Advanced Probabilistic Machine Learning and Applications

Caterina De Bacco and Martina Contisciani

1 Tutorial 9: Bethe approximation and BP

Exercise 1: representing models using factor graphs

Write the following problems (i) in terms of a probability distribution and (ii) in terms of a graphical model by drawing an example of the corresponding factor graph.

(a) **p-spin model**

One model that is commonly studied in physics is the so-called Ising 3-spin model. The Hamiltonian of this model is written as

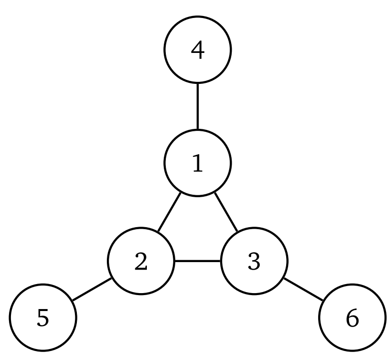
$$H(\mathbf{s}) = -\sum_{(ijk)\in E} J_{ijk} s_i s_j s_k - \sum_{i=1}^{N} h_i s_i \tag{1}$$

where E is a given set of (unordered) triplets $i \neq j \neq k$, J_{ijk} is the interaction strength for the triplet $(ijk) \in E$, and h_i is a magnetic field on spin i. The spins are Ising, which in physics means $s_i \in \{+1, -1\}$.

(b) **Independent set problem**

Independent set is a problem defined and studied in combinatorics and graph theory. Given a (unweighted, undirected) graph $G(V,E)$, an independent set $S \subseteq V$ is defined as a subset of nodes such that if $i \in S$ then for all $j \in \partial i$ we have $j \notin S$. In other words in for all $(ij) \in E$ only i or j can belong to the independent set.

For example, suppose we have the following graph:



Draw the corresponding factor graph.

(iii) Write a probability distribution that is uniform over all independent sets on a given graph.
(iv) Write a probability distribution that gives a larger weight to larger independent sets, where the size of an independent set is simply $|S|$.

[Solution]

(a,i) The Boltzmann distribution of the Ising p-spin model is:

$$\begin{aligned} P(\mathbf{s}) &= \frac{1}{Z(\beta)} \exp[-\beta H(\mathbf{s})] \\ &= \frac{1}{Z(\beta)} \prod_{(ijk)\in E} e^{\beta J_{ijk} s_i s_j s_k} \prod_{i=1}^{N} e^{\beta h_i s_i}, \end{aligned} \tag{2}$$

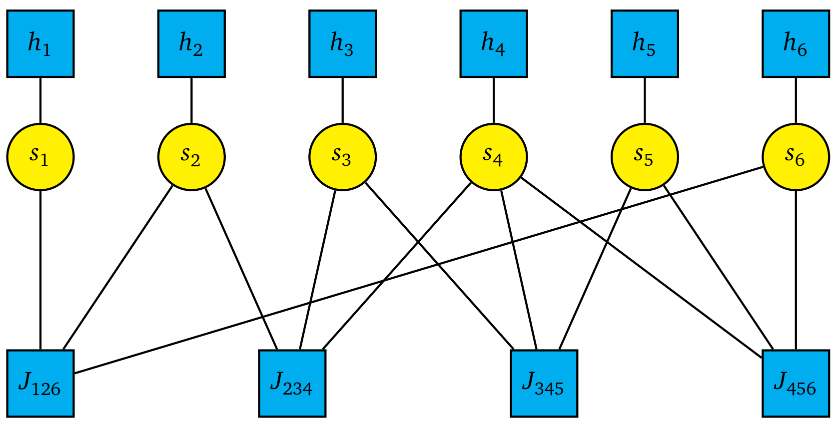
where β is the inverse temperature and:

$$Z(\beta) = \sum_{\mathbf{s}} \exp[-\beta H(\mathbf{s})],$$

is the partition function under inverse temperature β to guarantee that $P(\mathbf{s})$ sum to one for all possible configurations.

(a,ii) From eqn (2) we can see the probability distribution (without normalization) is the product of $|E|$ interaction terms and N local magnetic field terms.

