Name: Haoze He

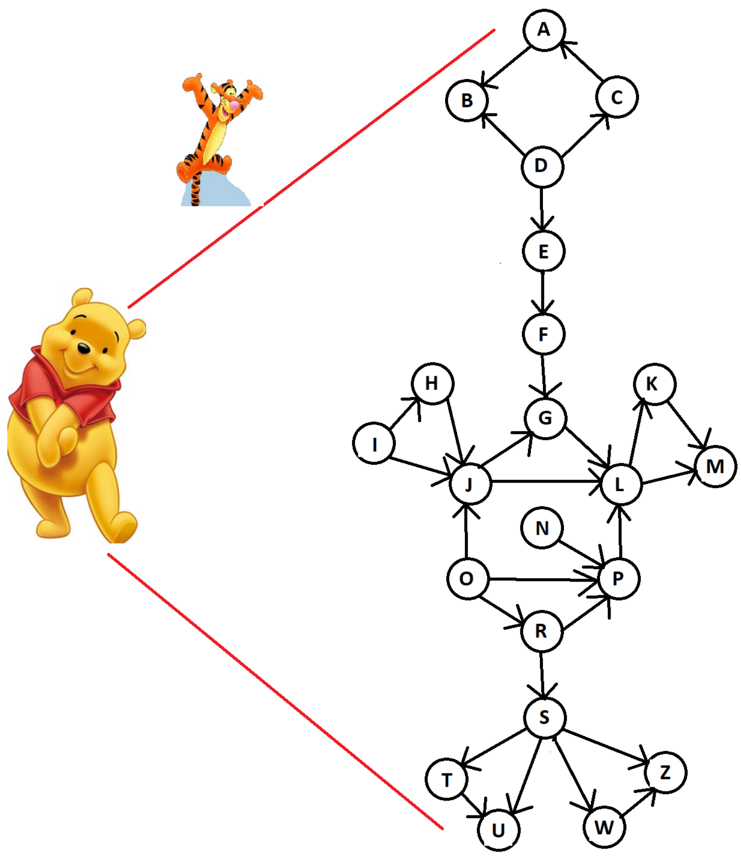
ID : hh253

Final exam

Introduction to Machine Learning
Fall 2021
Instructor: Anna Choromanska

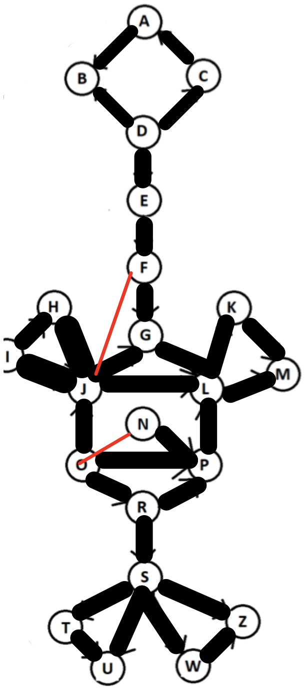
Problem 1 (100 points)

Winnie the Pooh is looking for Tiger in the forest but he can't find his friend. Winie the Pooh decided to perform the junction-tree algorithm to obtain cyber representation of his friend and post his digital photo online. Help him out by designing a junction-tree from the graph below which Winnie the Pooh should use for Tiger. Show ALL steps of creating the junction tree (including the table for the Kruskal algorithm).

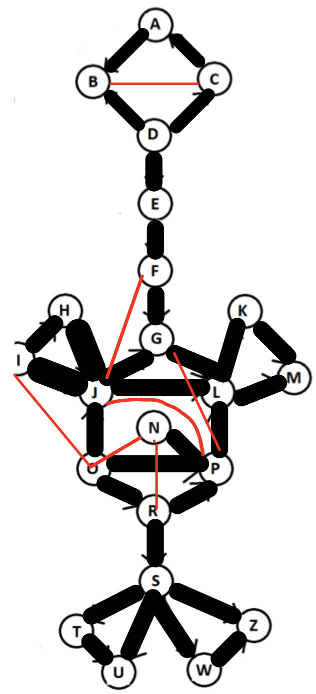


Q1. Ans.

