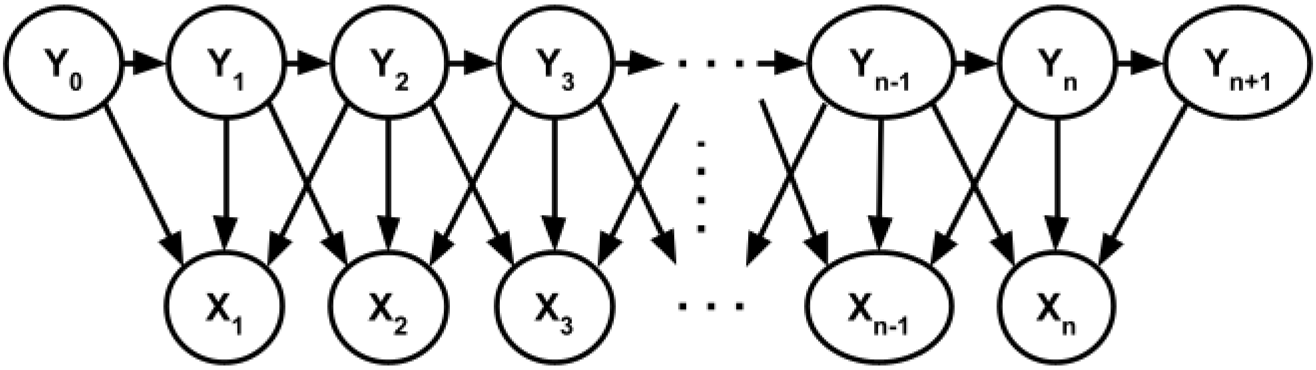
Probabilistic Graphical Models Final Exam - Spring 1399 (2020) Khordad 1399 - July 2020	Instructor: B. Nasihatkon	دانشگاه صنعتی خواجه نصیرالدین طوسی K. N. TOOSI UNIVERSITY OF TECHNOLOGY

Question 1 - Active Trails

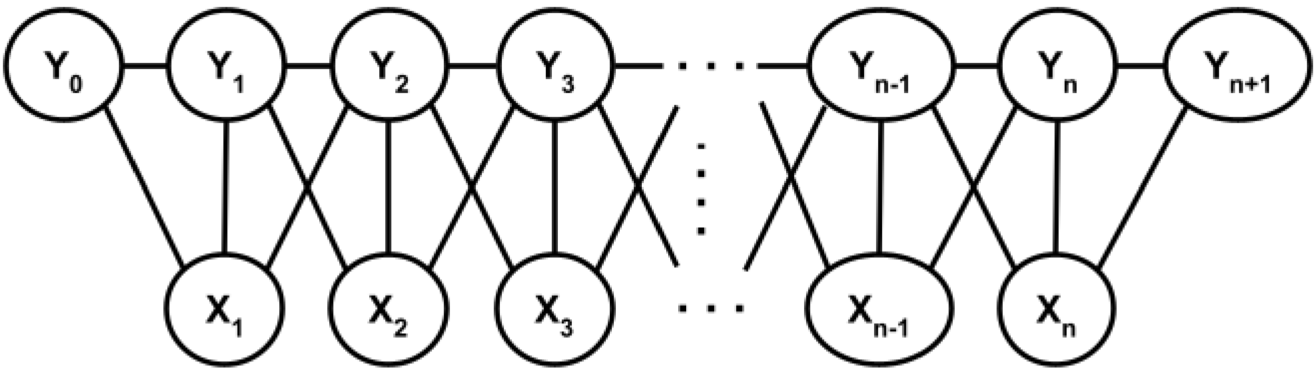
Consider the Bayes Network below.



A) Which of the following statements are True, and which are False (in general). Next to each statement, write the word "True" or an Active Trail rejecting the statement.

