability is given by the Boltzmann distribution with inverse temperature :

where , and the normalization constant

# References

[1] <https://en.wikipedia.org/wiki/Ising_model>

[2] <https://en.wikipedia.org/wiki/Maximum_cut>

[3] <https://en.wikipedia.org/wiki/Partition_function_(statistical_mechanics)>

[4] [Information Theory and Statistical Mechanics, E.T. Jaynes, Department of Physics, Stanford University, 1957](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/articles/Information_theory_and_statistical_mechanics_part1_Jaynes_1957.pdf)

[5] [Information Theory and Statistical Mechanics, E.T. Jaynes, Department of Physics, Stanford University II, 1957](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/articles/Information_theory_and_statistical_mechanics_part2_Jaynes_1957.pdf)

[6] [Exactly Solved Models in Statistical Mechanics, R.J. Baxter, The Australian National University, 1989](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/books/Exactly_Solved_Models_In_Statistical_Mechanics_Baxter_1989.pdf)

[7] [Statistical Mechanics of Deep Learning Neural Networks - The Back-Propagating Kernel Renormalization, Q. Li, H. Sompolinsky, 2021](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Statistical_Mechanics_of_Deep_Linear_Neural_Networks-The_Back-Propagating_Kernel_Renormalization_Li_Sompolinsky_2021.pdf)

[8] [Statistical Mechanics of Learning from Examples, HS Seung, H. Sompolinsky, 1994](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Statistical_mechanics_of_learning_from_examples_1992-seung-somplinsky.pdf)

[9] [Visualizing the Loss Landscape of Neural Nets, H. Li et al, 2018](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Visualizing_the_Loss_Landscape_of_Neural_Nets_Li_2018.pdf)

[10] [Qualitatively Characterizing Neural Network Optimization Problems, Ian Goodfellow et al, 2015](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Qualitatively_Characterizing_Neural_Network_Optimization_Problems_Goodfellow_2015.pdf)

[11] [The Loss Surfaces of Multilayer Networks, A. Choromanska et al, 2015](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/The_Loss_Surfaces_of_Multilayer_Networks_Choromanska_2015.pdf)

[12] [Spin Glasses and the Statistical Mechanics of Protein Folding, JD Bryngelson, PG Wolynes, 1987](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/bryngelson-wolynes-1987-spin-glasses-and-the-statistical-mechanics-of-protein-folding.pdf)

[13] [Mimicking The Folding Pathway to Improve Homology-Free Protein Structure Prediction, J. DeBartolo, 2008](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/debartolo-et-al-2009-mimicking-the-folding-pathway-to-improve-homology-free-protein-structure-prediction.pdf)

[14] [Funnels in Energy Landscapes, K. Klemm et al, 2007](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Funnels_in_Energy_Landscapes_Klemm_2007.pdf)

[15] [Understanding Protein Folding with Energy Landscape Theory, K. Plotkin, JN Onuchic, 2002](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Understanding_protein_folding_with_energy_landscape_theory_PlotkinOnuchic_part1.pdf)

[16] [Landscape Statistics of low autocorrelated binary string problem, F. Ferreira et al, 2000](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Landscape_statistics_of_the_low_autocorrelated_binary_string_problem_Ferreira_2000.pdf)

[17] [Energy Landscapes, Supergraphs, and "Folding Funnels" in Spin Systems, Piotr Garstecki et al, 1999](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Energy_Landscapes_Supergraphs_and_Folding_Funnels_in_Spin_Systems_Garstecki_1999.pdf)

[18] [From Levinthal to pathways to funnels, K. Dill, HS Chan, 1997](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/From_Levintal_to_pathways_to_funnels_Dill_1997.pdf)

[19] [Towards a new Theory of Learning: Statistical Mechanics of Deep Neural Networks, 2019, Charles H Martin](https://calculatedcontent.com/2019/12/03/towards-a-new-theory-of-learning-statistical-mechanics-of-deep-neural-networks/)

[20] [Why does Deep Learning work?, Charles H Martin, 2015](https://calculatedcontent.com/2015/03/25/why-does-deep-learning-work/)

# Appendix

## Microscopic Definitions of Thermodynamic Concepts

### Microstates and Macrostates

**Definition** *Microstate* –

A specific configuration of a system that describes the precise positions and momenta of all the individual particles or components that make up the system. Each microstate has a certain probability of occurring during the course of the system’s thermal fluctuations.

**Definition** *Macrostate* -

Macrostate of a system refers to its macroscopic properties, such as its temperature, pressure, volume, and density. More precisely, a microstate is a particular set of values of energy, the number of particles, and the volume of an isolated thermodynamic system. Microstates represent specific possible ways in which the system can achieve a particular microstate. Thus, a microstate is characterized by a probability distribution of possible states across a certain statistical ensemble of all microstates. This distribution describes the probability of finding the system in a certain microstate. In the *thermodynamic limit*, the microstates visited by a macroscopic system during its fluctuations all have the same macroscopic properties.

The *Thermodynamic (macroscopic) limit* of system is the limit for a large number of particles where the volume is taken to grow in proportion with the number of particles. Thus, the thermodynamic limit is defined as the limit of a system with a large volume, where the particle density is held fixed:

Example: Tossing a coin twice

Statistical Mechanics links the empirical properties of a system to the statistical distribution of an ensemble. All macroscopic thermodynamic properties of a system may be calculated from the partition function that sums of all its microstates.

## On Partition Functions

The *Partition function* describes the statistical properties of a system in thermodynamic equilibrium. Partition functions are functions of the thermodynamic state variables such as temperature and volume. Most of the aggregate thermodynamic variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives. The partition function is dimensionless.

Each partition function represents a particular statistical ensemble which in turn corresponds to a particular free energy. The most common statistical ensembles have named partition functions-

*Canonical partition function* : applies to the canonical ensemble, in which the system is allowed to exchange heat with the environment at fixed temperature, volume and number of particles.

*Grand canonical partition function* : applies to grand canonical ensemble , in which the system can exchange both heat and particles with the environment, at fixed temperature, volume and chemical potential.

Classical discrete system

For a canonical ensemble that is classical and discrete, the canonical partition function is defined as

where

* is index for the microstates of the system;
* is the thermodynamic beta coefficient given with where is the Boltzmann constant;
* is the total energy of the system in the respective microstate.

is also known as the Boltzmann factor

Derivation of canonical partition function for classical discrete system

According to the second law of thermodynamics, a system assumes a configuration of maximum entropy at thermodynamic equilibrium. We seek a probability distribution of states that maximizes the discrete Gibbs entropy

subject to two physical constraints:

1. the probabilities of all states add to unity (second axiom of probability):

2. in the canonical ensemble, the system is in thermal equilibrium, so the average energy does not change over time;