STAT260 Mean Field Asymptotics in Statistical Learning		Lecture 3 - 01/27/2021
	Lecture 3: Ising Models	
Lecturer: Song Mei	*Scriber: Dimitris Papadimitriou*	*Proof reader: Taejoo Ahn*

1 Introduction

In this lecture we will study Ising models related to ferromagnetism and more specifically we will study the Curie-Weiss and the spin-glass model. We will also briefly be introducing basic concepts in statistical decision theory.

2 The Curie-Weiss model

2.1 The model

The Curie-Weiss model is used to model Ferromagnets, which at low temperatures become magnetized but in high temperatures they lose their magnetization. The critical temperature is the threshold when this phenomenon occurs. The configuration space of the model is $\Omega = \{+1, -1\}^n$ and the perturbed Hamiltonian is given by

$$H_\lambda(\boldsymbol{\sigma}) = -\frac{1}{2n}\sum_{i,j=1}^{n}\sigma_i\sigma_j + \lambda\sum_{i=1}^{n}\sigma_i, \tag{1}$$

We are interested in the average magnetization and we define the extensive magnetization to be

$$M(\boldsymbol{\sigma}) = \sum_{i=1}^{n}\sigma_i.$$

The perturbed Gibbs measure is given by

$$P_{\beta,\lambda}(\boldsymbol{\sigma}) \propto \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\}. \tag{2}$$

We are interested in calculating the average magnetization

$$m_\star(\beta,\lambda) = \lim_{n\to\infty}\langle M\rangle_{\beta,\lambda}/n. \tag{3}$$

2.2 Intuitions at limiting temperature

To obtain some intuition for $m_\star$ we consider the following scenaria. In the high temperature limit, i.e. $\beta \to 0, T \to \infty$, all the states have the same probability which means that $P_{0\lambda} = \mathrm{Unif}(\{\pm 1\}^n)$ and $m_\star(0,\lambda) = 0$. On the other hand, in the low temperature limit ($\beta \to \infty$, $T \to 0$) we consider three cases.

1. $\lambda = 0$: The minimum value in (1) is achieved at $\arg\min H_0(\boldsymbol{\sigma}) = \{\mathbf{1}_n, -\mathbf{1}_n\}$. The Gibbs measure converges to $P_{\infty,0} = (1/2)\delta_{\mathbf{1}_n} + (1/2)\delta_{-\mathbf{1}_n}$ and the average magnetization is $m_\star(\infty, 0) = 0$.

2. $\lambda = 0^+$: In this case a "slightly" positive λ breaks the symmetry in (1) and hence $\arg\min H_\lambda(\boldsymbol{\sigma}) = -\mathbf{1}_n$, $P_{\infty,0} = \delta_{-\mathbf{1}_n}$ and $\lim_{\lambda\to 0^+} m_\star(\infty,\lambda) = -1$.

3. $\lambda = 0^-$: This case is symmetric to the previous one, $\arg\min H_\lambda(\sigma) = \mathbf{1}_n$, $P_{\infty,0} = \delta_{\mathbf{1}_n}$ and $\lim_{\lambda\to 0^-} m_\star(\infty,\lambda) = 1$.

Intuitively λ represents an external magnetic field. Now we know how $m_\star(\beta,\lambda)$ behaves on the limit. In the following, we will derive an expression of $m_\star$ for arbitrary values of β and λ. From the analysis above, we may postulate that there exists a critical value of β_c at which the material changes properties.

3 Analysis of the Ising model

We will follow the following two steps:

- Calculate the free energy density $f(\beta,\lambda)$.
- Differentiate to obtain $m_\star(\beta,\lambda) = \partial_\lambda f(\beta,\lambda)$.

3.1 Calculating the free energy density

By definition $f(\beta,\lambda) = -(1/\beta)\phi(\beta,\lambda)$ where

$$\phi(\beta,\lambda) = \lim_{n\to\infty} (1/n)\log Z_n(\beta,\lambda)$$

and

$$Z_n(\beta,\lambda) = \sum_{\sigma\in\{\pm 1\}^n} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\}.$$

We start by computing the partition function Z_n. For that we re-write (1) as follows

$$H_\lambda(\boldsymbol{\sigma}) = -\frac{1}{2n}\Big(\sum_{i=1}^n \sigma_i\Big)^2 + \lambda\sum_{i=1}^n \sigma_i. \tag{4}$$

Now we define as the instantaneous magnetization as $\bar{m}(\boldsymbol{\sigma}) = (1/n)\sum_{i=1}^n \sigma_i$. Then we can further re-write the Hamiltonian as a function of $\bar{m}$

$$H_\lambda(\boldsymbol{\sigma}) = -\frac{n}{2}\bar{m}(\boldsymbol{\sigma})^2 + n\lambda\bar{m}(\boldsymbol{\sigma}). \tag{5}$$

