These notes have not received the scrutiny of publication. They could be missing important references, etc.

Cavity method, Belief propagation & TAP I

1 Recall and Model Definition

Recall from last time, we had a model $Y = Xw^* + \xi$ where $w^* \sim \text{Uni}\{\pm 1\}, \xi \sim \mathcal{N}(0, \sigma^2 I)$.

The posterior distribution on $w^*|Y, X$ is an Ising model.

Definition 1 (Ising model)**.** Parameters: A symmetric matrix $J \in \mathbb{R}^{n\times n}$ and $h \in \mathbb{R}^n$. A distribution μ on $x \in \{\pm 1\}^n$ is given by:

$$\mu(x) = \frac{1}{z} \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right)$$

where z is the partition function.

Remark 1. Below we assume $X = (X_1, \cdots, X_n) \sim \mu$.

2 Goal and Examples

Goal: We want to "solve" μ in terms of J and h for "nice" Js.

Example 1. We want to estimate the expectation $\mathbb{E}_{X\sim\mu}[X]$.

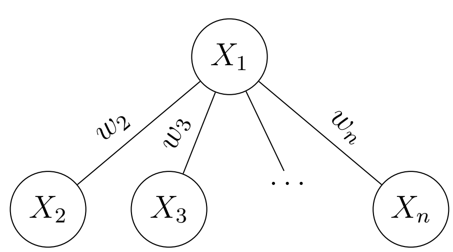
Example 2 (Sherrington-Kirkpatrick (SK) model)**.** Let M be a i.i.d. symmetric matrix where $M_{ij} \sim \mathcal{N}(0, \frac{1}{n})$. The distribution is:

$$p(x) \propto \exp(\beta\langle x, Mx\rangle + \langle h, x\rangle)$$

where $\beta > 0$ is the inverse temperature.

Remark 2. The SK model is the historical origin for some of these techniques.

Example 3 (Ising model on a star graph)**.** This is an "easier" case and represents the heart of the cavity method. The graph (shown below) has a central node X_1 connected to leaf nodes $X_2, \ldots, X_n$. The edge weights are $w_2, \ldots, w_n$.



The distribution is:

$$p(x) = \frac{1}{z} \exp\left(x_1 \sum_{j=2}^{n} w_j x_j + \langle h, x\rangle\right)$$

3 Cavity Method on a Star Graph (Example 3 in Part 2)

3.1 Cavity Method

Definition 2 (Cavity ("Hole") measure)**.** Let $x_{-1} := (x_2, \dots, x_n)$ and $h_{-1} := (h_2, \dots, h_n)$. The cavity measure p_{-1} is the distribution on x_{-1} (i.e., with node 1 removed):

$$p_{-1}(x_{-1}) = \frac{1}{z_{-1}} \exp(\langle h_{-1}, x_{-1} \rangle)$$

Remark 3. Observe that p_{-1} is a product measure.

3.2 Warm up 1: Compute z_{-1}

The partition function for the cavity measure is:

$$\begin{aligned} z_{-1} &= \sum_{x_{-1} \in \{\pm 1\}^{n-1}} \exp(\langle h_{-1}, x_{-1} \rangle) \\ &= \sum_{x_{-1} \in \{\pm 1\}^{n-1}} \prod_{j=2}^{n} \exp(h_j x_j) \\ &= \prod_{j=2}^{n} \left(e^{h_j} + e^{-h_j} \right) \\ &= 2^{n-1} \prod_{j=2}^{n} \cosh(h_j) \end{aligned}$$

where $\cosh(x) := \dfrac{e^x + e^{-x}}{2}$. (Similarly, $\sinh(x) := \dfrac{e^x - e^{-x}}{2}$ and $\tanh(x) := \dfrac{\sinh(x)}{\cosh(x)}$.)

