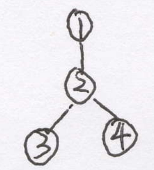
Homework-4 Yi Li yxl121030@utdallas.edu

Question 1. Part 1: using a simple graph G:



(thus $|V| = 4$)

and assume a test original weight vector $w = [1\ 2\ 3]$, thus total $K=3$;

Step 1: we can use the Gibbs sampling code from Homework-3 (gibbs_HW4.m) to generate m number of samples;

from these m samples, we can calculate the $|V| \times K = 4\times3 = 12$ empirical singleton marginals, $\breve{p}_1(x_1=1)$, $\breve{p}_1(x_1=2)$, $\breve{p}_1(x_1=3)$, $\breve{p}_2(x_2=1)$, $\breve{p}_2(x_2=2)$, $\breve{p}_2(x_2=3)$, $\breve{p}_3(x_3=1)$, $\breve{p}_3(x_3=2)$, $\breve{p}_3(x_3=3)$, $\breve{p}_4(x_4=1)$, $\breve{p}_4(x_4=2)$, $\breve{p}_4(x_4=3)$.

which can be arrayed as $|V| \times K$ matrix

$$\begin{array}{c|c} & 1\ \ 2\ \cdots\ K \\ \hline 1 & \\ 2 & \text{DATA} \\ \vdots & \\ |V| & \end{array};$$

for example, using gibbs_HW4.m, and burnin 1000, its 100000, we generated 100000 samples, and the p_empirical matrix is as:

$$\text{each vertex } V \left\{ \begin{array}{ccc} \multicolumn{3}{c}{\overbrace{\hphantom{0.0962\quad 0.1837\quad 0.7200}}^{\text{each color }(K)}} \\ 0.0962 & 0.1837 & 0.7200 \\ 0.3411 & 0.5316 & 0.1273 \\ 0.0987 & 0.1859 & 0.7154 \\ 0.0971 & 0.1857 & 0.7172 \end{array} \right. \leftarrow \breve{P}\_\text{empirical}$$

Step 2: from these 100000 samples, we try to use MLE to estimate the weight parameters Θ, which should also be a $|V| \times K$ matrix:

$\|V\| \backslash K$	1	2	3
1	$w_1(x_1=1)$	$w_2(x_1=2)$	$w_3(x_1=3)$
2	$w_4(x_2=1)$	$w_5(x_2=2)$	$w_6(x_2=3)$
3	$w_7(x_3=1)$	$w_8(x_3=2)$	$w_9(x_3=3)$
4	$w_{10}(x_4=1)$	$w_{11}(x_4=2)$	$w_{12}(x_4=3)$

However, in this context we did assume that $w_1(x_1=1) = w_4(x_2=1) = w_7(x_3=1) = w_{10}(x_4=1)$, and $w_2(x_1=2) = w_5(x_2=2) = w_8(x_3=2) = w_{11}(x_4=2)$, and $w_3(x_1=3) = w_6(x_2=3) = w_9(x_3=3) = w_{12}(x_4=3)$, Thus, we only need to generate 3 weight parameters;

Step 3: The gradient ascend method is as: for each clique (in this case, each node)

$$\phi_i(X_i)\ (t{+}1\ \text{step}) \leftarrow \phi_i(X_i)\ (t\ \text{step}) + \eta \cdot \left(\frac{\hat{P}\_\text{empirical}_{,i}(X_i) - P\_\text{inference}_{,i}(X_i)}{\phi_i(X_i)\ (t\ \text{step})} \right)$$

we choose an initial guess for the W vector as [1 1 1] arbitrarily.

Thus, for the samples, the $\phi_i(X_i)$ will be a $|V| \times K$ matrix:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ 1 & \phi_1(X_1=1) & \phi_1(X_1=2) & \phi_1(X_1=3) \\ \vdots & \phi_2(X_2=1) & \phi_2(X_2=2) & \phi_2(X_2=3) \\ \vdots & \phi_3(X_3=1) & \phi_3(X_3=2) & \phi_3(X_3=3) \\ |V| & \phi_4(X_4=1) & \phi_4(X_4=2) & \phi_4(X_4=3) \end{array}$$

, here $\phi_i(X_i) = e^{W(X_i)}$ by definition.

Thus for [1, 1, 1] weight, the $\phi_i(X_i)$ matrix is:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ 1 & e & e & e \\ \vdots & e & e & e \\ \vdots & e & e & e \\ |V| & e & e & e \end{array}$$

Step 4: since $W = [1, 1, 1]$, we then can use the sumprod_HW4.m from Homework 2 to calculate the p_inference,i (X_i) for each singleton marginal $P_i(X_i)$ (or $b_i(X_i)$);

this matrix here is:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ 1 & 0.3333 & 0.3333 & 0.3333 \\ \vdots & 0.3333 & 0.3333 & 0.3333 \\ \vdots & 0.3333 & 0.3333 & 0.3333 \\ |V| & 0.3333 & 0.3333 & 0.3333 \end{array} \leftarrow P\_\text{inference}$$

and we choose η arbitrarily $= 0.1$; so the next step $\phi_i(X_i)$ matrix will equal to:

$$\begin{bmatrix} e & e & e \\ e & e & e \\ e & e & e \\ e & e & e \end{bmatrix} + 0.1 \cdot \left(\frac{\begin{bmatrix} 0.0962 & 0.1837 & 0.7200 \\ 0.3411 & 0.5316 & 0.1273 \\ 0.0987 & 0.1859 & 0.7154 \\ 0.0971 & 0.1857 & 0.7172 \end{bmatrix} - \begin{bmatrix} 0.3333 & 0.3333 & 0.3333 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.3333 & 0.3333 & 0.3333 \end{bmatrix}}{\begin{bmatrix} e & e & e \\ e & e & e \\ e & e & e \\ e & e & e \end{bmatrix}} \right)$$

note*: this division is dot division!