Step1: Moralization



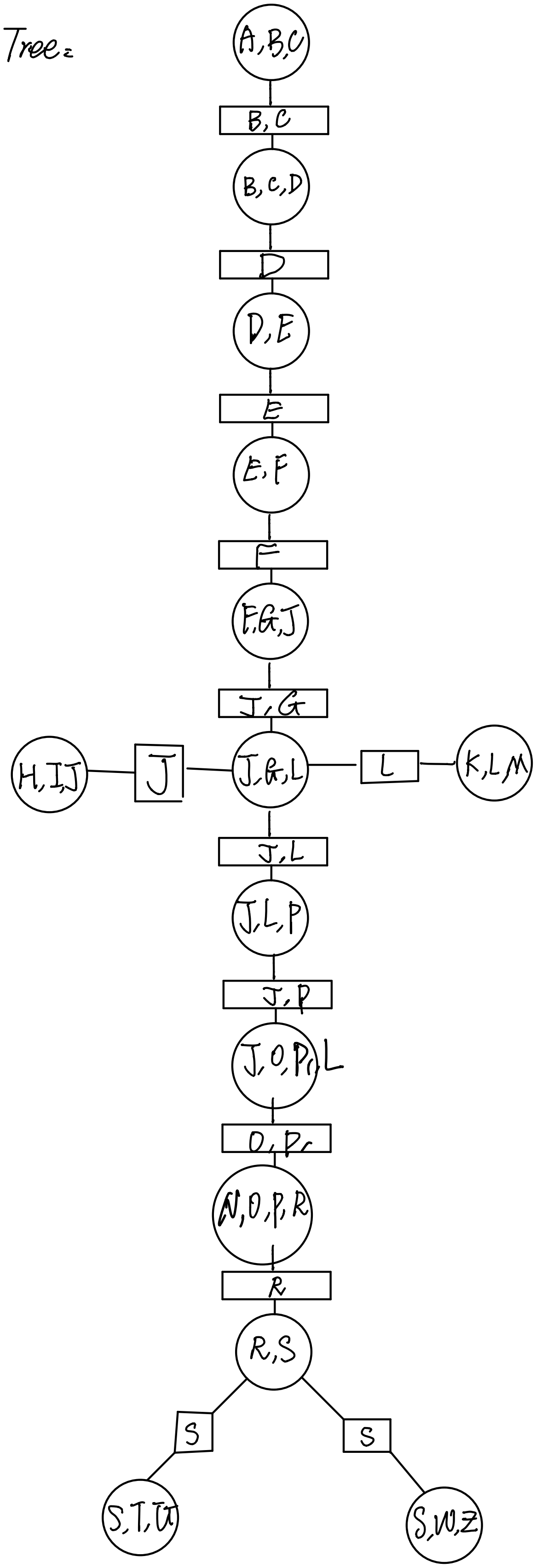
Step2: Moralization + Triangulation



Go to the next page to find Table!

