

An Outline of (Classical) Statistical Mechanics and Related Concepts in Machine Learning

Chang Liu

Tsinghua University

June 1, 2016

What is covered

- What is Statistical mechanics developed for? What is its basic principle? What does it achieve? What tools/techniques is employed?
- What is the origin of some concepts in machine learning? Entropy, Boltzmann distribution (canonical distribution), (Helmholtz) free energy, etc.

Overview

- 1 Introduction
- 2 The Microcanonical Ensemble
- 3 The Canonical Ensemble
- 4 Free Energies

Introduction

What is statistical mechanics

On the study of statistical properties of a macroscopic system in equilibrium that consists of a huge number of particles:

- Thermodynamics: based on the four thermodynamic laws summarized from experiments. Weak dependence on microscopic laws of motion. Studies the Relations of thermodynamic quantities (e.g. $PV = Nk_B T$, $E = \frac{3}{2}Nk_B T$)
- (Equilibrium) statistical mechanics: based on statistics, conclusions from microscopic laws of motion, and ergodic hypothesis. Studies the ensembles (an ensemble is the distribution over microstates of a specific macroscopic system). Relations of thermodynamic quantities in thermodynamics can be derived by taking the large system limit in statistical mechanics.

The basic principle of statistical mechanics

The basic principle of statistical mechanics: the probability density over microstates of an isolated system in equilibrium is constant.

The principle is supported by:

- A microstate of an isolated (classical) system comprising N particles: the most detailed description of the system at some instant:

$s = (Q, P) = (q_{1x}, q_{1y}, q_{1z}, q_{2x}, \dots, q_{Nz}, p_{1x}, p_{1y}, p_{1z}, \dots, p_{Nz})$ is a point in phase-space \mathbb{R}^{6N} . This description arises from classical mechanics.

- The time evolution of a microstate is governed by Hamilton's equation: $\dot{q} = \nabla_p H(s), \dot{p} = -\nabla_q H(s)$, where $H(s)$ is the Hamiltonian, a function of a microstate s , and usually the total energy of the system.

- Hamiltonian is conserved while evolving:

$$dH(s)/dt = \nabla_q H \cdot \dot{q} + \nabla_p H \cdot \dot{p} = \nabla_q H \cdot \nabla_p H - \nabla_p H \cdot \nabla_q H = 0.$$

- Liouville's theorem: $d\rho/dt = 0$.

The basic principle of statistical mechanics

- Ergodic hypothesis (first proposed by L. E. Boltzmann): informally, from any initial microstate, the time evolution will lead the system to pass every other microstates of the same energy.
For some systems like a three-body system, although chaotic, but not ergodic. For some systems like the hard sphere gas, ergodicity can be proven. We believe ergodicity holds for most systems with large numbers of interacting particles.
- For the equilibrium distribution over microstates, starting with state $s(0)$, for any other microstate A , there is a time t s.t. $s(t) = A$ (ergodicity), and $\rho(s(0), 0) = \rho(s(t), t) = \rho(A, t)$ (Liouville's theorem), $\rho(A, t) = \rho(A, 0)$ (equilibrium), so $\rho(A, 0) = \rho(s(0), 0)$. The equilibrium distribution is constant over microstates for an isolated system.

The Microcanonical Ensemble

The microcanonical ensemble

- The distribution over microstates of an isolated system (fixed energy, volume and particle number).
- According to the basic principle, the density $\rho(s) \propto \delta(E - H(s))$. Define $\Omega(E) = \int \delta(E - H(s)) ds = \frac{d}{dE} \int_{H(s) \leq E} ds$ (the volume of constant-energy region in the phase space / the number of microstates of constant energy), then $\rho(s) = \frac{1}{\Omega(E)} \delta(E - H(s))$.

Relation between $\rho(s)$ and $\rho(E)$

From the basic principle, $\rho(s) = f(H(s))$.

- For a general function $g(E)$, we have

$$\begin{aligned} \int g(H(s))ds &= \int \left(\int g(E)\delta(E - H(s))dE \right) ds \\ &= \iint g(E)\delta(E - H(s))dsdE = \int g(E)\Omega(E)dE. \end{aligned} \quad (1)$$

- Apply the above conclusion:

$$\begin{aligned} \rho(E) &= \int \rho(s, E)ds = \int f(H(s))\delta(E - H(s))ds \\ &\stackrel{\text{Eqn. (1)}}{=} \int f(y)\delta(E - y)\Omega(y)dy = f(E)\Omega(E). \end{aligned} \quad (2)$$

Introducing temperature and equilibrium entropy

Consider an isolated system consisting of two subsystems that share the total energy E (in heat contact). What is the energy of each subsystem when in equilibrium?