We can also re-write

$$\begin{aligned} Z_n(\beta,\lambda) &= \sum_{m\in M_n}\sum_{\bar{m}(\boldsymbol{\sigma})=m} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\} &(6)\\ &= \sum_{m\in M_n} N_n(m)\exp\{\beta n m^2/2 - n\beta\lambda m\} &(7)\\ &\doteq \sum_{m\in M_n} \exp\{n\phi_{\mathrm{mf}}(m;\beta,\lambda)\}. &(8) \end{aligned}$$

Here $M_n = \{-1, \frac{-n+2}{n}, \ldots, \frac{n-2}{n}, 1\}$ is the possible value of average magnetization when the system size is n, and $N_n(m)$ represents the number of configurations such that $\bar{m}(\boldsymbol{\sigma}) = m$. By simple combinatorics, we have

$$N_n(m) = |\{\boldsymbol{\sigma} : \bar{m}(\boldsymbol{\sigma}) = m\}| = \binom{n}{(n(1+m))/2} \doteq \exp\{nH((1+m)/2)\}$$

where in the last step we use Stirling's formula and $H(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function. We used the notation $a_n \doteq b_n$, which means that

$$\lim_{n\to\infty}(1/n)\log(a_n) = \lim_{n\to\infty}(1/n)\log(b_n).$$

Finally

$$\phi_{\rm mf}(m;\beta,\lambda) = \frac{\beta}{2}m^2 - \beta\lambda m + H((1+m)/2)$$

is the mean field free entropy.

Using the Laplace method, we obtain

$$\phi(\beta,\lambda) = \lim_{n\to\infty} \frac{1}{n}\log \sum_{m\in M_n} \exp\{n\phi_{\rm mf}(m;\beta,\lambda)\} \tag{9}$$

$$= \max_{m\in[-1,1]} \phi_{\rm mf}(m;\beta,\lambda). \tag{10}$$

Finally we obtain that

$$f(\beta,\lambda) = -\frac{1}{\beta}\phi(\beta,\lambda) = \min_{m\in[-1,1]} \Big[-\frac{1}{2}m^2 + \lambda m - \frac{1}{\beta}H\Big(\frac{1+m}{2}\Big)\Big] \tag{11}$$

3.2 Calculating the average magnetization

Now in the second step we compute the derivative $m_\star(\beta,\lambda) = \partial_\lambda f(\beta,\lambda)$. Given that this expression contains the min function we cannot differentiate right away. For this reason we will be using the envelope theorem in our derivation.

Remark 1 (Envelope theorem (informal, without clarifying the assumptions))**.** *Let* $f(\lambda) = \min_m f_{\rm mf}(m;\lambda)$ *and the value that minimizes the argument inside the* min *operation* $m^* = \arg\min_m f_{\rm mf}(m;\lambda)$*. Then*

$$f'(\lambda) = \frac{\partial f_{\rm mf}(m;\lambda)}{\partial\lambda}\Big|_{m=m_\star}. \tag{12}$$

Using the envelope theorem we start by computing the arg min of the term appearing inside the min in $f(\beta;\lambda)$. Differentiating that term we obtain that the solution

$$m_\star = \arg\min_{m\in[-1,1]} \Big[-\frac{1}{2}m^2 + \lambda m - \frac{1}{\beta}H((1+m)/2)\Big]$$

satisfies the following self-consistent equation

$$m_\star = \tanh(\beta m_\star - \beta\lambda). \tag{13}$$

To obtain some visual intuition on $m_\star$ using its definition, we obtain

$$m_\star(\beta,\lambda) = \arg\max_{m\in[-1,1]} \phi_{\rm mf}(m;\beta,\lambda) = \arg\max_{m\in[-1,1]} \Big[\frac{\beta}{2}m^2 - \beta\lambda m + H\Big(\frac{1+m}{2}\Big)\Big]. \tag{14}$$

We plot the $\phi_{\rm mf}$ with respect to m (left, shifted by $\beta/2$), and $m_\star$ with respect to β (right). In the left plot we see that for high temperature ($\beta < 1$) there is one maximizer of $\phi_{\rm mf}$ around zero while in the high temperature setting there are two maximixing values which means that the arg max operation will be returning a set. This is because for $\lambda = 0$ the envelope theorem fails. For λ slightly less or greater than zero there is only one maximizer of $\phi_{\rm mf}$ less or greater than zero respectively. Hence there is a clear phase transition at $\beta = 1$ which is a property of ferromagnets.

