Contents

1 Tutorial 6: Bethe approximation and BP

1.1 Exercise 1: graph coloring problem (part I)

> Coloring is a classical problem of graph theory. Given a (unweighted, undirected) graph $G(V, E)$, a coloring $M \subseteq V$ is an assignment of labels, called colors, to the vertices of a graph such that no two adjacent vertices share the same color.
> (i) Write a probability distribution for the coloring problem.
> (ii) 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.
> (iii) Draw a factor graph corresponding to it.

(i) 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:

$$\psi_{ij}(s_i, s_j) = \mathbb{I}(s_i \neq s_j) \quad . \tag{1}$$

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

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

$$\psi_{ij}(s_i, s_j) = \mathrm{e}^{-\beta\,\mathbb{I}(s_i = s_j)} \quad , \tag{2}$$

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

The Boltzmann distribution of the factor graph is

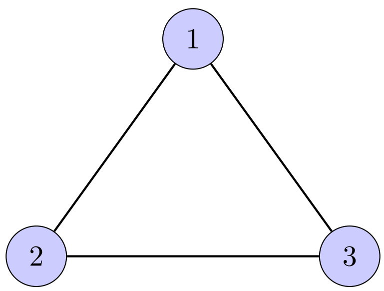
$$P(\mathbf{s}) = \frac{1}{Z} \prod_{(ij)\in\tilde{F}} \psi_{ij}(s_i, s_j) = \frac{1}{Z} \prod_{(ij)\in E} \mathrm{e}^{-\beta\,\mathbb{I}(s_i = s_j)} \quad .$$

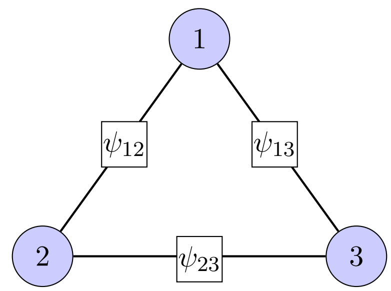
Notice that in this factor graph, every factor node has exactly degree 2, i.e. this type of model is an example of *pair-wise model*.

(iii) The factor graph is similar to that for the independent set problem below. Namely, each pair of connected nodes $i \sim j$ interacts via the compatibility function $\psi_{ij}(s_i, s_j)$. Let's take the example of a triangular graph with $q = 3$ colors (suppose $s_i \in \{\text{red, green, blue}\}$). The representation of the problem is the following.

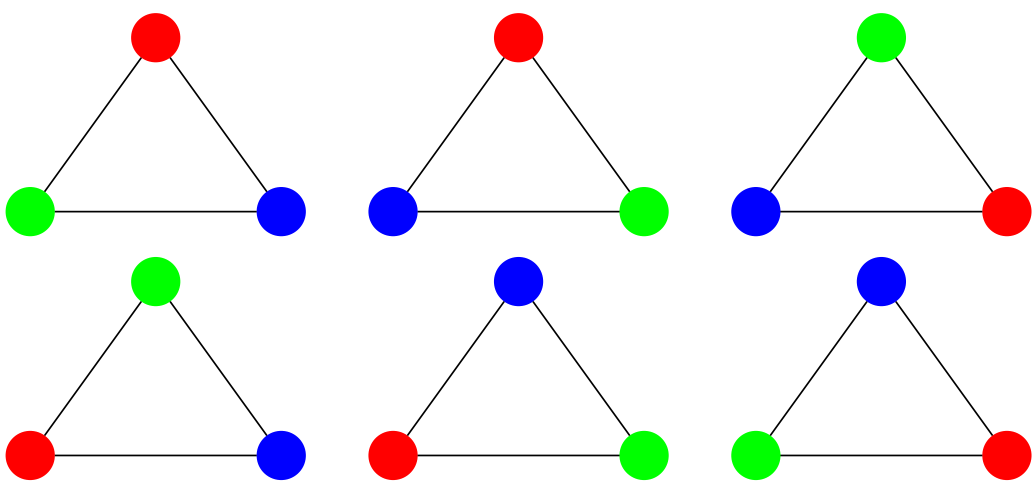
Graph: $G(V = \{1, 2, 3\}, E = \{(12), (13), (23)\})$

