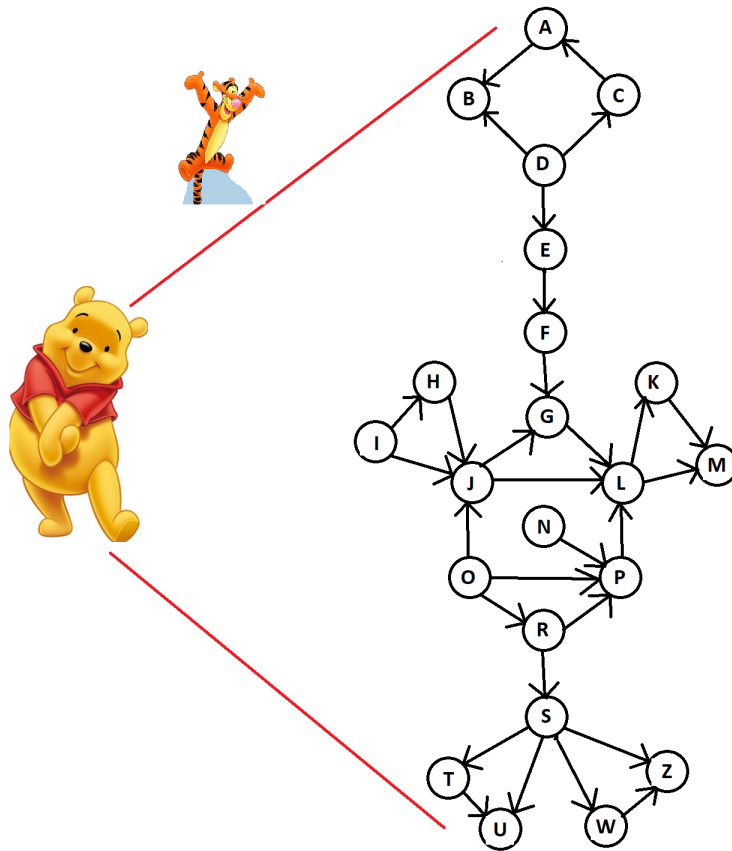


# Final exam

Introduction to Machine Learning  
Fall 2021  
Instructor: Anna Choromanska

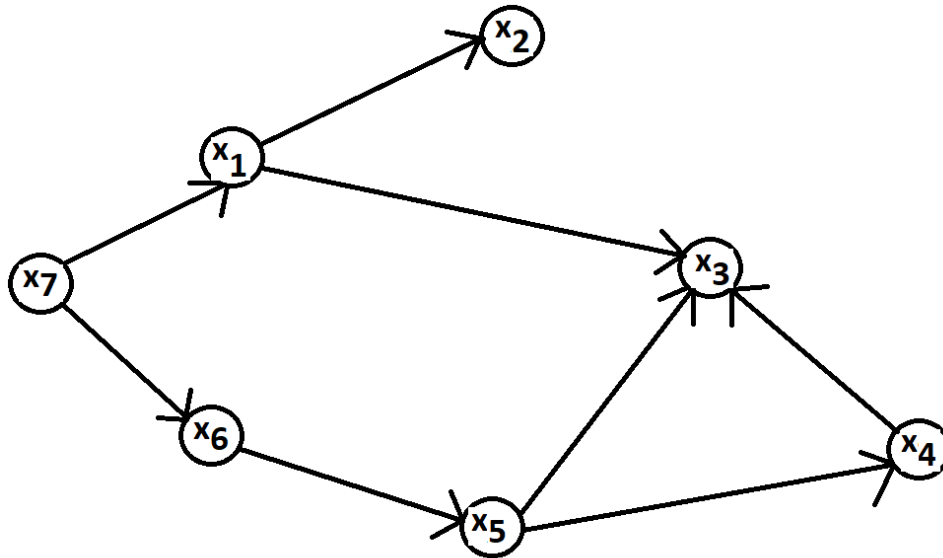
## Problem 1 (100 points)

Winnie the Pooh is looking for Tiger in the forest but he can't find his friend. Winnie 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).



## Problem 2 (60 points)

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



Write out the factorization of the probability distribution  $p(x_1, \dots, x_7)$  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)

- $x_2$  and  $x_6$  are independent.
- $x_2$  and  $x_6$  are conditionally independent given  $x_1, x_3$ , and  $x_5$ .
- $x_1$  and  $x_4$  are conditionally independent given  $x_5$ .
- $x_5$  and  $x_2$  are conditionally independent given  $x_1$  and  $x_3$ .
- $x_5$  and  $x_1$  are conditionally independent given  $x_3, x_2$ , and  $x_4$ .
- $x_4$  and  $x_7$  are conditionally independent given  $x_6$ .
- $x_4$  and  $x_7$  are conditionally independent given  $x_5$ .
- $x_1$  and  $x_5$  are conditionally independent given  $x_6$  and  $x_7$ .
- $x_5$  and  $x_1$  are independent.
- $x_2$  and  $x_4$  are conditionally independent given  $x_1$ .

### 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.

$$\pi = p(q_0) = \begin{array}{cc} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{bmatrix} 1/8 & 7/8 \end{bmatrix} \end{array}$$

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

### 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  $(4, -5)$ .

### 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.

## Problem 6 (50 points)

Consider the fragment of the convolutional architecture given below:

- Input image:  $1 \times x \times y$
- Convolutional layer:  $\underbrace{1 \rightarrow 5}_{\text{number of input and output channels}}, \underbrace{3 \times 3}_{\text{filter size}}, \underbrace{2 \times 3}_{\text{stride}}$
- ReLU
- MaxPooling:  $\underbrace{2 \times 2}_{\text{region size}}, \underbrace{2 \times 2}_{\text{stride}}$
- Convolutional layer:  $5 \rightarrow 7, 3 \times 3, 2 \times 2$
- ReLU
- MaxPooling:  $2 \times 2, 3 \times 3$
- Flattening (3D to 1D):  
 $\underbrace{7 \times 14 \times 8}_{\text{number of feature maps} \times \text{size of the feature map (14} \times 8)}} \rightarrow 784$

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