	True/Active Trail		True/Active Trail
$Y_0 \perp Y_3$	Y0,Y1,Y2,Y3	$X_1 \perp X_3 \mid X_2, Y_2$	**X_1,Y_1,X_2,Y_3,X_3**
$Y_0 \perp Y_3 \mid Y_2$	True	$Y_0 \perp Y_6 \mid Y_3, X_3$	Y0,Y1,Y2,X3,Y4,Y5,Y6
$Y_0 \perp Y_3 \mid Y_2, X_2$	Y0,Y1,X2,Y3	$Y_0 \perp Y_6 \mid Y_3, X_4$	True

B) Repeat part A for the Markov Network below

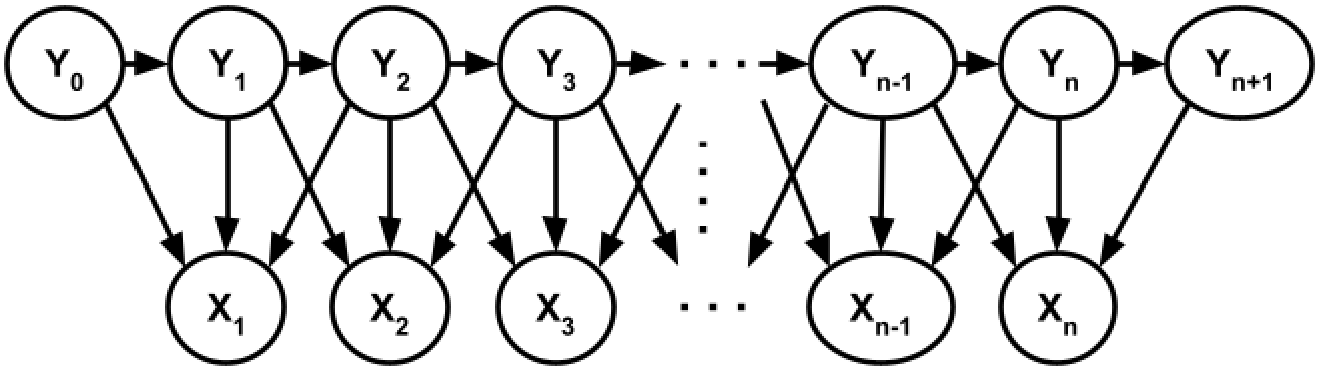


	True/Active Trail		True/Active Trail
$Y_0 \perp Y_3$	Y0,Y1,Y2,Y3	$X_1 \perp X_3 \mid X_2, Y_2$	True
$Y_0 \perp Y_3 \mid Y_2$	Y0,Y1,X2,Y3	$Y_0 \perp Y_6 \mid Y_3, X_3$	True

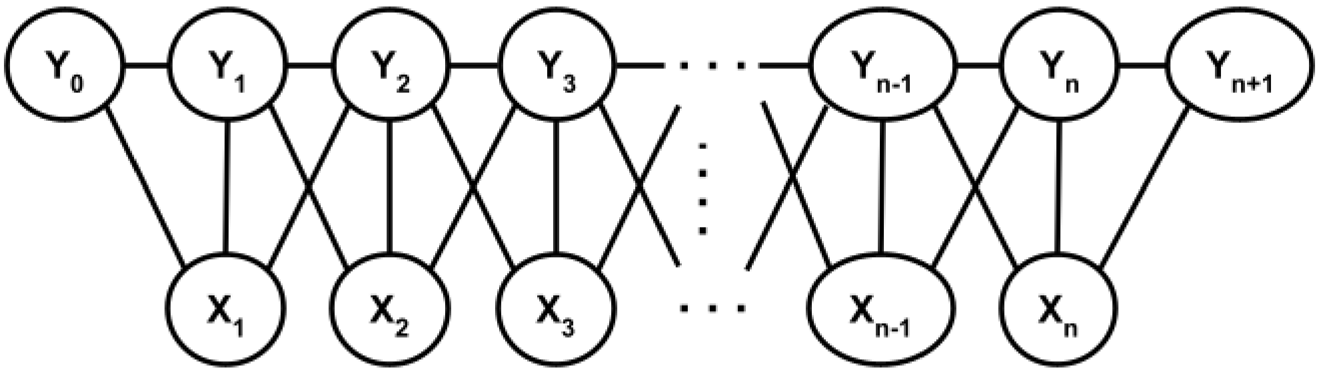
$Y_0 \perp Y_3 \mid Y_2, X_2$	True	$Y_0 \perp Y_6 \mid Y_3, X_4$	Y0,Y1,Y2,X3,Y4,Y5,Y6

Question 2 - Induced Markov Networks - Junction tree

Consider the Bayes Network below.