3.3 Warm up 2: Compute $\mathbb{E}_{p_{-1}}[X_j]$

By Remark 3, for any $j \in \{2, \dots, n\}$: The marginal $p_{-1}(x_j)$ is:

$$p_{-1}(x_j) = \frac{e^{h_j x_j}}{2 \cosh(h_j)}$$

The expectation is:

$$\mathbb{E}_{p_{-1}}[X_j] = \frac{e^{h_j} \cdot (1) + e^{-h_j} \cdot (-1)}{2 \cosh(h_j)} = \frac{e^{h_j} - e^{-h_j}}{2 \cosh(h_j)} = \tanh(h_j)$$

3.4 Deriving the marginal $p(x_1)$

Lemma 1. *The marginal distribution for the central node x_1 is:*

$$\begin{aligned} p(x_1) &\propto e^{h_1 x_1} \prod_{j=2}^{n} \frac{\cosh(w_j x_1 + h_j)}{\cosh(h_j)} \\ &\propto e^{h_1 x_1} \prod_{j=2}^{n} (1 + x_1 \tanh(w_j) \tanh(h_j)) \end{aligned}$$

Proof. We can write the full distribution $p(x)$ using the cavity measure p_{-1}:

$$
\begin{aligned}
p(x) = p(x_1, x_{-1}) &= \frac{1}{z} \exp\left(h_1 x_1 + x_1 \sum_{j=2}^{n} w_j x_j + \langle h_{-1}, x_{-1} \rangle \right) \\
&= \frac{z_{-1}}{z} p_{-1}(x_{-1}) \exp\left(h_1 x_1 + x_1 \sum_{j=2}^{n} w_j x_j \right)
\end{aligned}
$$

To find the marginal $p(x_1)$, we sum over all x_{-1}:

$$
\begin{aligned}
p(x_1) &= \sum_{x_{-1}} p(x_1, x_{-1}) \\
&= \frac{z_{-1}}{z} \sum_{x_{-1}} p_{-1}(x_{-1}) \exp\left(h_1 x_1 + x_1 \sum_{j=2}^{n} w_j x_j \right) \\
&= \frac{z_{-1}}{z} \mathbb{E}_{p_{-1}} \left[\exp\left(h_1 x_1 + x_1 \sum_{j=2}^{n} w_j X_j \right) \right]
\end{aligned}
$$

Since p_{-1} is a product measure, the expectation splits:

$$
\mathbb{E}_{p_{-1}} \left[\prod_{j=2}^{n} e^{x_1 w_j X_j} \right] = \prod_{j=2}^{n} \mathbb{E}_{p_{-1}} [e^{x_1 w_j X_j}]
$$

Each term in the product is computed as:

$$
\mathbb{E}_{p_{-1}}[e^{x_1 w_j X_j}] = \sum_{x_j \in \{\pm 1\}} p_{-1}(x_j) e^{x_1 w_j x_j} = \sum_{x_j \in \{\pm 1\}} \frac{e^{h_j x_j}}{2\cosh(h_j)} e^{x_1 w_j x_j} = \frac{e^{h_j + x_1 w_j} + e^{-h_j - x_1 w_j}}{2\cosh(h_j)} = \frac{\cosh(w_j x_1 + h_j)}{\cosh(h_j)}
$$

Moreover,

$$
\cosh(w_j x_1 + h_j) = \cosh(h_j)\cosh(w_j) + x_1 \sinh(h_j)\sinh(w_j)
$$

so

$$
\frac{\cosh(h_j + x_1 w_j)}{\cosh(h_j)} = \cosh(w_j)(1 + x_1 \tanh(h_j)\tanh(w_j))
$$

Substituting this back gives the lemma. □

Remark 4. If we have $C_j, u_{j\to 1}$, s.t.

$$
1 + x_1 \tanh(w_j)\tanh(h_j) = C_j e^{u_{j\to 1} x_1}
$$

for $x_1 \in \{\pm 1\}$.

We can solve for $C_j, u_{j\to 1}$:

$$
C_j^2 = (1 + \tanh(w_j)\tanh(h_j))(1 - \tanh(w_j)\tanh(h_j))
$$

i.e.

$$
C_j = \sqrt{1 - \tanh(w_j)^2 \tanh(h_j)^2}
$$

$$
e^{2u_{j\to 1}} = \frac{1 + \tanh(w_j)\tanh(h_j)}{1 - \tanh(w_j)\tanh(h_j)}
$$

so

$$
u_{j\to 1} = \tanh^{-1}(\tanh(w_j)\tanh(h_j))
$$

Proposition 1 (Marginal for x_1)**.** The marginal $p(x_1)$ can be written in a simple form:

$$p(x_1) \propto e^{Hx_1}$$

where $H = h_1 + \sum_{j=2}^{n} u_{j\to 1}$.

