HW3. Problem 1. Part 1. Yi Li yxl121030@utdallas.edu

Problem 1. Part 1:

Since $b_i(x_i) \propto \phi_i(x_i) \cdot \prod_{k \in N(i)} m_{k \to i}(x_i)$

$\propto \phi_i(x_i) \cdot \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \cdot m_{j \to i}(x_i)$

note $m_{j \to i}(x_i) \propto \max_{x_j} (\phi_j(x_j) \cdot \psi_{ij}(x_i, x_j) \cdot \prod_{k \in N(j) \setminus i} m_{k \to j}(x_j))$;

take this into $b_i(x_i)$, then:

$b_i(x_i) \propto \phi_i(x_i) \cdot \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \cdot \max_{x_j} (\phi_j(x_j) \cdot \psi_{ij}(x_i, x_j) \cdot \prod_{k \in N(j) \setminus i} m_{k \to j}(x_j))$;

note $\phi_i(x_i) \cdot \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i)$ does not involve or depends on index j, we can reorganize as:

$b_i(x_i) \propto \max_{x_j} (\psi_{ij}(x_i, x_j) \cdot \phi_i(x_i) \cdot \phi_j(x_j) \cdot \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \cdot \prod_{k \in N(j) \setminus i} m_{k \to j}(x_j))$;

note $b_{ij}(x_i, x_j) \propto \psi_{ij}(x_i, x_j) \cdot \phi_i(x_i) \cdot \phi_j(x_j) \cdot \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \cdot \prod_{k \in N(j) \setminus i} m_{k \to j}(x_j)$;

Thus $b_i(x_i) \propto \max_{x_j} b_{ij}(x_i, x_j)$;

Thus for converged messages, $\max_{x_j} b_{ij}^*(x_i, x_j) \propto b_i^*(x_i)$;

Problem-1, Part 2:

Since $b_{ij}(x_i, x_j) = \dfrac{\psi_{ij}(x_i, x_j)\cdot\phi_i(x_i)\cdot\phi_j(x_j)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)}{\sum_{x_i x_j}\psi_{ij}(x_i, x_j)\cdot\phi_i(x_i)\cdot\phi_j(x_j)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)}$

and $b_i(x_i) = \dfrac{\phi_i(x_i)\cdot\prod_{k\in N(i)} m_{k\to i}(x_i)}{\sum_{x_i}\phi_i(x_i)\cdot\prod_{k\in N(i)} m_{k\to i}(x_i)}$

$= \dfrac{\phi_i(x_i)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot m_{j\to i}(x_i)}{\sum_{x_i}\phi_i(x_i)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot m_{j\to i}(x_i)}$

note $m_{j\to i}(x_i) = \eta_{j\to i}\sum_{x_j}\left[\phi_j(x_j)\cdot\psi_{ij}(x_i, x_j)\cdot\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)\right]$

$\eta_{j\to i}$ is some normalization constant;

take this into $b_i(x_i)$, then:

$b_i(x_i) = \dfrac{\phi_i(x_i)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot\eta_{j\to i}\sum_{x_j}\left[\phi_j(x_j)\cdot\psi_{ij}(x_i, x_j)\cdot\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)\right]}{\sum_{x_i}\phi_i(x_i)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot\eta_{j\to i}\sum_{x_j}\left[\phi_j(x_j)\cdot\psi_{ij}(x_i, x_j)\cdot\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)\right]}$

note $\phi_i(x_i)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)$ does not involve or depends on index j;

thus $b_i(x_i)$ can be reorganized and proportional to (drop the normalization factor):

$b_i(x_i) \propto \dfrac{\sum_{x_j}\psi_{ij}(x_i, x_j)\cdot\phi_i(x_i)\cdot\phi_j(x_j)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)}{\sum_{x_i}\sum_{x_j}\psi_{ij}(x_i, x_j)\cdot\phi_i(x_i)\cdot\phi_j(x_j)\cdot\prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)\cdot\prod_{k\in N(j)\setminus i} m_{k\to j}(x_j)}$

now plug in equation ①, this means $b_i(x_i) \propto \sum_{x_j} b_{ij}(x_i, x_j)$ when messages converged;

or $\sum_{x_j} b_{ij}^*(x_i, x_j) \propto b_i^*(x_i)$.

HW3 . Problem 1. Part 3 Yi Li yxl121030@utdallas.edu

Problem 1. Part 3.

