CMT311-Principles of Machine learning

- According to the changes on the examination procedure, the exam will be **open book.**

  - **That is, you are allowed to consult all the material available**

  - **Nevertheless, it is advised to know well the material**

    - **so you know what are you looking for while consulting the material, and where to find it.**

  - **The material on learning central (slides, and recordings) covers everything relevant to the exam**

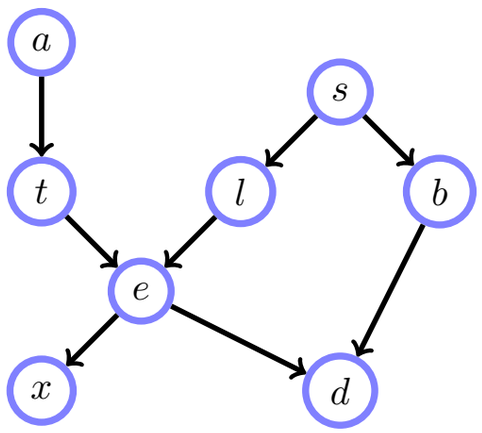
Examination Guidance

- What may be examined?
  - **Synthesis: e.g.,** the student should be able to apply a definition seen in class to a particular problem/ scenario.
    - Questions asking for computing answers using numeric values given
    - Drawing/defining a model capturing a given scenario
    - Applying an algorithm covered in class to a particular instance
  - **Multiple choice questions testing the knowledge of students regarding definitions or concepts covered in class.**
    - **Besides choosing the right answer, students should provide a justification for their choice to get full marks.**
- **Note that some examples of these kind of questions were consider in class. (c.f. exercises in slides)**

Please remember that everything covered in this module is examinable!

Question Samples

1. The Chest Clinic network concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither), a PGM capturing this scenario is shown in the figure below.



$x =$ Positive X-ray
$d =$ Dyspnea (Shortness of breath)
$e =$ Either Tuberculosis or Lung Cancer
$t =$ Tuberculosis
$l =$ Lung Cancer
$b =$ Bronchitis
$a =$ Visited Asia
$s =$ Smoker

In this model a visit to Asia is assumed to increase the probability of tuberculosis. State if the following conditional independence relationships are true or false (**Justify your answer**)

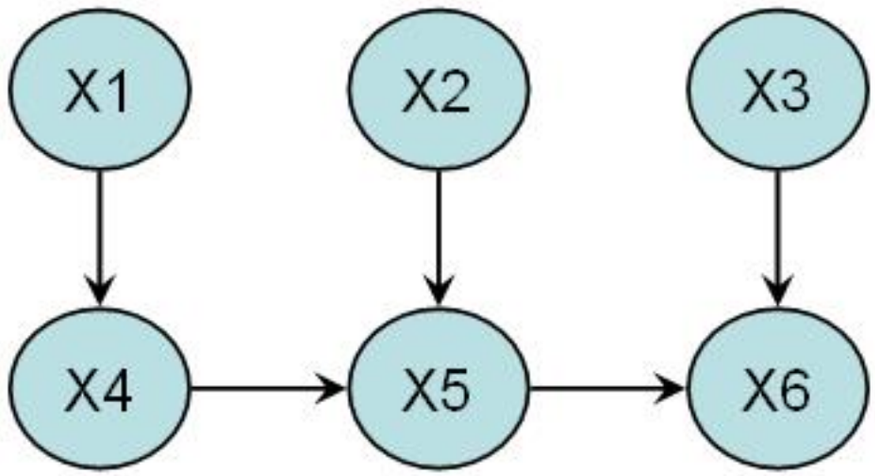
(a) tuberculosis $\perp$ smoking | shortness of breath

(b) lung cancer $\perp$ bronchitis | smoking

(c) visit to Asia $\perp$ smoking | lung cancer

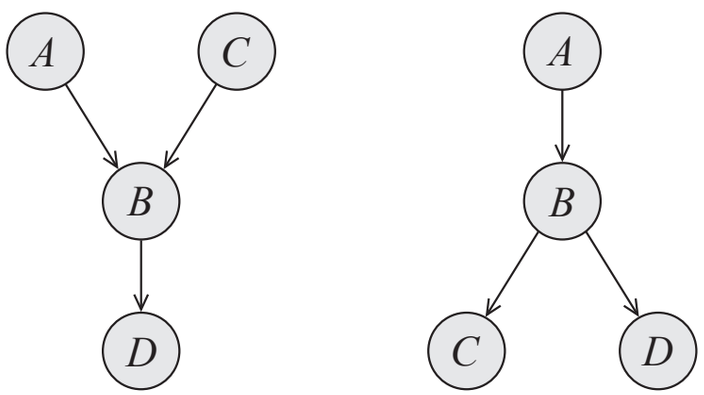
(d) visit to Asia $\perp$ smoking | lung cancer, shortness of breath

2. (a) Consider running variable elimination on the following Bayesian network over binary variables. Which of the nodes, if eliminated first, results in the largest intermediate factor? By largest factor we mean the factor with the largest number of entries.



•

2. (b) Consider the two networks:



(a) (b)

For each of them, determine whether there can be any other Bayesian network that is I-equivalent to it.