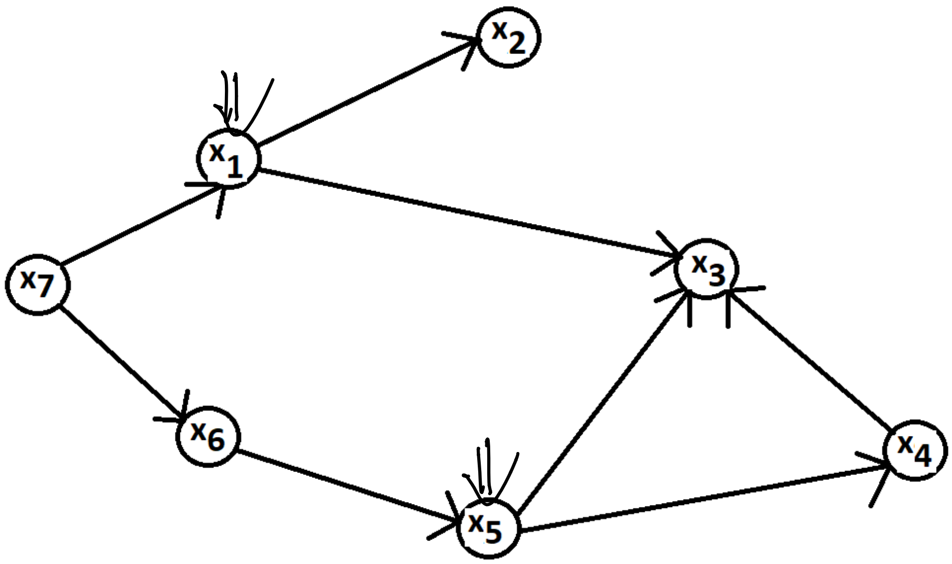
	ABC	BCD	DE	EF	FG	GJPL	LMK	JONP	NORP	HIJ	IJO	RS	STU	SZW
ABC		2												
BCD			1 1											
DE				1										
EF					1									
FG						1								
GJPL							1	2	1	1	1			
LMK														
JONP														
NORP										3	2			
HIJ											1 2	1		
IJO														
RS													1	1
STU														1
SZW														

Step3: Junction Tree:



Problem 2 (60 points)

Consider the Bayesian network below with binary variables $x_1, x_2, \ldots, x_7$.



Write out the factorization of the probability distribution $p(x_1, ..., x_7)$ implied by this directed graph. (10 points) Then, using the Bayes ball algorithm, indicate for each statement below if it is True or False and justify your answers (50 points)

Ans:

F (a) x_2 and x_6 are independent. False: Balls go through 6-7-1-2

T (b) x_2 and x_6 are conditionally independent given x_1, x_3, and x_5. True

T (c) x_1 and x_4 are conditionally independent given x_5. True

T (d) x_5 and x_2 are conditionally independent given x_1 and x_3. True

F (e) x_5 and x_1 are conditionally independent given x_3, x_2, and x_4. False: Balls go through 5-3-1

T (f) x_4 and x_7 are conditionally independent given x_6. True

T (g) x_4 and x_7 are conditionally independent given x_5. True

T (h) x_1 and x_5 are conditionally independent given x_6 and x_7. True

F (i) x_5 and x_1 are independent. False: Balls go through 1-7-6-5

T (j) x_2 and x_4 are conditionally independent given x_1. True

The factorization of the probability distribution

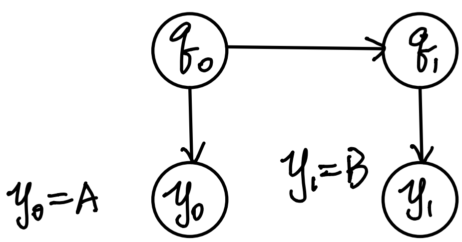
$$p(x_1, x_2, \cdots, x_7) = p(x_1|x_7) \cdot p(x_2|x_1) \cdot p(x_3|x_1, x_4, x_5) \cdot p(x_4|x_5) \cdot p(x_5|x_6)\, p(x_6|x_7) \cdot p(x_7)$$

Problem 3 (100 points)

You are given the parameters of a 2-state HMM. You observed the input sequence AB (from a 2-symbol alphabet A or B). In other words, you observe two symbols from your finite state machine, A and then B. Using the junction tree algorithm, evaluate the likelihood of this data $p(y)$ given your HMM and its parameters. Also compute (for decoding) the individual marginals of the states after the evidence from this sequence is observed: $p(q_0|y)$ and $p(q_1|y)$. The parameters for the HMM are provided below. They are the initial state prior $p(q_0)$, the state transition matrix given by $p(q_t|q_{t-1})$, and the emission matrix $p(y_t|q_t)$, respectively.