Factor graph: $\mathsf{FG}(\tilde{V} = \{1, 2, 3\}, \tilde{F} = \{12, 13, 23\}, \tilde{E} = \{(12), (13), (23)\})$





Admissible colorings are:



1.2 Exercise 2: representing models using factor graphs

Write **both** 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

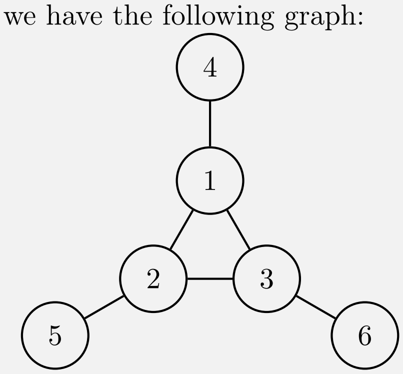
$$H(\mathbf{s}) = -\sum_{(ijk)\in E} J_{ijk} s_i s_j s_k - \sum_{i=1}^{N} h_i s_i \tag{3}$$

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:



Moreover, for problem (b):

(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 $|S|$.

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

$$\begin{aligned} P(\mathbf{s}) &= \frac{1}{Z(\beta)} \exp\left[-\beta H(\mathbf{s})\right] \\ &= \frac{1}{Z(\beta)} \prod_{(ijk)\in E} \mathrm{e}^{\beta J_{ijk} s_i s_j s_k} \prod_{i=1}^{N} \mathrm{e}^{\beta h_i s_i}, \end{aligned} \tag{4}$$

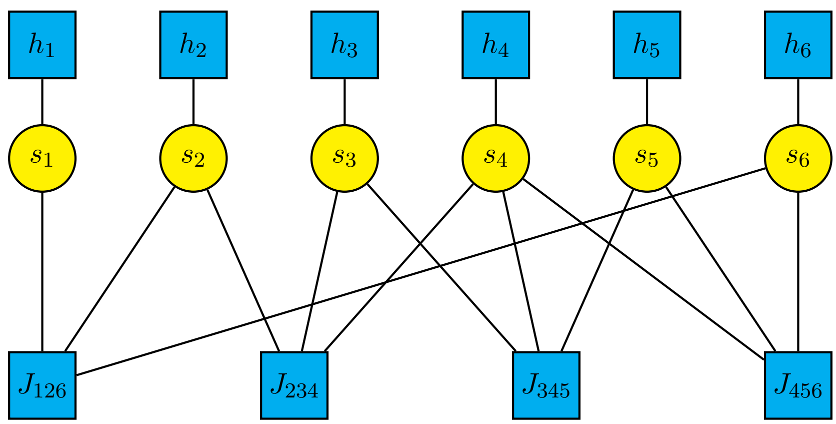
where β is the inverse temperature and:

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

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 (4) 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:

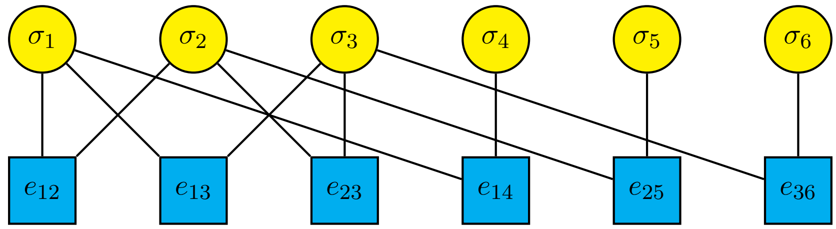
$$\sigma_i^S = \begin{cases} +1, & \text{if node } i \text{ is in } S \\ -1, & \text{if node } i \text{ is not in } S \end{cases}$$

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

$$P(\sigma) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(\sigma_i, \sigma_j) = \frac{1}{Z} \prod_{(ij)\in E} \mathbb{I}(\sigma_i + \sigma_j < 2). \tag{5}$$

(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.

$$P(\sigma) = g\left(\frac{N + \sum_{i=1}^{N} \sigma_i}{2}\right) \times \frac{1}{Z} \prod_{(ij)\in E} \mathbb{I}(\sigma_i + \sigma_j < 2). \tag{6}$$

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 (5) and eqn (6).

