Inverse Ising via Pseudolikelihood Maximization

Fangjian Guo

CCAST Stat Phy Summer School 2012

July, 2012
What do we have?

- We observe a lot of states sampled from

\[ P(\sigma) = \frac{1}{Z} \exp(\beta \sum_i h_i \sigma_i + \beta \sum_{i<j} J_{ij} \sigma_i \sigma_j), \]

just like

\[
\sigma^{(1)} = \{+1, -1, +1, \cdots, +1\} \\
\sigma^{(2)} = \{-1, -1, +1, \cdots, -1\} \\
\sigma^{(3)} = \{-1, -1, +1, \cdots, -1\} \\
\vdots \]

\[ \sigma^{(M)} = \{+1, +1, +1, \cdots, +1\} \]

Without knowing the actual structure, we can only assume a fully-connected graph.
What do we have?

This is a *Sherrington-Kirkpatrick Model*.

Moments can be extracted from the observed data:

- Average magnetization of a site: \( m_i = \langle \sigma_i \rangle \)
- Correlation between a pair of sites: \( c_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \)
Maximum likelihood estimation

Probability of the observed samples (strong assumption on independence!)

\[ P = \prod_k P(\sigma^{(k)}) \]

Rewrite and take the log,

\[ \log(P) = \beta \sum_i \sum_k \sigma_i^{(k)} + \beta \sum_{i<j} J_{ij} \sum_k \sigma_i^{(k)} \sigma_j^{(k)} - M \log Z \]

\[ l = \frac{1}{M} \log(P) = \beta \sum_i h_i m_i + \beta \sum_{i<j} J_{ij}(m_i m_j + c_{ij}) - \log Z \]

- Only the moments are found here!

\{m_i\} and \{c_{ij}\} are sufficient statistics, and this is the standard formulation.
Painful computability

\[ \log(P) = \beta \sum_i \sum_k \sigma_i^{(k)} + \beta \sum_{i<j} J_{ij} \sum_k \sigma_i^{(k)} \sigma_j^{(k)} - M \log Z \]

Partition function \( Z \) *hinders computation.*
Approximate \( Z \)?
Pseudolikelihood

(Erik Aurell and Magnus Ekeberg, *Inverse Ising Inference Using All the Data*, PRL (2012))

Can we write down a probability without $Z$?

- Aha, how about the conditional probability?

$$P(\sigma = 1|R) = \frac{P(\sigma | \sigma = 1)}{P(\sigma | \sigma = 1) + P(\sigma | \sigma = -1)},$$

and

$$P(\sigma | \sigma = 1) = \frac{1}{Z} \exp(\beta(\sum_{i \neq r} h_i \sigma_i + h_r) + \beta(\sum_{i \neq r} J_{ir} \sigma_i + \frac{1}{2} \sum_{k,l \neq r} J_{kl} \sigma_k \sigma_l))$$

- So $Z$ is gone!

$$P(\sigma | \sigma \setminus r) = \frac{1}{1 + \exp(-2\beta \sigma_r (h_r + \sum_{i \neq r} J_{ir} \sigma_i))}$$
Pseudolikelihood maximization

For the site $r$, we can maximize

$$P(\sigma_r|\sigma_{\setminus r}) = \frac{1}{1 + \exp(-2\beta \sigma_r (h_r + \sum_{i \neq r} J_{ir} \sigma_i))}$$

to estimate related parameters $h_r$ and $\{J_{ir}\}$.

- Using all the samples, the objective function is simply

$$f_r = -\frac{1}{M} \sum_k \log(P(\sigma_r^{(k)}|\sigma_{\setminus r}^{(k)}))$$

Parameters can be estimated by minimize this objective function.
How to compute?

We have $N$ objective functions,

$$\{f_1, f_2, \cdots, f_N\}$$

Then how to minimize them simultaneously?

- Minimize — *Gradient descent*
- Multi-objective — *Working in-turn*

Rewriting

$$X_r^{(k)} = \sigma_r^{(k)} \sum_{i \neq r} J_{ir}^* \sigma_i^{(k)}$$

$$P_r^{(k)} = \frac{1}{1 + \exp(-2\beta X_r^{(k)})}$$
How to compute?

Gradient (going down the hill...)

\[
\frac{\partial f_r}{\partial J_{ir}} = - \frac{1}{M} \sum_k \frac{1}{P_r^{(k)}} \frac{\partial P_r^{(k)}}{\partial J_{ir}}
\]

\[
\frac{\partial P_r^{(k)}}{\partial J_{ir}} = \frac{2 \beta \sigma_r^{(k)} \sigma_i^{(k)} \exp(-2 \beta X_r^{(k)})}{(1 + \exp(-2 \beta X_r^{(k)}))^2}
\]

- Formulating them in the form of matrix could make it better...
Experiments and results

Experimental setup

- $N = 32$
- Zero external field $h_i = 0$
- $J_{ij}$ sampled from a Gaussian $N(0, \frac{1}{pN})$
- Evaluated by reconstruction error $\Delta = \sqrt{\frac{\sum_{i<j} (J_{ij}^* - J_{ij})^2}{N(N-1)/2}}$
Effective minimization?

Approaching convergence

![Graph showing reconstruction error decreasing with iteration, indicating effective minimization.](image_url)
Coupling recovery

Real $J_{ij}$ vs. Estimated $J_{ij}$

- Real $J_{ij}$ values range from $-0.4$ to $0.6$.
- Estimated $J_{ij}$ values range from $-0.4$ to $0.6$.

The plots show the comparison between real and estimated coupling coefficients across different indices $i$ and $j$. The visual representation helps in understanding the recovery of coupling coefficients in the model.
Error vs Temperature

![Graph showing error vs temperature with a curve indicating \( \propto T^2 \).]
Error vs Sample Number

![Graph showing the relationship between reconstruction error and number of samples. The x-axis represents the number of samples, ranging from 0 to 8000, and the y-axis represents the reconstruction error, ranging from 0 to 1. The graph shows a decreasing trend as the number of samples increases.]
Thank you!