which is: $\begin{bmatrix} 2.7096 & 2.7128 & 2.7325 \\ 2.7186 & 2.7256 & 2.7107 \\ 2.7097 & 2.7129 & 2.7324 \\ 2.7096 & 2.7129 & 2.7324 \end{bmatrix}$

since we assume $W_i(x_i = k)$ is the same for all i; we average up each column and the exponentials of the weight vectors are:

$$[2.7119,\ 2.7160,\ 2.7270],$$

now take log, and we get the updated weight vector as:

$$[\ln 2.7119,\ \ln 2.7160,\ \ln 2.7270] = [0.9976,\ 0.9916,\ 1.0032];$$

This concludes one round of the calculation;

Step 5: We repeated the above steps 1-4 10,000 times;

and the final weight vector is $[-0.5344, 0.4575, 1.4813]$;

when we use sumprod_HW4.m and this weight vector to calculate the p_inference,

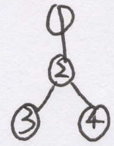
it's:

$$\begin{array}{cc} & K \\ |V| & \left\{ \begin{matrix} 0.0960 & 0.1797 & 0.7243 \\ 0.3425 & 0.5352 & 0.1223 \\ 0.0960 & 0.1797 & 0.7243 \\ 0.0960 & 0.1797 & 0.7243 \end{matrix} \right. \end{array} \leftarrow \text{p\_inference}.$$

we can see that this matches very well with the observed $\hat{p}$_empirical from Step 1;

The codes are colormle.m and colormle_test.m.

Question 1. Part 2: using a simple graph G:



(thus $|V| = 4$),

and assume the original weight vector $\omega = [1, 2, 3]$, thus total $K=3$.

Step 1: Expectation step:

for illustration purpose, suppose we only generate 5 samples using the gibbs-HW4.m, as:

		x1	x2	x3	x4
m=5 samples	1	3	2	3	3
	2	3	1	3	3
	3	3	2	3	3
	4	3	2	3	3
	5	3	2	3	3

, assume x2 is the latent / missing data;

we initially guess the weight as [1, 1, 1], so for sample 1:

x1	x2	x3	x4
3	?	3	3

we need to calculate the probabilities for $P_x(X_1=3, X_2=1, X_3=3, X_4=3)$, and
$P_x(X_1=3, X_2=2, X_3=3, X_4=3)$, and
$P_x(X_1=3, X_2=3, X_3=3, X_4=3)$.
in condition to $P(X_1=3, X_3=3, X_4=3)$

So the $P(X_1=3, X_2=1, X_3=3, X_4=3 \mid X_1=3, X_3=3, X_4=3) = 0.5$;
$P(X_1=3, X_2=2, X_3=3, X_4=3 \mid X_1=3, X_3=3, X_4=3) = 0.5$;
$P(X_1=3, X_2=3, X_3=3, X_4=3 \mid X_1=3, X_3=3, X_4=3) = 0$;

Thus we split sample 1 into 3 new samples, as:

x1	x2	x3	x4	obeserved numbers*
3	1	3	3	0.5
3	2	3	3	0.5
3	3	3	3	0

We repeat this with all m=5 samples, and the new sample is as:

	x1	x2	x3	x4	observed numbers/frequencies *
from sample 1	3	1	3	3	0.5
	3	2	3	3	0.5
	3	3	3	3	0
from sample 2	3	1	3	3	0.5
	3	2	3	3	0.5
	3	3	3	3	0
from sample 3	3	1	3	3	0.5
	3	2	3	3	0.5
	3	3	3	3	0
from sample 4	3	1	3	3	0.5
	3	2	3	3	0.5
	3	3	3	3	0
from sample 5	3	1	3	3	0.5
	3	2	3	3	0.5
	3	3	3	3	0

now we can calculate the empirical p matrix,

as an example, $\check{p}_2(x_2=1) = \dfrac{(0.5+0.5+0.5+0.5+0.5)}{5} = 0.5$;

and the $\check{p}$_empirical matrix is:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ \hline 1 & 0 & 0 & 1 \\ \vdots & 0.5 & 0.5 & 0 \\ \vdots & 0 & 0 & 1 \\ |V| & 0 & 0 & 1 \end{array}$$

Step 2: Once we have the $\check{p}$_empirical, we can use the ulormle.m method to perform the MLE maximization, and generate a new weight vector w^*;

Step 3: We then go back to Step 1, using w^* as the estimated weight; and we repeat this process;

For this specific graph, we generated 10,000 samples, using weight [1,2,3] (burnin, 1000); then we performed 100 cycles of the EM methods;

The final weight vector is [-0.5646, 0.5327, 1.4712];

the corresponding p-inference matrix using sumprod_HW4.m is:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ \hline 1 & 0.0952 & 0.1940 & 0.7108 \\ \vdots & 0.3320 & 0.5338 & 0.1342 \\ \vdots & 0.0952 & 0.1940 & 0.7108 \\ |V| & 0.0952 & 0.1940 & 0.7108 \end{array} \quad \longleftarrow \text{p\_inference.}$$

which matches well with the original p-empirical as:

$$\begin{array}{c|ccc} & 1 & \cdots & K \\ \hline 1 & 0.0978 & 0.1903 & 0.7119 \\ \vdots & 0.3200 & 0.5418 & 0.1382 \\ \vdots & 0.0997 & 0.1883 & 0.7120 \\ |V| & 0.1001 & 0.1865 & 0.7134 \end{array} \quad \longleftarrow \text{p\_empirical.}$$

The codes are colorem.m and colorem_test.m.