Department of Computer Science

Examination paper for TDT4171 - Artificial Intelligence Methods

Examination date: May 10th, 2022

Examination time (from-to): 1500-1900

Permitted examination support material: A / All support material is allowed

Academic contact during examination:
Use Blackboard Collaborate on the course's webpage

Technical support during examination: Orakel support services
Phone: 73 59 16 00

If you experience technical problems during the exam, contact Orakel support services as soon as possible before the examination time expires/the test closes. If you don't get through immediately, hold the line until your call is answered.

OTHER INFORMATION

Do not open Inspera in multiple tabs, or log in on multiple devices, simultaneously. This may lead to errors in saving/submitting your answer.

Get an overview of the question set before you start answering the questions.

Read the questions carefully, make your own assumptions and specify them in your answer. Only contact academic contact if you think there are errors or insufficiencies in the question set.

Cheating/Plagiarism: The exam is an individual, independent work. Examination aids are permitted, but make sure you follow any instructions regarding citations. During the exam it is not permitted to communicate with others about the exam questions or distribute drafts for solutions. Such communication is regarded as cheating. All submitted answers will be subject to plagiarism control. *Read more about cheating and plagiarism here.*

Notifications: If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspera. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.

Weighting:

Please note the following:

- *The question-set can give you up to 80 points. The remaining 20 points have been available through assignments. It is the sum of points from the exam and the assignments that gives your grade.*
- *The grade is assigned according to the rule that you will get a Pass if and only if you have at least 41 points in total.*
- *It is noted for each question how many points you get from that question, and (if applicable) how sub-questions are weighted.*

ABOUT SUBMISSION

Answering in Inspera: If the question set contains questions that are not upload assignment, you must answer them directly in Inspera. In Inspera, your answers are saved automatically every 15 seconds.

NB! We advise against pasting content from other programs, as this may cause loss of formatting and/or entire elements (e.g. images, tables).

File upload:

All files must be uploaded before the examination time expires.

The file types allowed are specified in the upload assignment(s). Note that it is only possible to upload one file per upload assignment.

30 minutes are added to the examination time to manage the sketches/calculations/files. The additional time is included in the remaining examination time shown in the top left-hand corner.

NB! You are responsible to ensure that the file(s) are correct and not corrupt/damaged. Check the file(s) you have uploaded by clicking “Download” when viewing the question. All files can be removed or replaced as long as the test is open.

How to digitize your sketches/calculations

How to create PDF documents

Remove personal information from the file(s) you want to upload

Automatic submission: Your answer will be submitted automatically when the examination time expires and the test closes, as long as you have answered at least one question. This will happen even if you do not click “Submit and return to dashboard” on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted. This is considered as “did not attend the exam”.

Withdrawing from the exam: If you become ill during the exam or wish to submit a blank answer/withdraw from the exam for another reason, go to the menu in the top right-hand corner and click “Submit blank”. This cannot be undone, even if the test is still open.

Accessing your answer post-submission: You will find your answer in Archive when the examination time has expired.

1 Part I:

Model the following domain using a Bayesian network. Make your model as simple and easy to understand as possible:

Peter wakes up in the morning with a sore throat. It can be either due to a common cold or due to an angina. Peter wants to know what is wrong with him, as he will not go to work if he has the angina. To better understand his problem, he can:

- *Measure his temperature to detect if he has a fever*
- *Look for yellow spots in his throat*

Both the common cold as well as the angina can result in a fever, but only angina gives rise to the yellow spots in the throat.

You are not asked to give the conditional probability distributions in this part, only the qualitative of the Bayesian network (the graph).

Part II:

What modelling assumptions about independences and conditional independences can be read off your model structure?

Part III:

Which conditional probability tables are required for the modelling of the domain to be complete? Suggest some relevant numbers, and calculate the probability for Peter to have a common cold given that he has a fever (38 degrees body temperature), but no yellow spots in his throat.