1. $\beta < 1$ (High temperature), **there is no residual magnetization**:

$$\lim_{\lambda\to 0^+} m_\star(\beta,\lambda) = \lim_{\lambda\to 0^-} m_\star(\beta,\lambda) = 0.$$

2. $\beta > 1$ (Low temperature), **presence of residual magnetization**:

$$\lim_{\lambda \to 0^+} m_\star(\beta, \lambda) > 0, \quad \text{and,} \quad \lim_{\lambda \to 0^-} m_\star(\beta, \lambda) < 0.$$

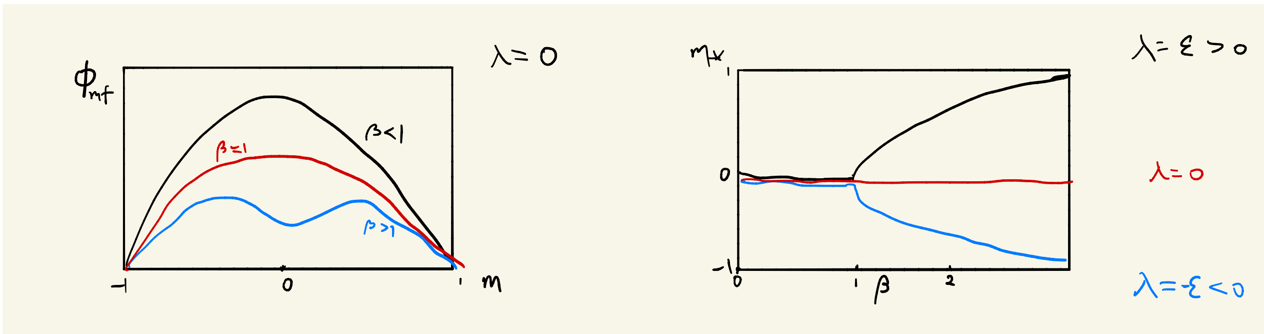

Figure 1: (left) Plot of ϕ_{mf} with respect to m. (right) Plot of $m_\star$ with respect to λ.

4 The Sherrington-Kirkpatrick model

For the Sherrington-Kirkpatrick model $\Omega = \{+1, -1\}^n$ and the Hamiltonian is given by

$$H_\lambda(\boldsymbol{\sigma}) = -\frac{1}{2} \sum_{i,j=1}^{n} W_{ij} \sigma_i \sigma_j - \frac{\lambda}{2n} \sum_{i,j=1}^{n} \sigma_i \sigma_j, \tag{15}$$

where $W_{ij} \sim_{i.i.d.} \mathcal{N}(0, 1/n)$, $1 \le i < j \le n$, $W_{ii} \sim_{i.i.d.} \mathcal{N}(0, 2/n)$ for $1 \le i \le n$, and $W_{ij} = W_{ji}$. The free entropy is given by

$$\Phi_n(\beta, \lambda) = \log \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\} \tag{16}$$

and its density by

$$\phi(\beta, \lambda) = \lim_{n \to \infty} \mathbb{E}[\Phi_n(\lambda, \beta)/n]. \tag{17}$$

The following holds (for some interesting region of β and λ)

$$\phi(\beta, \lambda) = \inf_{q,b} \phi_{\text{mf}}(q, b; \beta, \lambda), \tag{18}$$

where

$$\phi_{\text{mf}}(q, b; \beta, \lambda) = \frac{1}{4} \beta^2 (1-q)^2 - \frac{1}{2} \beta \lambda b^2 + \mathbb{E}_{G \sim \mathcal{N}(0,1)}[\log(2 \cosh(\beta(\lambda b + qG))].$$

5 Statistical decision theory

The high level correspondence between statistical physics and statistical decision theory is summarized in Table 1.

Statistical Physics	Statistical Decision theory
Configuration space Ω	Parameter space Θ
Reference measure ν_0	Bayes prior Q
Hamiltonian $H : \Omega \to \mathbb{R}$	log-Likelihood function $\log \mathbb{P}_\theta(Y)$
Gibbs-measure $\propto \exp(-\beta H(\boldsymbol{\sigma})\nu_0 d\boldsymbol{\sigma})$	Bayes posterior $\propto \mathbb{P}_{\boldsymbol{\sigma}}(Y)Q(d\boldsymbol{\sigma})$
Average configuration under Gibbs measure (finite β)	Bayes estimator
Average configuration under Gibbs measure ($\beta = \infty$)	M-estimator (MLE)
Ensemble average of an observable (finite β)	Risk of Bayes estimator
Ensemble average of an observable ($\beta = \infty$)	Risk of M-estimator
Free energy $-(1/\beta)\log\int \exp\{-\beta H(\boldsymbol{\sigma})\}\nu_0 \mathrm{d}\boldsymbol{\sigma}$	Mutual information I(observation; signal)

Table 1: Statistical physics and decision theory correspondence.