Let s_1 and s_2 be a microstate of each subsystem. The whole system is a microcanonical ensemble: $\rho(s_1, s_2) = \frac{1}{\Omega(E)}\delta(E - H_1(s_1) - H_2(s_2))$. So

$$\begin{aligned} \rho(s_1) &= \int \rho(s_1, s_2) ds_2 = \frac{1}{\Omega(E)} \int \delta(E - H_1(s_1) - H_2(s_2)) ds_2 \\ &\stackrel{\text{Eqn. (1)}}{=} \frac{1}{\Omega(E)} \int \delta(E - H_1(s_1) - E_2) \Omega_2(E_2) dE_2 \\ &= \frac{1}{\Omega(E)} \Omega_2(E - H_1(s_1)), \end{aligned} \quad (3)$$

and from Eqn. (2) $\rho(E_1) = \frac{1}{\Omega(E)} \Omega_1(E_1) \Omega_2(E - E_1)$. The mode of the energy distribution should satisfy:

$$0 = \frac{d\rho(E_1)}{dE_1} = \frac{1}{\Omega(E)} \left(\frac{d\Omega_1(E_1)}{dE_1} \Omega_2(E_2) - \Omega_1(E_1) \frac{d\Omega_2(E_2)}{dE_2} \right)$$

Introducing temperature and equilibrium entropy

It can be shown that $\rho(E_1)$ is very sharp so at equilibrium E_1 stays at the mode:

$$\frac{d \log \Omega_1(E_1)}{dE_1} = \frac{d \log \Omega_2(E_2)}{dE_2}.$$

If $\frac{d \log \Omega_1(E_1)}{dE_1} < \frac{d \log \Omega_2(E_2)}{dE_2}$, then $\frac{d\rho(E_1)}{dE_1} < 0$, so E_1 tends to decrease. So the subsystem with a smaller $\frac{d \log \Omega(E)}{dE}$ tends to give out its energy. We know that the two systems have the same temperature in equilibrium, and a higher temperature indicates the tend to give out energy, so we define the temperature T as

$$\frac{1}{T} = \frac{d \log \Omega(E)}{dE}.$$

Define the equilibrium entropy $S(E, V, N) = k_B \log \Omega(E, V, N)$ and the temperature is formally defined as

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_{V, N}.$$

An intuitive interpretation of $\rho(E_1)$ to be a very sharp distribution

Expand $\rho(E_1)$ around its mode E_1^* :

$$\begin{aligned} \rho(E_1) &\propto \Omega_1(E_1)\Omega_2(E - E_1) = \exp(S_1(E_1)/k_B + S_2(E - E_1)/k_B) \\ &= \exp\left[\left(S_1(E_1^*) + \frac{1}{2}\frac{\partial^2 S_1}{\partial E_1^2}(E_1 - E_1^*)^2 + S_2(E_2^*) + \frac{1}{2}\frac{\partial^2 S_2}{\partial E_2^2}(E_1 - E_1^*)^2\right)/k_B\right] \\ &\propto \exp[-(E_1 - E_1^*)^2/(2\sigma_E)], \end{aligned}$$

where $\sigma_E^2 = -k_B/(\partial^2 S_1/\partial E_1^2 + \partial^2 S_2/\partial E_2^2)$. E and S are extensive quantities (scale linearly with system size N), so $\partial^2 S/\partial E^2 \propto 1/N$ and the relative fluctuation $\sigma_E/E_1^* \propto \sqrt{N}/N = 1/\sqrt{N}$ vanishes in the limit $N \rightarrow \infty$.