For example, consider a 3-spin model with $N = 6$ spins and triplet set $E = (126), (234), (345), (456)$, the corresponding factor graph looks like:



(b,i,iii) First let's construct the factor graph for this problem. Denote $N = |V|$ as the number of nodes in the graph, we can use a length-N spin configuration σ^S to represent any set $S \subseteq V$ by:

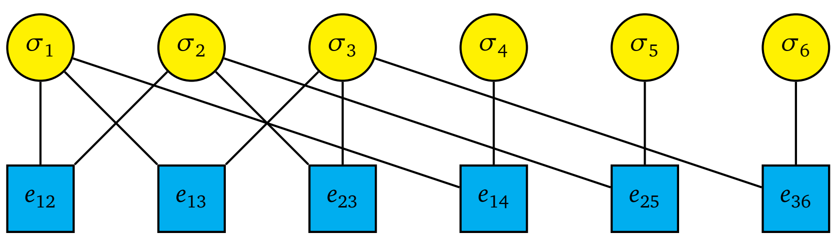
$$\sigma_i^S = \begin{cases} +1, & \text{if node } i \text{ is in } S \\ -1, & \text{if node } i \text{ is not in } S \end{cases}$$

Besides, for any edge $(ij) \in E$, we associated it with a function node e_{ij} whose compatibility function is $\psi_{ij}(\sigma_i, \sigma_j) = \mathbb{I}(\sigma_i + \sigma_j < 2)$, which equals to 0 whenever $i, j \in S$ and $(ij) \in E$.

A node set $S \subseteq V$ is an independent set if and only if $\psi_{ij}(\sigma_i^S, \sigma_j^S) = 0$ for all distinct $i, j \in S$. The probability distribution that is uniform over all independent sets

$$P(\sigma) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(\sigma_i, \sigma_j) = \frac{1}{Z} \prod_{(ij)\in E} \mathbb{I}(\sigma_i + \sigma_j < 2). \tag{3}$$

(b,ii) The corresponding factor graph looks like:



(b,iv) Note that $|S| = (N + \sum_{i=1}^{N} \sigma_i^S)/2$ (hint: write $N = n_- + n_+$, where $n_+ \equiv |S|$). If we want a probability distribution that gives a larger weight to larger independent sets, we can simply introduce a positive increasing function $g(\cdot)$ and multiply $g(|S|)$ to the probability distribution in part (a), i.e.

$$P(\sigma) = g\left(\frac{N + \sum_{i=1}^{N} \sigma_i}{2}\right) \times \frac{1}{Z} \prod_{(ij)\in E} \mathbb{I}(\sigma_i + \sigma_j < 2). \tag{4}$$

For example, we can choose $g(x) = \exp(\mu x)$ for some $\mu > 0$. The last thing to notice is that the normalizing constant Z is different in eqn (3) and eqn (4).

2 Tutorial 9: Bethe approximation and BP

Exercise 2: graph coloring problem and BP

Coloring is another classical problem of graph theory. Given a (unweighted, undirected) graph $G(V, E)$ a coloring $M \subseteq E$ is an assignment of labels, called colors, to the vertices of a graph such that no two adjacent vertices share the same color.

(a) Write a probability distribution for the coloring problem.

(b) Consider a "soft" constraint instead which relaxes the "hard" one and write the corresponding interaction function of the factor node. Hint: the soft constraint allows two neighboring nodes to have the same color, but penalized this a lot.

(c) Draw a factor graph corresponding to it.

(d) Using BP to model marginals of the coloring assignment, denote as:

- $\nu_{s_i}^{(ij)\to i}$ the messages from function node (ij) to variable node i.
- $\chi_{s_i}^{i\to(ij)}$ the message from variable node i to function node (ij).

Note that they are both functions of the state s_i of variable node i.
Write BP equations for this model.

(e) Find a fix point of these equations. Hint: what would a random guess do?
PS: recall that there might be more than one fixed point.

(f) Write the equation for the one-point marginal $P(s_i)$ and the two-point marginal $P(s_i, s_j)$ obtained from BP.

[Solution]

(a) Given graph $G(V, E)$, let's define its associated factor graph $\mathsf{FG}(\tilde{V}, \tilde{F}, \tilde{E})$, where:

- $\tilde{V} = V$ is the set of variable nodes in factor graph, the value is the color of the corresponding node.
- $\tilde{E} = \bigcup_{(ij)\in E} \{(i, ij), (j, ij)\}$ is the set of edges in factor graph.
- $\tilde{F}$ is the set of factor nodes in factor graph, which correspond to the constraint function:

$$\psi_{ij}(s_i, s_j) = \mathbb{I}(s_i \neq s_j) \quad . \tag{5}$$

The Boltzmann distribution of the factor graph is

$$P(\mathbf{s}) = \frac{1}{Z} \prod_{(ij)\in \tilde{F}} \psi_{ij}(s_i, s_j) = \frac{1}{Z} \prod_{(ij)\in E} \mathbb{I}(s_i \neq s_j) \quad ,$$

where $s_i \in \{1, 2, \ldots, q\}$ and q is the number of colors.