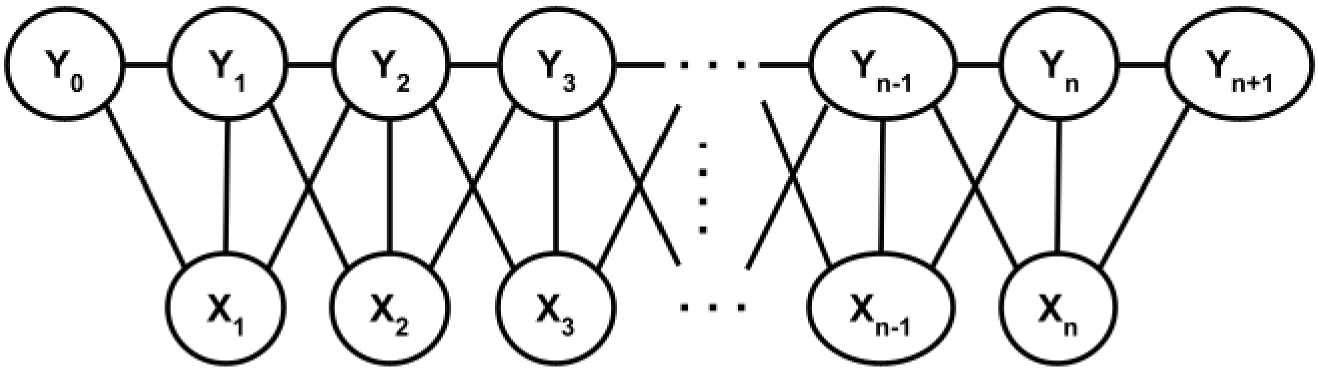


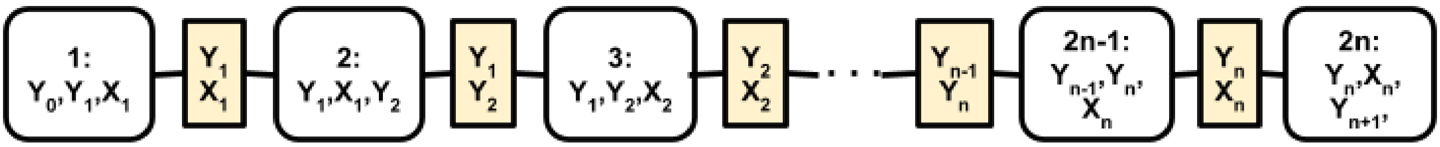
A) Draw the corresponding induced Markov Network.
B) Using part A, build a junction-tree (clique-tree) in which each cluster contains at most 4 variables. Do not forget to determine the sepsets.
C) Build a junction-tree for the Markov Network below for which each cluster has at most 3 variables. Do not forget to specify the sepsets.



Question 3 - Message Passing - Junction Tree

Consider the Markov Network below and its corresponding Junction-tree.





Assume that all variables are binary $Y_0, Y_1, ..., Y_{n+1}, X_1, X_2, ..., X_n \in \{0, 1\}$,and the joint distribution is defined as

$$P(Y_0, Y_1, ..., Y_{n+1}, X_1, X_2, ..., X_n) = \frac{1}{Z}\prod_{t=1}^{n} \phi_t(Y_{t-1}, Y_t, X_t)\ \ \sigma_t(Y_t, X_t, Y_{t+1}),$$

where
$\phi_t(Y_{t-1}, Y_t, X_t) = \exp(\, 1(Y_{t-1} = Y_t \neq X_t)\,)$, and
$\sigma_t(Y_t, X_t, Y_{t+1}) = \exp(\, Y_t\ X_t Y_{t+1})$.

We intend to perform max-sum message passing.

A) Derive the max-sum messages $\delta_{1\to 2}(Y_1, X_1)$ and $\delta_{2\to 3}(Y_1, Y_2)$

$$\delta_{1\to 2}(Y_1, X_1) = \max_{Y_0}\ 1(Y_0 = Y_1 \neq X_1) = \max\,(\,1(0 = Y_1 \neq X_1),\ 1(1 = Y_1 \neq X_1)\,) = 1(Y_1 \neq X_1)$$

$$\delta_{2\to 3}(Y_1, Y_2) = \max_{X_1}\ 1(Y_1 \neq X_1) + Y_1\ X_1 Y_2 = \max\ \ (1(Y_1 = 1)\,, 1(Y_1 = 0) + Y_1\ Y_2)$$

$$= \max\ (Y_1\,, 1 - Y_1 + Y_1\ Y_2) = 1$$

B) Derive $\delta_{2n\to 2n-1}(Y_n, X_n)$ and $\delta_{2n-1\to 2n-2}(Y_{n-1}, Y_n)$