For a tree, the joint probability can be facturized as:

$$p(x_1 \cdots x_n) = \frac{1}{Z} \prod_{i \in V} p_i(x_i) \cdot \prod_{(i,j)\in E} \frac{p_{ij}(x_i, x_j)}{\prod_{k \in (i,j)} p_k(x_k)}$$

For the sum-product loopy BP process, once the process converges, we could obtain the converged beliefs $b_i^*(x_i)$ and $b_{ij}^*(x_i, x_j)$, as well as the approximated partition function Z (see HW2, problem 2);

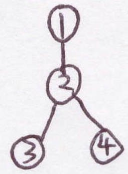
Thus, the $p(x)$ could be approximated by these converged beliefs as:

$$p(x) = \frac{1}{Z} \prod_{i \in V} b_i^*(x_i) \cdot \prod_{(i,j)\in E} \frac{b_{ij}^*(x_i, x_j)}{b_i^*(x_i) \cdot b_j^*(x_j)}$$

Therefore we could plug these $p(x)$ values into a function $f(T) = (\sum_x p(x)^{\frac{1}{T}})^T$;

Finally, we can use the matlab function limit (f(T), T, 0, 'right') which is to calculate $\lim_{T\to 0} (\sum_x p(x)^{\frac{1}{T}})^T$; and this will get the $\max_x p(x)$;

HW3. Problem 2. Yi Li yxl121030@utdallas.edu

1. use a G as example:  , its adjacent matrix is $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

assume the weight vector as [1,2,3], thus each node can take three posible colors (K=3).

Step 1: construct an initial assignment for X, which is to assign a value from {1,2,3} to each node (x_1, x_2, x_3, x_4).

For this specific coloring problem, naively, in order to avoid -zero probability assignment, we could artificially assign consecutively 1, 2, 3 to each node, which means that the inital assignment $X = [\underset{x_1}{1}, \underset{x_2}{2}, \underset{x_3}{3}, \underset{x_4}{1}]$

Step 2: perform each round of Gibbs sampling:

for each sample's specific X_i (at round t), we will calculate

$p(x_i \mid x_1^t, \cdots x_{i-1}^t, x_{i+1}^{t-1}, \cdots x_{|V|}^{t-1})$;

Thus in above example, round 1 (t=1), $x = [\underset{x_1}{1}, \underset{x_2}{2}, \underset{x_3}{3}, \underset{x_4}{1}]$

part 1: for round 2 (t=2), when we want to update x_1, we need to calculate

$p(x_1 \mid x_2^{t=1}, x_3^{t=1}, x_4^{t=1})$;

specifically, for $p(x_1 = 1 \mid x_2^{t=1}, x_3^{t=1}, x_4^{t=1}) = \dfrac{p(x_1=1, x_2=2, x_3=3, x_4=1)}{p(x_2=2, x_3=3, x_4=1)}$

$$= \frac{p(x_1=1, x_2=2, x_3=3, x_4=1)}{p(x_1=1, x_2=2, x_3=3, x_4=1) + p(x_1=2, x_2=2, x_3=3, x_4=1) + p(x_1=3, x_2=2, x_3=3, x_4=1)}$$

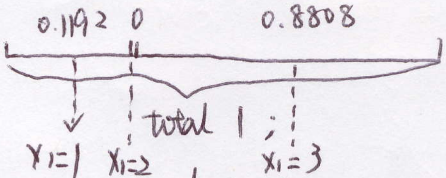
note $$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \cdot \underbrace{e^{x_1} \cdot e^{x_2} \cdot e^{x_3} \cdot e^{x_4}}_{\prod_{i \in V} \phi_i(x_i)} \cdot \underbrace{1_{x_1 \neq x_2} \cdot 1_{x_2 \neq x_3} \cdot 1_{x_2 \neq x_4}}_{\prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)}$$

for example, $p(x_1=1, x_2=2, x_3=3, x_4=1) = \frac{1}{Z} \cdot e^1 \cdot e^2 \cdot e^3 \cdot e^1 \cdot 1 \cdot 1 \cdot 1 = \frac{e^7}{Z}$;

Thus, we found that $p(x_1 = 1 \mid x_2^{t=1}, x_3^{t=1}, x_4^{t=1}) = \dfrac{e^7}{e^7 + 0 + e^9} = 0.1192$;

$p(x_1 = 2 \mid x_2^{t=1}, x_3^{t=1}, x_4^{t=1}) = \dfrac{0}{e^7 + 0 + e^9} = 0$;

$p(x_1 = 3 \mid x_2^{t=1}, x_3^{t=1}, x_4^{t=1}) = \dfrac{e^9}{e^7 + 0 + e^9} = 0.8808$;



Now we use rand function to generate a random number betwen [0,1], to say it's 0.3000

then it's falling within the bucket of $x_1=3$, thus we assign $x_1^{t=2}=3$;

part 2: now we keep updating next one, x_2, we need to calculate

$$p(x_2 \mid x_1^{t=2}, x_3^{t=1}, x_4^{t=1}) = \frac{p(x_2, x_1=3, x_3=3, x_4=1)}{p(x_1=3, x_3=3, x_4=1)}$$

using same procedures as in part 1;

we find that $p(x_2=1 \mid x_1^{t=2}, x_3^{t=1}, x_4^{t=1}) = 0$

$p(x_2=2 \mid x_1^{t=2}, x_3^{t=1}, x_4^{t=1}) = 1$

$p(x_2=3 \mid x_1^{t=2}, x_3^{t=1}, x_4^{t=1}) = 0$

we draw another random number from [0,1], to say it's 0.4000;

then we assign $x_2^{t=2}=2$;

part 3: repeat the same to X_3, X_4, to say we update $x_3^{t=2}=3$, $X_4^{t=2}=3$;

then we complete one round of sampling, and the sample is $x=[\underset{X1}{3}, \underset{X2}{2}, \underset{X3}{3}, \underset{X4}{3}]$.