The history of entropy

- Entropy in thermodynamics: Entropy is first proposed in thermodynamics by Rudolf Clausius. The fact that $\oint_C \frac{\delta Q}{T} = 0$ for any reversible loop C indicating that there exists a state function S and the difference for two states A and B is $S(B) - S(A) = \int_l \frac{\delta Q}{T}$ for any reversible path l from A to B .
- Entropy in equilibrium statistical mechanics: Namely the above definition $S = k_B \log \Omega$, which is proposed by Planck based on the idea of Boltzmann. For thermodynamic entropy, $dS = \frac{\delta Q}{T}$. For Boltzmann's entropy, $T = \left(\frac{\partial E}{\partial S}\right)_{V,N} = \frac{\delta Q}{dS}$, so $dS = \frac{\delta Q}{T}$ also holds. That is the relation between the two definitions.

The history of entropy

- Shannon's entropy in information theory:

$S = -k_S \sum_i p_i \log p_i = - \int \rho(x) \log \rho(x) dx$. Boltzmann's entropy:

$S = k_B \log \Omega = -k_B \log(1/\Omega) = -k_B \sum_{i=1}^{\Omega} (1/\Omega) \log(1/\Omega)$.

Shannon's entropy is the only continuous function of $p = (p_1, \dots, p_N)$ that satisfies: 1) $S(\frac{1}{N}, \dots, \frac{1}{N}) \geq S(p_1, \dots, p_N)$, and the equality holds if and only if p is uniform; 2) $S(p_1, \dots, p_{N-1}, 0) = S(p_1, \dots, p_{N-1})$; 3) for a joint distribution over (x, y) , $\mathbb{E}_y[S(x|y)] = S(x, y) - S(y)$.

The Canonical Ensemble

The canonical ensemble

- The distribution over microstates of a system with fixed temperature, volume and particle number.
- The only theoretical way to implement such a system is to let it in contact with another system large enough to maintain temperature while exchanging energy with the concerned system. The large enough system is called a heat bath (thermal energy reservoir).
- Deriving the distribution over microstates: denote the concerned system as 1 and the heat bath as 2. Then according to Eqn. (3), $\rho(s_1) \propto \Omega_2(E - H_1(s_1))$. According to the large heat bath assumption, $\partial S_2/\partial E_2 = 1/T$ is fixed. So $\partial \log \Omega_2/\partial E_2 = 1/k_B T$, solved as $\Omega_2(E_2) \propto \exp(E_2/k_B T)$. So $\rho(s_1) \propto \exp(-H_1(s_1)/k_B T)$, and

$$\rho(s) = \frac{1}{Z(\beta)} \exp(-\beta H(s)), \quad (4)$$

where $\beta = 1/k_B T$ and $Z(\beta)$ is the partition function. This is the Boltzmann distribution, or canonical distribution.

The canonical ensemble

- The Boltzmann distribution is in the exponential family with sufficient statistics $H(s)$ and natural parameter $-\beta$. So $\mathbb{E}[E] = -\partial \log Z / \partial \beta$. Entropy $S = -\int \rho(s) \log \rho(s) ds = \mathbb{E}[E]/T + k_B \log Z$.
- Define the Helmholtz free energy $A = -k_B T \log Z = \mathbb{E}[E] - TS$, then $S = -\frac{\partial A}{\partial T}$. Generally thermodynamic properties of a canonical ensemble are determined by the Helmholtz free energy through its derivatives.

Free Energies

Free energies

- The free energy of a specific system determines thermodynamic properties of the system.
- The free energy of a specific system is the part of internal energy that can be converted into useful work.

E.g. The Helmholtz free energy: $A = \mathbb{E}[E] - TS$ is the total internal energy excluding the heat that has to be transmitted into a cold bath when doing work, thus the available work.

- A free energy $F(x)$ integrates out other degrees of freedom (variables). $\exp(-F(x)/k_B T)$ is the volume of the region in phase space with constant x .

E.g. For the Helmholtz free energy, $\exp(-A(T)/k_B T) = Z(T)$ is the volume of constant-temperature region in phase space. It is the part that region with constant given temperature T takes in the phase space.

Free energies

- Free energies are Legendre transforms of energy. The Legendre transform of a convex function $f(x)$ is $f^*(p) = \min_x f(x) - xp$. It can be extended to Fenchel transform for convex conjugate. For differentiable $f(x)$, $f^*(p) = (f(x) - xp)_{p=f'(x)}$. The Helmholtz free energy $A(T)$ is the Legendre transform of $E(S)$:
 $A(T) = (E(S) - ST)$ since $T = \partial E / \partial S$.

Thanks!