$$\delta_{2n\to 2n-1}(Y_n, X_n) = \max_{Y_{n+1}}\ Y_n\ X_n Y_{n+1} = Y_n\ X_n$$

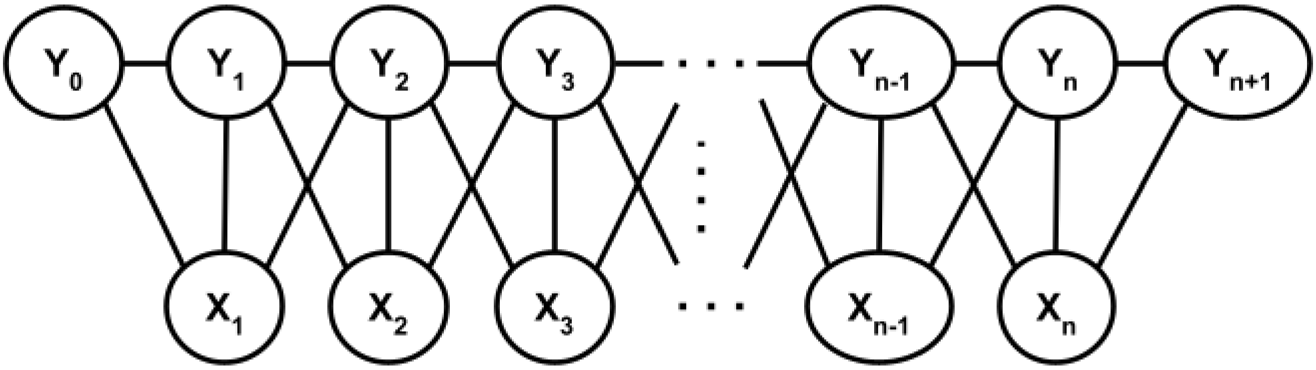
$$\delta_{2n-1\to 2n-2}(Y_{n-1}, Y_n) = \max_{X_n}\ 1(Y_{n-1} = Y_n \neq X_n) + Y_n\ X_n$$

$$= \max\ (1(Y_{n-1} = Y_n = 1), 1(Y_{n-1} = Y_n = 0) + Y_n)$$

$$= \max\ (Y_{n-1}Y_n, (1 - Y_{n-1})(1 - Y_n) + Y_n) = Y_{n-1}Y_n + 1 - Y_{n-1}$$

Question 4 - Gibbs sampling

Consider the Markov Network below,



with all binary random variables $Y_0, Y_1, ..., Y_{n+1}, X_1, X_2, ..., X_n \in \{0, 1\}$,and the joint distribution

$$P(Y_0, Y_1, ..., Y_{n+1}, X_1, X_2, ..., X_n) = \frac{1}{Z}\prod_{i=1}^{n} \phi_i(Y_{i-1}, Y_i, X_t)\ \ \sigma_i(Y_i, X_i, Y_{i+1}),$$

where

$\phi_i(Y_{i-1}, Y_i, X_i) = \exp(\,1(Y_{i-1} = Y_i \neq X_i)\,)$, and

$\sigma_i(Y_i, X_i, Y_{i+1}) = \exp(\,Y_i\ X_i\ Y_{i+1})$.

We need to use Gibbs sampling to draw samples from the above joint distribution. To perform a single transition

$Y_0^{\ t}, Y_1^{\ t}, ..., Y_{n+1}^{\ t}, X_1^{\ t}, X_2^{\ t}, ..., X_n^{\ t} \rightarrow Y_0^{\ t+1}, Y_1^{\ t+1}, ..., Y_{n+1}^{\ t+1}, X_1^{\ t+1}, X_2^{\ t+1}, ..., X_n^{\ t+1}$

of the Gibbs sampling algorithm, each $Y_i^{\ t+1}$ is sampled from the distribution $Q_i(Y_i)$ and each $X_i^{\ t+1}$is sampled from the distribution $R_i(X_i)$.

A) Derive $Q_0(Y_0)$ and $Q_{n+1}(Y_{n+1})$.