$$\pi = p(q_0) = \begin{array}{cc} 1 & 2 \\ \end{array} \left[\begin{array}{cc} 1/8 & 7/8 \end{array}\right]$$

$$a^T = p(q_t \mid q_{t-1}) = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1/3 & 1/4 \\ 2 & 2/3 & 3/4 \end{array} \qquad \eta^T = p(y_t \mid q_t) = \begin{array}{c|cc} & 1 & 2 \\ \hline A & 1/4 & 1/2 \\ B & 3/4 & 1/2 \end{array}$$



The Junction Tree:

Step1: Initialize the Junction Tree.

	$y_0=$ A	B
$q_0=1$	$\frac{1}{32}$	$\frac{3}{32}$
$q_0=2$	$\frac{7}{16}$	$\frac{7}{16}$

$q_0=$ 1	2
1	1

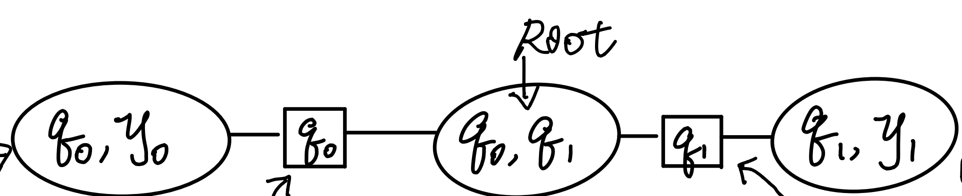
	$q_1=$ 1	2
$q_0=1$	$\frac{1}{3}$	$\frac{2}{3}$
$q_0=2$	$\frac{1}{4}$	$\frac{3}{4}$

$q_1=$ 1	2
1	1

	$y_1=$ A	B
$q_1=1$	$\frac{1}{4}$	$\frac{3}{4}$
$q_1=2$	$\frac{1}{2}$	$\frac{1}{2}$

Step2 Collect:

Root



	$y_0=$ A	B
$q_0=1$	$\frac{1}{32}$	$\frac{3}{32}$
$q_0=2$	$\frac{7}{16}$	$\frac{7}{16}$

$q_0=1$	$\frac{1}{32}$
$q_0=2$	$\frac{7}{16}$

	$q_1=$ 1	2
$q_0=1$	$\frac{1}{3}\times\frac{1}{32}\times\frac{3}{4}$	$\frac{2}{3}\times\frac{1}{32}\times\frac{1}{2}$
$q_0=2$	$\frac{1}{4}\times\frac{7}{16}\times\frac{3}{4}$	$\frac{3}{4}\times\frac{7}{16}\times\frac{1}{2}$

$q_1=$ 1	2
$\frac{3}{4}$	$\frac{1}{2}$

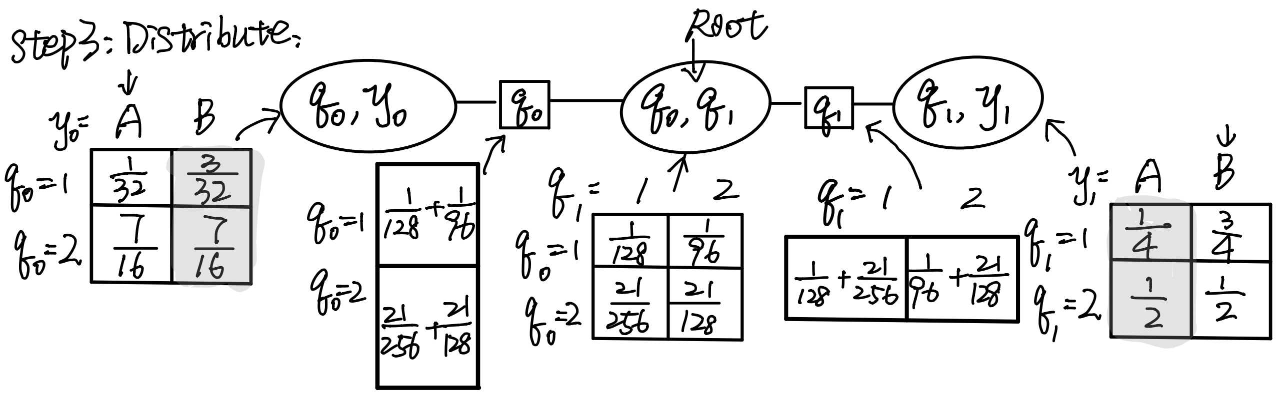
	$y_1=$ A	B
$q_1=1$	$\frac{1}{4}$	$\frac{3}{4}$
$q_1=2$	$\frac{1}{2}$	$\frac{1}{2}$