2 Q&A

Question 1

During the tutorial some choices of the soft distribution of Ex. 2b where discussed. Here I want to precise something that I believe could have been not explained properly

Answer 1

Our proposal is to relax $\psi(s_i, s_j) = \mathbb{I}(s_i \neq s_i)$ using the approximation $\psi(s_i, s_j) \simeq e^{-\beta \mathbb{I}(s_i \neq s_i)}$, with $\beta \gg 1$. One of you proposed an atomic measure of the form:

$$\psi(s_i, s_j) = \begin{cases} 1 & \text{if } s_i \neq s_j \\ \varepsilon & \text{if } s_i = s_j \end{cases}$$

with $\varepsilon \ll 1$. This choice is equivalently valid. In fact, as remarked by the same student who proposed the latter relaxation, since we evaluate $\psi(s_i, s_j)$ only on two points, i.e. when $s_i = s_j$ and when $s_i \neq s_j$, we are not interested in the values attained by the function on a continuous interval (not even well defined). In this sense, notice that our choice and the choice proposed by the students are practically equivalent. In detail, we can rewrite $e^{-\beta \mathbb{I}(s_i \neq s_i)}$ when $s_i = s_j$ and when $s_i \neq s_j$ as:

$$e^{-\beta \mathbb{I}(s_i \neq s_i)} = \begin{cases} 1 & \text{if } s_i \neq s_j \\ e^{-\beta} & \text{if } s_i = s_j \end{cases}$$

and rebelling $e^{-\beta} =: \varepsilon$ we show that the two choices are indeed equivalent.
We stress that our representation has a few advantages:

1. it allows for a simple generalization to continuous color palettes, where the colors takes values on a continuous interval. In this case, a choice of the form $e^{-\beta d(s_i, s_j)}$, with d being a distance evaluating the difference between two colors, may be helpful since it has some nice continuity properties.
2. Our representation is the natural choice in case we want to represent the distribution as a Boltzmann distribution.

3 Errata

- The coloring problem is defined on the nodes, thanks for pointing it out.
- I was not precise when discussing the conditions we impose on the variational distribution (see Sec 2.1 of the lecture notes):

$$Q_{\text{Bethe}}(\mathbf{s}) = \prod_{ij} b_{ij}(s_i, s_j) \prod_{i=1}^{N} b_i(s_i)^{1-d_i} \quad .$$

  In particular, if we want $b_i(\cdot)$ and $b_{a\text{``}=\text{''}ij}(\cdot)$ to be distributions, we impose $b_i(\cdot) \geq 0$, $b_a(\cdot) \geq 0, \forall i \in V, a \in F$, where F is the set of factor nodes, and the normalization conditions:

$$\begin{array}{r}\sum_{s_i} b_i\left(s_i\right)=1 \quad \forall i \in V \\ \sum_{s_{\partial a}} b_a\left(s_{\partial a}\right)=1 \quad \forall a \in F\end{array} \quad .$$

Moreover, if we want the distribution to be *locally consistent* we impose:

$$\sum_{s_{\partial a \backslash i}} b_a\left(s_{\partial a}\right)=b_i\left(s_i\right) \quad \forall a \in F, \forall i \in \partial a \quad .$$

As discussed in Obs1 of the lecture notes, in general it is possible to find a set of b_i, b_a that are not distributions but satisfy the latter condition. During the tutorial I mistakenly called the three conditions together *locally consistent conditions*, ignoring the important conceptual motivation the constraints have.