We assume that samples are taken in the following order:

$Y_0\ , Y_1\ , ..., Y_{n+1}\ , X_1\ , X_2\ , ..., X_n$

then we have

$$Q_0(Y_0) = P(Y_0 | \, Y_1^{\ t}, ..., Y_{n+1}^{\ t}, X_1^{\ t}, ..., X_n^{\ t})$$

$$= \exp(1(Y_0 = Y_1^{\ t} \neq X_1^{\ t})\,)\,/\,\sum_{Y_0}\exp(1(Y_0 = Y_1 \neq X_1)\,)$$

$$= \exp(1(Y_0 = Y_1 \neq X_1))\,/\,(\exp(1(0 = Y_1 \neq X_1)) + \exp(1(1 = Y_1 \neq X_1))$$

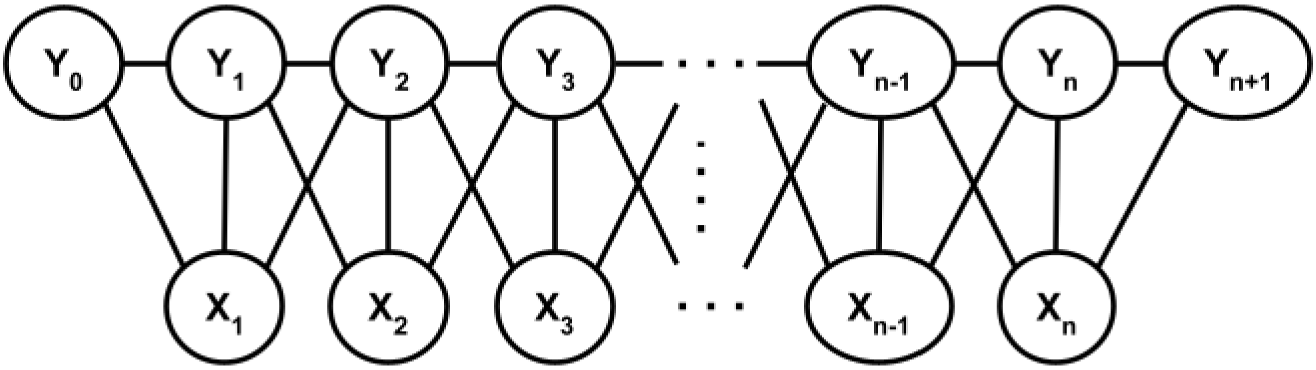
B) Derive $Q_i(Y_i)$ for $i \in \{1, 2,..., n\}$.

C) Derive $R_i(X_i)$.

- In all cases, **simplify** your answer as much as possible.

Question 3 - MRF Parameter Estimation

Consider the Markov Network below,



with all binary random variables $Y_0, Y_1, \ldots, Y_{n+1}, X_1, X_2, \ldots, X_n \in \{0, 1\}$, and the joint distribution

$$P(Y_0, Y_1, \ldots, Y_{n+1}, X_1, X_2, \ldots, X_n) = \frac{1}{Z(u,w)} \prod_{i=1}^{n} \phi_i(Y_{i-1}, Y_i, X_t) \;\; \sigma_i(Y_i, X_i, Y_{i+1}),$$

where

$\phi_i(Y_{i-1}, Y_i, X_i) = \exp(w\, 1(Y_{i-1} = Y_i \neq X_i))$, and

$\sigma_i(Y_i, X_i, Y_{i+1}) = \exp(\, Y_t \, X_t \, Y_{t+1})$.

The training data is

$$Y_0^1 = 0, Y_1^1 = 0, \ldots, Y_{n+1}^1 = 0, X_1^1 = 1, X_2^1 = 1, \ldots, X_n^1 = 1$$
$$Y_0^2 = 1, Y_1^2 = 1, \ldots, Y_{n+1}^2 = 1, X_1^2 = 0, X_2^2 = 0, \ldots, X_n^2 = 0$$
$$Y_0^3 = 0, Y_1^3 = 0, \ldots, Y_{n+1}^3 = 0, X_1^3 = 1, X_2^3 = 1, \ldots, X_n^3 = 1$$
$$Y_0^4 = 0, Y_1^4 = 0, \ldots, Y_{n+1}^4 = 0, X_1^4 = 0, X_2^4 = 0, \ldots, X_n^4 = 0$$

A) Derive the log-likelihood function as a function of w and u. Simplify as much as you can.
B) Compute the log-likelihood function at $w = u = 0$.
C) Repeat parts A,B for the following CRF (compute conditional log-likelihood instead of log-likelihood)

$$P(Y_0, Y_1, \ldots, Y_{n+1} | X_1, X_2, \ldots, X_n) = \frac{1}{Z(u,w,X_1,X_2,\ldots,X_n)} \prod_{i=1}^{n} \phi_i(Y_{i-1}, Y_i, X_t) \;\; \sigma_i(Y_i, X_i, Y_{i+1}),$$