According to step2: evaluate the likelihood of p(y):

Ans = $p(y) = \frac{1}{3}\times\frac{1}{32}\times\frac{3}{4} + \frac{2}{3}\times\frac{1}{32}\times\frac{1}{2} + \frac{1}{4}\times\frac{7}{16}\times\frac{3}{4} + \frac{3}{4}\times\frac{7}{16}\times\frac{1}{2} = \frac{1}{128} + \frac{1}{96} + \frac{21}{256} + \frac{21}{128}$

$= \frac{44}{256} + \frac{21}{256} + \frac{1}{96} = \frac{132}{768} + \frac{63}{768} + \frac{8}{768} = \frac{203}{768} \approx 0.2643$

step3: Distribute:



According to step3: Decode

$$P(q_0=1|y) = \frac{\frac{1}{128}+\frac{1}{96}}{\frac{1}{128}+\frac{1}{96}+\frac{21}{256}+\frac{21}{128}} = \frac{\frac{6}{768}+\frac{8}{768}}{\frac{203}{768}} = \frac{\frac{14}{768}}{\frac{203}{768}} = \frac{14}{203} = 0.0690$$

$$P(q_0=2|y) = \frac{\frac{21}{256}+\frac{21}{128}}{\frac{1}{128}+\frac{1}{96}+\frac{21}{256}+\frac{21}{128}} = \frac{\frac{63}{768}+\frac{126}{768}}{\frac{203}{768}} = \frac{\frac{189}{768}}{\frac{203}{768}} = \frac{189}{203} = 0.9310$$

$$P(q_1=1|y) = \frac{\frac{1}{128}+\frac{21}{256}}{\frac{1}{128}+\frac{1}{96}+\frac{21}{256}+\frac{21}{128}} = \frac{\frac{6}{768}+\frac{63}{768}}{\frac{203}{768}} = \frac{\frac{69}{768}}{\frac{203}{768}} = \frac{69}{203} = 0.3340$$

$$P(q_1=2|y) = \frac{\frac{1}{96}+\frac{21}{128}}{\frac{1}{128}+\frac{1}{96}+\frac{21}{256}+\frac{21}{128}} = \frac{\frac{8}{768}+\frac{126}{768}}{\frac{203}{768}} = \frac{\frac{134}{768}}{\frac{203}{768}} = \frac{134}{203} = 0.6601$$

Problem 4 (40 points)

Show the first two iterations (after the initialization) of the k-means clustering algorithm (show centers and assignments of data points to clusters) for the following 2D data set: $(-5, 3), (-3, 2), (-4, 5), (-3, 4), (3, -4), (4, -2),$ $(6, -6), (8, -3)$. Assume the number of centers is equal to 2 and the centers are initialized to $(-4, 2)$ and $(4, -5)$.

