PGM-Assignment #4

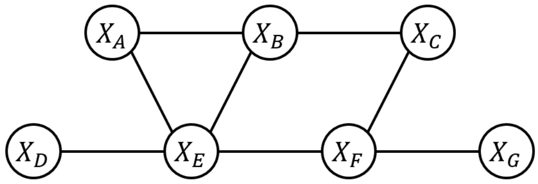
1. The differences between Bayesian Networks (BNs, denoted by G) and Markov Networks (MNs, denoted by H):

(a) Please prove that no MN can PERFECTLY* represent a v-structure in BN;

(b) Please prove that no BN can PERFECTLY represent a polygon (use quadrangle** as the example) in MN;

Hint:
* : PERFECTLY: refer to Perfect-Map: $I(H) = I(G)$
** : quadrangle: 四边形

2. For a Markov network as below (all the variables are 0/1 binary):



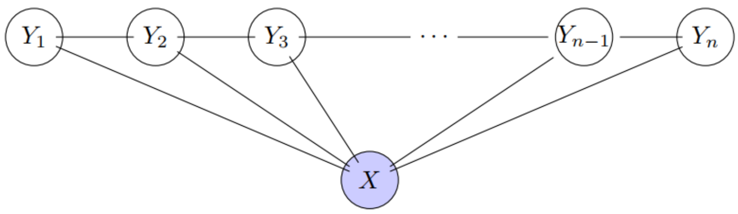
Q: Please write down the Gibbs distribution of this MN in **log-linear format** based on the constructive solution by Julian Beseg (1974).

(**Hint:** for each term, the coefficient has different values for the different configurations of the associated variables.)

3. In a simplified linear-chain like Markov network as below, all the variables are 0-1 binary

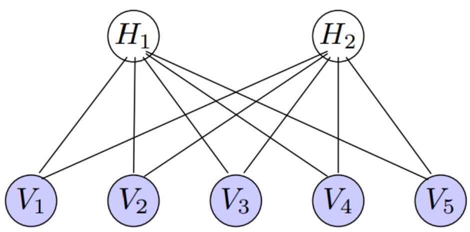
(1) Please write down its Gibbs distribution.

(2) If all X is always observed, please write down its conditional distribution (see the hint below).



Hints for above conditional models: because X is always observed, you do not need to model the marginal distribution over X.

4. An RBM (Resctricted Boltzmann Machine) is a bipartite Markov network consisting of a visible (observed) layer and a hidden layer, where each node is a **binary** random variable. Consider the following RBM:



The joint distirbution of a configuation is given by:

$$P(\boldsymbol{H}=\boldsymbol{h},\boldsymbol{V}=\boldsymbol{v})=\frac{1}{Z}e^{-E(\boldsymbol{h},\boldsymbol{v})},\quad \boldsymbol{h}=(h_1,h_2)^T,\quad \boldsymbol{v}=(v_1,...,v_5)^T$$

And:

$$E(\boldsymbol{h},\boldsymbol{v})=-\sum_{i=1}^{2}a_ih_i-\sum_{j=1}^{5}b_jv_j-\sum_{ij}w_{ij}h_iv_j$$

(1) Using the above joint distirbution form, **show that $p(\boldsymbol{H}|\boldsymbol{V})$ could be factorized as:**

$$p(\boldsymbol{H}|\boldsymbol{V})=\prod_{i=1}^{2}p(H_i|\boldsymbol{V})$$

And also:

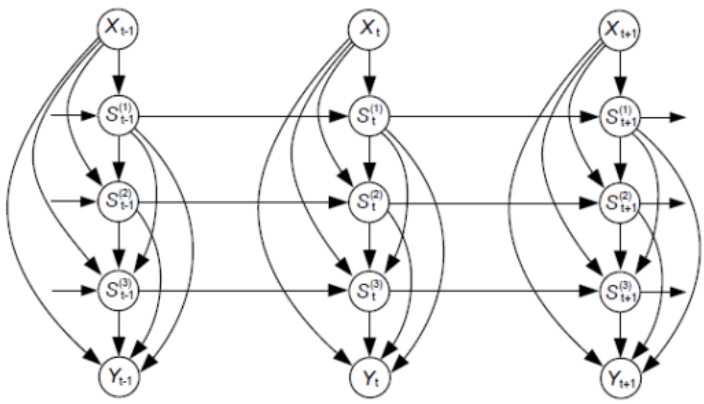
$$p(\boldsymbol{V}|\boldsymbol{H})=\prod_{j=1}^{5}p(V_j|\boldsymbol{H})$$

(2) Can the marginal distribution over the hidden variables $p(\boldsymbol{H})$ be factorized? Please give your reason according to the graph.

(Optional) Try derive $p(\boldsymbol{H})$ to confirm your judgement.

(3) (Optinal) Now let $\boldsymbol{V}$ denotes the evaluations of users on some songs. $V_j=1/0$ means the user likes/dislikes this song. Suppose V_1: Hero(Mariah Carey song), V_2: Perfect(Ed Sheeran song), V_3: IDOL(BTS song), V_4: Horner: For The Love Of A Princess, V_5: Kung Fu Piano: Cello Ascends. After training, suppose $H_1=1$ corresponds to the pop music, and $H_2=1$ corresponds to the classical music. **Which w_{ij} do you expect to be positive?**

5. Consider a tree-structured HMM (simply treat it as a Bayesian network with a variable index for time!):



(1) Please write down the factorization of the joint distribution P according to the figure.

(2) (Optional) Suppose all variables are discrete random variables. $|val(S)|=k,|val(X)|=r,|val(Y)|=l$. (The number of possible values of all S is k, all X is r, and all Y is l.) This model could be parameterized by:

a) Initial probability vector $\boldsymbol{v}\in\mathbb{R}^r$: $\boldsymbol{v}_x=P(X=x)$

b) State-state transmission tensor $\boldsymbol{T}^{(1)}\in\mathbb{R}^{r\times k\times k}$, $\boldsymbol{T}^{(2)}\in\mathbb{R}^{r\times k\times k\times k}$, $\boldsymbol{T}^{(3)}\in\mathbb{R}^{r\times k\times k\times k\times k}$:

$$\boldsymbol{T}^{(1)}_{x,i_1,i}=P(S^{(1)}_t=i|X_t=x,S^{(1)}_{t-1}=i_1)$$
$$\boldsymbol{T}^{(2)}_{x,i_1,i_2,i}=P(S^{(2)}_t=i|X_t=x,S^{(1)}_t=i_1,S^{(2)}_{t-1}=i_2)$$
$$\boldsymbol{T}^{(3)}_{x,i_1,i_2,i_3,i}=P(S^{(3)}_t=i|X_t=x,S^{(1)}_t=i_1,S^{(2)}_t=i_2,S^{(3)}_{t-1}=i_3)$$

c) State-observation emission tensor $\boldsymbol{O} \in \mathbb{R}^{r\times k\times k\times k\times l}$: $\boldsymbol{O}_{x,i_1,i_2,i_3,y} = P(Y=y|X=x, S^{(m)}=i_m, m=1,2,3)$

Q: Please derive the form of $p(S_t^{(2)}|S^{(1)}, S_{-t}^{(2)}, S^{(3)}, X, Y)$ in terms of above parameters, where:

$$S^{(i)} = (..., S_{t-1}^{(i)}, S_t^{(i)}, S_{t+1}^{(i)}, ...), i = 1, 3$$
$$S_{-t}^{(2)} = (..., S_{t-1}^{(2)}, S_{t+1}^{(2)}, ...)$$
$$X = (..., X_{t-1}, X_t, X_{t+1}, ...)$$
$$Y = (..., Y_{t-1}, Y_t, Y_{t+1}, ...)$$