3. A local supermarket specialising in breakfast cereals decides to analyse the buying patterns of its customers. They make a small survey asking 6 randomly chosen people their age (older or younger than 60 years) and which of the breakfast cereals (Cornflakes, Frosties, Sugar Puffs, Branflakes) they like. Each respondent provides a vector with entries 1 or 0 corresponding to whether they like or dislike the cereal. Thus a respondent with (1101) would like Cornflakes, Frosties and Branflakes, but not Sugar Puffs.

- The older than 60 years respondents provide the following data (1000), (1001), (1111), (0001).

- The younger than 60 years old respondents responded (0110), (1110).

A novel customer comes into the supermarket and says she only likes Frosties and Sugar Puffs. Using naive Bayes trained with maximum likelihood, what is the probability that she is younger than 60?

4. **Parameter Estimation with Missing Data.** The process of learning Bayesian Network parameters with missing data (partially observed instances) is more difficult than learning with complete data for which of the following reasons? You may select one or more options.

(a) While there is still always a single optimal value for the parameters, it can only be found using an iterative method.

(b) Because there can be multiple optimal values, we must always run our learning algorithm multiple times from different initialisations to make sure we find ALL of them.

(c) We lose local decomposition, whereby each CPD can be estimated independently.

(d) We require more training data, because we must throw out all incomplete instances.

5. Consider Boolean concept learning with the object set X containing all points (x,y) on the square grid with coordinates from 0 to 10, and hypotheses of the form (a≤x≤b ∧ c≤y≤d) with a,b,c,d integers in [0,10]

- What is the most general hypothesis in this space, and what is the most specific one?

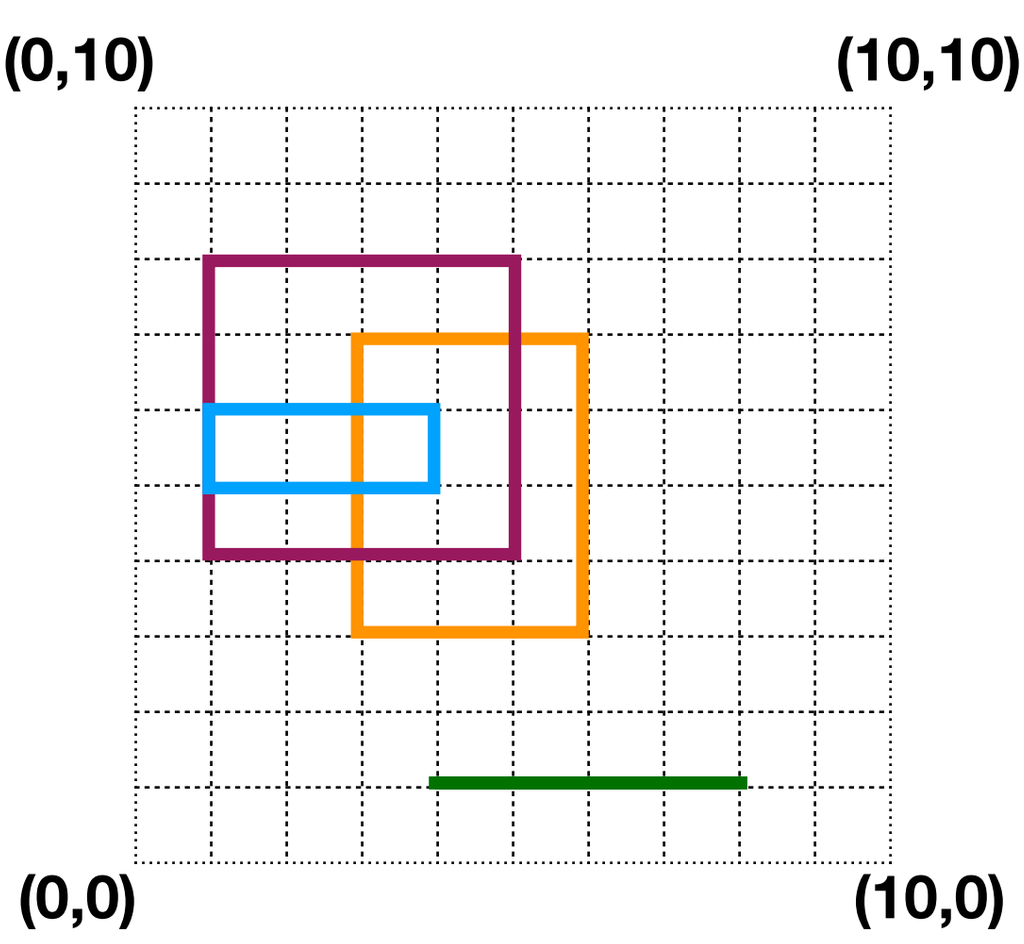
- Give a graphical interpretation of the "more general than" order for this space.

(1≤x≤4 ∧ 5≤y≤6)

(3≤x≤6 ∧ 3≤y≤7)

(1≤x≤5 ∧ 4≤y≤8)

(4≤x≤8 ∧ 1≤y≤1)



- Consider again the space of rectangles ($a \le x \le b \wedge c \le y \le d$) on the [0,10]x[0,10] grid.
- Trace the FIND-S algorithm for the following sequence of examples:
  (2,4) 1
  (7,4) 1
  (5,1) 0
  (5,3) 1
  (2,6) 0
  (6,5) 1