Iteration 1: According to distance:

	(−5,3)	(−3,2)	(−4,5)	(−3,4)	(3,−4)	(4,−2)	(6,−6)	(8,−3)
center A: (−4,2)	A	A	A	A				
center B: (4,−5)					B	B	B	B

$$\text{new center } A = \left(\frac{-5-3-4-3}{4}, \frac{3+2+5+4}{4}\right) = (-3.75, 3.5)$$

$$\text{new center } B = \left(\frac{3+4+6+8}{4}, \frac{-4-2-6-3}{4}\right) = (5.25, -3.75)$$

Iteration 2: According to distance:

	(−5,3)	(−3,2)	(−4,5)	(−3,4)	(3,−4)	(4,−2)	(6,−6)	(8,−3)
center A: (−3.75,3.5)	A	A	A	A				
center B: (5.25,−3.75)					B	B	B	B

$$\text{new center } A' = \left(\frac{-5-3-4-3}{4}, \frac{3+2+5+4}{4}\right) = (-3.75, 3.5)$$

$$\text{new center } B' = \left(\frac{3+4+6+8}{4}, \frac{-4-2-6-3}{4}\right) = (5.25, -3.75)$$

Problem 5 (50 points)

Prove (using Jensen's inequality) that KL-divergence defined below is non-negative:

$$KL(p||q) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

where $p(x)$ and $q(x)$ are two probability distributions.

According to the Jensen's inequality, for any convax fiunction f,

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \le \sum_{i=1}^{n} \lambda_i f(x_i)$$

where $\lambda_i \ge 0$ and $\sum_{i=1}^{n} \lambda_i = 1$

We set $A = \{x : p(x) > 0\}$ be the support of $p(x)$. Then:

$$-KL(p||q) = -\sum_{x\in A} p(x) \cdot \log \frac{p(x)}{q(x)} = \sum_{x\in A} p(x) \log \frac{q(x)}{p(x)}$$

$$\le \log \sum_{x\in A} p(x) \frac{q(x)}{p(x)} = \log \sum_{x\in A} q(x)$$

$$\le \log \sum_{x\in X} q(x) = \log 1 = 0$$

Since $-KL(p||q) \le 0$, we have $KL(p||q) \ge 0$. nonnegative

It is proved.

Problem 6 (50 points)

Consider the fragment of the convolutional architecture given below:

- Input image: $1 \times x \times y$
- ① Convolutional layer: $\underbrace{1 \to 5}_{\text{number of input and output channels}}$, $\underbrace{3 \times 3}_{\text{filter size}}$, $\underbrace{2 \times 3}_{\text{stride}}$
- ReLU
- ② MaxPooling: $\underbrace{2 \times 2}_{\text{region size}}$, $\underbrace{2 \times 2}_{\text{stride}}$
- ③ Convolutional layer: $5 \to 7, 3 \times 3, 2 \times 2$
- ReLU
- ④ MaxPooling: $2 \times 2, 3 \times 3$
- ⑤ Flattening (3D to 1D):

$$\underbrace{7 \times 14 \times 8}_{\text{number of feature maps} \times \text{size of the feature map } (14 \times 8)} \to 784$$

What is the size of the input (in other words what is x and y)?

We calculate from the last step ⑤ and backward to the first step 1.
x_i, y_i means the size of x and y in current step.

⑤ → ④

$$= \frac{x_4 - 2}{3} + 1 = 14, \quad \frac{y_4 - 2}{3} + 1 = 8 \Rightarrow x_4 = 41, \ y_4 = 23$$

④ → ③

$$= \frac{x_3 - 3}{2} + 1 = 41, \quad \frac{y_3 - 3}{2} + 1 = 23 \Rightarrow x_3 = 83, \ y_3 = 47$$

③ → ② $=$

$$\frac{x_2 - 2}{2} + 1 = 83, \quad \frac{y_2 - 2}{2} + 1 = 47 \Rightarrow x_2 = 166, \ y_2 = 94$$

② → ①

$$\frac{x - 3}{2} + 1 = 166, \quad \frac{y - 3}{3} + 1 = 94 \Rightarrow x = 333, \ y = 282$$

Size of input: $x = 333$, $y = 282$