(b) Since the indicator constraint function is hard to deal with, usually we soften the constraint as:

$$\psi_{ij}(s_i, s_j) = \mathrm{e}^{-\beta\,\mathbb{I}(s_i=s_j)} \quad , \tag{6}$$

where we let $\beta \to \infty$.

The Boltzmann distribution of the factor graph is

$$P(\mathbf{s}) = \frac{1}{Z}\prod_{(ij)\in\tilde{F}} \psi_{ij}(s_i, s_j) = \frac{1}{Z}\prod_{(ij)\in E} \mathrm{e}^{-\beta\,\mathbb{I}(s_i=s_j)} \quad ,$$

Notice that in this factor graph, every factor node has exactly degree 2, i.e. this type of model is an example of *pair-wise model*.

(c) The factor graph is similar to that for the independent set problem as above.

(d) According to BP rule we have:

$$\begin{aligned}
\chi_{s_j}^{j\to(ij)} &= \frac{1}{\tilde{Z}^{j\to(ij)}}\prod_{k\in\partial j\setminus i} \nu_{s_j}^{(kj)\to j} &(7)\\
\nu_{s_i}^{(ij)\to i} &= \frac{1}{\tilde{Z}^{(ij)\to i}}\sum_{s_j}\psi_{ij}(s_i,s_j)\chi_{s_j}^{j\to(ij)} &(8)\\
&= \frac{1}{\tilde{Z}^{(ij)\to i}}\left[\psi_{ij}(s_i,s_i)\chi_{s_i}^{j\to(ij)} + \sum_{s_j\neq s_i}\psi_{ij}(s_i,s_j)\chi_{s_j}^{j\to(ij)}\right] &(9)\\
&= \frac{\mathrm{e}^{-\beta}\chi_{s_i}^{j\to(ij)} + \sum_{s_j\neq s_i}\chi_{s_j}^{j\to(ij)}}{\tilde{Z}^{(ij)\to i}} &(10)\\
&= \frac{1-\left(1-\mathrm{e}^{-\beta}\right)\chi_{s_i}^{j\to(ij)}}{\tilde{Z}^{(ij)\to i}} \quad , &(11)
\end{aligned}$$

where $\tilde{Z}^{j\to(ij)}$ and $\tilde{Z}^{(ij)\to i}$ are the normalization constant of $\chi_{s_j}^{j\to(ij)}$ and $\nu_{s_i}^{(ij)\to i}$ respectively. Hence, it is sufficient to only use one set of BP messages, here we choose χ's. This simplifies the equations to obtain:

$$\begin{aligned}
\chi_{s_j}^{j\to(ij)} &= \frac{1}{\tilde{Z}^{j\to(ij)}}\prod_{k\in\partial j\setminus i}\frac{1-\left(1-\mathrm{e}^{-\beta}\right)\chi_{s_j}^{k\to(kj)}}{\tilde{Z}^{(kj)\to j}} &(12)\\
&= \frac{1}{\tilde{Z}^{j\to i}}\prod_{k\in\partial j\setminus i}\left[1-\left(1-\mathrm{e}^{-\beta}\right)\chi_{s_j}^{k\to(kj)}\right] &(13)
\end{aligned}$$

Observations.

- The quantity $1-\left(1-\mathrm{e}^{-\beta}\right)\chi_{s_j}^{k\to(kj)}$ is the probability that neighbor k is fine with j taking color s_j.
- The quantity $\prod_{k\in\partial j\setminus i}\left[1-\left(1-\mathrm{e}^{-\beta}\right)\chi_{s_j}^{k\to(kj)}\right]$ is the probability that all the neighbors are fine that j taking color s_j (if i was excluded and (ij) erased).

(e) A fixed point is the uniform probability $\chi_{s_j}^{k\to(kj)} = \frac{1}{q}$. In fact, substituting inside (13) and calculating the normalization function explicitly, we get:

$$\frac{\left[1-\left(1-\mathrm{e}^{-\beta}\right)\frac{1}{q}\right]^{d_i-1}}{q\left[1-(1-\mathrm{e}^{-\beta})\frac{1}{q}\right]^{d_i-1}} = \frac{1}{q}$$

(f) The marginals are:

$$\chi_{s_j} = \frac{1}{Z^{(i)}} \prod_{k \in \partial j} \left[1 - \left(1 - \mathrm{e}^{-\beta} \right) \chi_{s_j}^{k \to (kj)} \right] \tag{14}$$

$$\chi_{s_j, s_j} = \frac{1}{Z^{(ij)}} \psi_{ij}(s_i, s_j) \chi_{s_j}^{j \to (ij)} \chi_{s_i}^{i \to (ij)} \tag{15}$$

which are valid at convergence.