Step 3: We will repeat step 2 (burnin + its) times. and will discard samples from the burnin stages since at this time, the sampling has not reached a stationary stage. finally, for samples from the its stage, for each x_i, we can count from all its samples how many times it's assigned as 1, 2, ... k; and divide those by total its, this will be the marginal probability for $P_i(x_i=k)$ from the sampling.

Matlab code notes:

(1). the C matrix is (1+ burnin + its) x N dimension; each row is a specific X assignment;

the first row will be the initial assignment;

the next burnin rows will be the burnin samples which won't be used for calculating the marginal distributions;

the last its rows will be the samples used for calculating the marginal probabilities.

(2). the Ma matrix is N x K dimension;

the element (i,j) will be the marginal probability $P(X_i=j)$;

(3). if we use burnin as 1000, its as 10,000;

the matlab produces reasonably good estimates of the actual marginals;

For example, the actual marginals for this specific graph are:

$P_1(x_1=1)=0.0973$, $P_1(x_1=2)=0.1841$, $P_1(x_1=3)=0.7187$, ← (manually calculated)

The matlab produced $\hat{P}_1(x_1=1)=0.1065$, $\hat{P}_1(X_1=2)=0.1925$, $\hat{P}_1(X_1=3)=0.7010$;

HW3, Question 3: Yi Li yxl121030@utdallas.edu

Problem 3:

$$p(x) = \frac{1}{Z}\prod_{i\in V}\exp(h_i x_i)\prod_{(i,j)\in E}\exp(J_{ij}x_i x_j).$$

for a three-node complete graph, where $h_i = 0$ and $J_{ij} = J$;

$$p(x) = \frac{1}{Z}\prod_{i\in V}\exp(0)\cdot\prod_{(i,j)\in E}\exp(J\cdot x_i x_j) \quad , \text{ since } x\in\{-1,1\}^{|V|};$$

First, $Z = \sum_x p(x) = \exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot 1\cdot 1) + \leftarrow x = [\underset{x_1}{1}, \underset{x_2}{1}, \underset{x_3}{1}]$

$\exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot 1\cdot(-1)) + \leftarrow x = [1,1,-1]$

$\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot(-1)\cdot 1) + \leftarrow x = [1,-1,1]$

$\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1)) + \leftarrow x = [1,-1,-1]$

$\exp(J\cdot(-1)\cdot 1)\cdot\exp(J\cdot(-1)\cdot 1)\cdot\exp(J\cdot 1\cdot 1) + \leftarrow x = [-1,1,1]$

$\exp(J\cdot(-1)\cdot 1)\cdot\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot 1\cdot(-1)) + \leftarrow x = [-1,1,-1]$

$\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot 1)\cdot\exp(J\cdot(-1)\cdot 1) + \leftarrow x = [-1,-1,1]$

$\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1)) \leftarrow x = [-1,-1,-1]$

$= 2e^{3J} + 6e^{-J};$

Thus, the $L(J)$ from the five samples are:

$L(J) = \left(\frac{\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot 1)\cdot\exp(J\cdot(-1)\cdot 1)}{Z}\right)\times \leftarrow \{-1,-1,1\};$

$\left(\frac{\exp(J\cdot(1)\cdot(-1))\cdot\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1))}{Z}\right)\times \leftarrow \{1,-1,-1\}$

$\left(\frac{\exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot 1\cdot 1)\cdot\exp(J\cdot 1\cdot 1)}{Z}\right)\times \leftarrow \{1,1,1\}$

$\left(\frac{\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1))}{Z}\right)\times \leftarrow \{-1,-1,-1\}$

$\left(\frac{\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot 1\cdot(-1))\cdot\exp(J\cdot(-1)\cdot(-1))}{Z}\right) \leftarrow \{1,-1,-1\}$

$$= \frac{e^{3J}}{(2e^{3J}+6e^{-J})^5}$$

$\log L(J) = 3J - 5\ln(2e^{3J} + 6e^{-J})$

when $\frac{d\log L(J)}{dJ} = 0 = 3 - \frac{5}{(2e^{3J}+6e^{-J})} \cdot (6e^{3J} - 6e^{-J})$

Thus $e^{3J} = 2e^{-J}$

$\hat{J} = \frac{1}{4}\ln 2$;

Thus the estimate for $\hat{J} = \frac{1}{4}\ln 2$ from these samples;