From this, the expectation is simply:

$$\mathbb{E}[X_1] = \tanh(H)$$

Proof. By Lemma 1 and Remark 4,

$$p(x_1) \propto e^{h_1 x_1} \prod_{j=2}^{n} e^{u_{j\to 1} x_1} = e^{Hx_1}$$

From this,

$$z_1 = \sum_{x\in\{\pm 1\}} e^{Hx} = e^{H} + e^{-H} = 2\cosh(H)$$

$$\mathbb{E}[X_1] = (+1)\cdot\left(\frac{e^{H\cdot(+1)}}{2\cosh(H)}\right) + (-1)\cdot\left(\frac{e^{H\cdot(-1)}}{2\cosh(H)}\right) = \frac{2\sinh(H)}{2\cosh(H)} = \tanh(H)$$

□

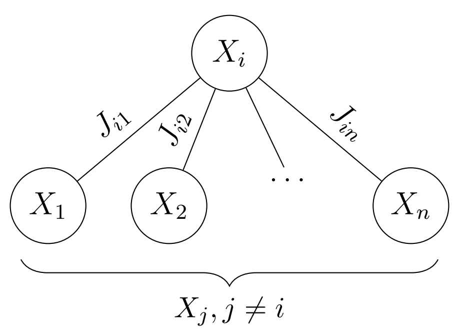
4 Cavity Method for the SK Model (Example 2 in Part 2)

Now we return to the SK model, which is on a fully connected graph.

$$p(x) \propto \exp\left(\sum_{i,j} J_{ij} x_i x_j + \sum_j h_j x_j\right)$$

where $J_{ij} \sim \mathcal{N}(0, \frac{\beta^2}{n})$.

Key Guess: Even in this dense graph, the nodes X_j (for $j \neq i$) are conditionally independent given X_i for any i.



This cannot be exactly true. In reality, there's a bunch of interactions between $X_j, j \neq i$, which are caused by all the other terms in J. However, J is a random matrix, and all of its rows are independent, so it's kind of reasonable to guess although they have some complicated relationship, for the respective calculating $\mathbb{E}[X_i]$, the interaction between $X_j, j \neq i$ actually not be very important.

Based on this guess, we expect the mean $\mathbb{E}[X_i]$ to be the form similar with the star graph:

$$\mathbb{E}[X_i] \approx \tanh\left(h_i + \sum_{j\neq i} u_{j\to i}\right)$$

where $u_{j\to i} = \tanh^{-1}(\tanh(J_{ij}) m_{j\to i})$. Here we don't use $\tanh(h_j)$. Instead, we use $m_{j\to i}$, which represents the cavity mean of X_j with i has been deleted. On the star graph, we can calculate that it is $\tanh(h_j)$, but in the SK model, it is an unknown quantity.

Observation: Since $J_{ij} \sim \mathcal{N}(0, \frac{\beta^2}{n})$, J_{ij} is small ($J_{ij}^2 = O(\frac{1}{n})$). Roughly speaking, $m_{j\to i} \approx m_j := \mathbb{E}[X_j]$.

5 Naïve Mean-Field

Now we try to accept the guess in Part 4. Basically, we want to come up with a system of equations which is called the **Naïve Mean-Field** such that their solution tells us the properties of the model.

For the n unknown means $m_1, \ldots, m_n$, we have n equations: (assuming $h_i = 0$ for simplicity)

$$\begin{cases} m_i = \tanh\left(\sum_{j \neq i} u_{j \to i}\right) \\ u_{j \to i} = \tanh(J_{ij}) m_j \end{cases}$$

In the modern day, we have hindsight to guess how to calculate the basic solutions to the equations and there's a relatively simple and clever way to prove that basically the equations do have a solution.

This works in many cases, but not in the SK model. And this is the observation behind TAP, which is the next thing we'll discuss.