Part IV:

Peter is wondering if he should go to the office or not. If it turns out he has the angina he will become very unpopular if he shows up at work, but get a lot of sympathy if he calls in sick. If he only has the common cold, his colleagues will think he has poor work-ethics if he chooses to stay home. Extend your model to give support about this decision. You do not have to enter any numerical values or make numerical calculations in this Part.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 15

2(a) Assume we have *n* variables in a Bayesian network, X_1, X_2 , ..., X_n and that the first *k* variables X_1, X_2 , ..., X_k for a given $k < n$ are such that they do not have any parents in the network. Then the model always asserts that
$P(x_1, x_2, \dots x_n) = P(x_1) \cdot P(x_2) \cdot P(x_3) \cdot \ \dots \ \cdot P(x_n)$.
Select one alternative:

- True
- False

Maximum marks: 1

2(b) Let X and Y be two discrete random variables, with N_x and N_y states respectively. Then you will *never* need more than $N_x \cdot N_y$ parameters to fully describe the joint distribution $\mathbf{P}(X, Y)$
Select one alternative:

- False
- True

Maximum marks: 1

2(c) Among the types of inference that can be done with a *Hidden Markov Model*, we find *filtering* and *smoothing*.
Select one alternative:

- False
- True

Maximum marks: 1

2(d) In a Bayesian network with *n* binary variables, the complexity will in the worst case grow exponentially in *n*. The complexity of the representation is measured as the total number of parameters in the conditional probability tables.

Select one alternative:

- False
- True

Maximum marks: 1

2(e) Name the inference-task in a Hidden Markov Model when one calculates $\mathbf{P}(\boldsymbol{X}_{t+k}|\mathbf{e}_{1:t})$ for $k > 0$

Select one alternative:

- Smoothing
- Prediction
- Filtering

Maximum marks: 1

2(f) If you make a Hidden Markov model, you have to assume *stationarity*.

Select one alternative:

- False
- True

Maximum marks: 1

2(g) Let X, Y and Z be three discrete random variables. Is it possible to calculate $P(X=0, Y=1|Z=0)$ from the joint distribution $\mathbf{P}(X, Y, Z)$?

Select one alternative:

- Yes, using the formula $P(X=0, Y=1|Z=0) = \frac{P(X=0,Y=1,Z=0)}{P(X=0,Y=1|Z=0)}$
- Yes, using the formula $P(X=0, Y=1|Z=0) = \frac{P(X=0,Y=1,Z=0)}{\sum_x \sum_y P(X=x,Y=y,Z=0)}$
- Yes, using the formula $P(X=0, Y=1|Z=0) = \frac{P(X=0,Y=1,Z=0)}{\sum_x \sum_y P(X=x,Y=y|Z=0)}$
- No

Maximum marks: 1

2(h) Deep learning cannot be used to analyze text data

Select one alternative:

- True
- False

Maximum marks: 1

2(i) When a learning algorithm for decision-trees uses reduction in remaining entropy to find the next split-node, then that is just a heuristic and does not guarantee that the learned decision tree will be optimal.

Select one alternative:

- True
- False

Maximum marks: 1

2(j) When solving a Markov Decision Problem with infinite time horizon by discounting, the solution no longer describes a rational behavior.

Select one alternative:

- True
- False

Maximum marks: 1

2(k) Which of the following claims holds for three discrete random variables *X*, *Y*, and *Z* ?

Select one alternative:

- $\mathbf{P}(X|Y,Z) = \frac{\mathbf{P}(Z|X,Y)\cdot\mathbf{P}(X|Y)}{\mathbf{P}(Z|Y)}$
- $\mathbf{P}(X,Y|Z) = \mathbf{P}(X,Y,Z)\cdot\frac{\mathbf{P}(X|Z)}{\mathbf{P}(Y|Z)}$
- $\mathbf{P}(X,Y,Z) = \mathbf{P}(X,Y) + \mathbf{P}(Y,Z) + \mathbf{P}(Y,Z)$

Maximum marks: 1

2(l) A Bayesian network will often give a compact representation of the joint distribution modelled. Nevertheless, the worst-case space complexity of the model is exponential in the number of variables in the domain.

Select one alternative:

- True
- False

Maximum marks: 1

2(m) If a system passes the Turing-test, it proves that the strong AI hypothesis is true
Select one alternative:

- True
- False

Maximum marks: 1

2(n) A deterministic boolean function *f* over the *k* boolean variables $X_1, X_2, \ldots, X_k$, i.e. the function $f(X_1, X_2, \ldots, X_k)$, can be represented by a Bayesian network if and only *f* is linearly separable.
Select one alternative:

- False
- True

Maximum marks: 1

2(o) If an intelligent agent is not getting the information it usually gets, it means that it at the same time looses the ability to be rational.
Select one alternative:

- True
- False

Maximum marks: 1

2(p) Name the inference-task in a Hidden Markov Model when one calculates $\mathbf{P}(X_t|\mathbf{e}_{1:t})$:
Select one alternative:

- Smoothing
- Prediction
- Filtering

Maximum marks: 1

2(q) The local and the global semantics of a Bayesian network are identical in the sense that you can guarantee one if the other holds.
Select one alternative:

- False
- True

Maximum marks: 1

2(r) In a Bayesian network with *n* binary variables, the complexity will in the worst case grow linearly in *n*. The complexity of the representation is measured as the total number of parameters in the conditional probability tables.
Select one alternative:

- True
- False

Maximum marks: 1

2(s) If A and B are discrete random variables with $P(A=a, B=b) = P(A=a)$ we know that $\frac{1}{2} \leq P(B=b|A=a) \leq \frac{1}{\sqrt{2}}$

Select one alternative:

- False
- True

Maximum marks: 1

2(t) Name the inference-task in a Hidden Markov Model when one calculates $\mathbf{P}(X_k|\mathbf{e}_{1:t})$ for $0 \leq k < t$:

Select one alternative:

- Filtering
- Prediction
- Smoothing

Maximum marks: 1

2(u) The algorithms "Value iteration" and "Policy iteration" can both be used to solve sequential decision problems, and rest on exactly the same set of assumptions.

Select one alternative:

- False
- True

Maximum marks: 1

2(v) If one succeeds in building a fully rational agent, this agent will pass the Turing test if and only if the agent has calculated that passing the Turing test maximises its expected utility.

Select one alternative:

- False
- True

Maximum marks: 1

2(w) Case-based reasoning is used for machine learning

Select one alternative:

- True
- False

Maximum marks: 1

2(x) Every *deterministic* function over *n* binary variables can be represented by a Bayesian network

Select one alternative:

- True
- False

Maximum marks: 1

2(y) Let X and Y be two discrete random variables, with N_x and N_y states, respectively. Then it *always* holds that $P(X = x) \geq P(X = x, Y = y)$

Select one alternative:

- True
- False

Maximum marks: 1

2(z) There always exists a utility-function that can be used to explain the behavior of a rational agent.

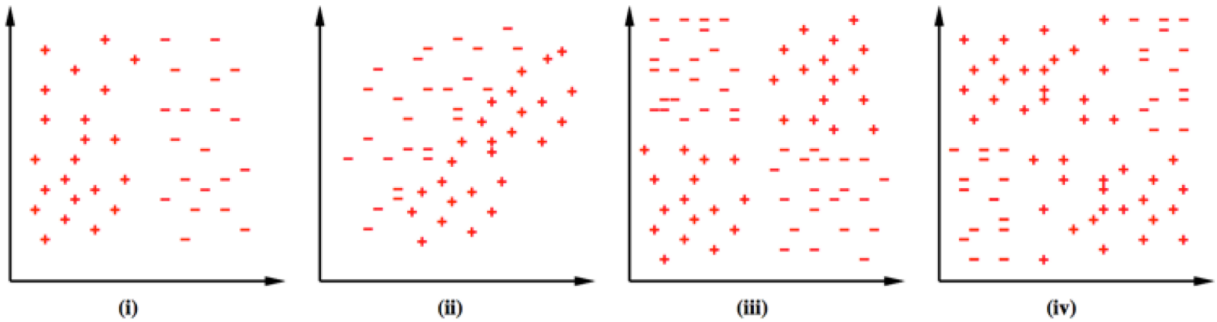
Select one alternative:

- False
- True

Maximum marks: 1

3(a) Note! You will get the full score for this question if you answer all sub-questions correctly. You get zero points from this question if you answer wrong (or leave your answer empty) for at least one sub-question.

The figure below shows four different datasets.



Common for all datasets is that each data-point consist of two continuous attributes (one shown along the *x*-axis, the other along the *y*-axis). Each example is also a member of a class, either the *positive* class (marked by "+") or the *negative* class (marked by "-").

You are now asked to choose the machine learning algorithm (among two alternatives you will be given for each dataset) that is most suitable for that dataset. When you do your deliberations you should think about each learning algorithm uses the data, and what kind of *decision boundary* each can easily represent. **Hint:** Remember that decision trees can handle continuous attributes by introducing *split points*.

Dataset (i):
Select one alternative:

- Perceptron
- Deep neural network using convolutions

Dataset (ii):
Select one alternative:

- Decision-tree
- Perceptron

Dataset (iii):
Select one alternative:

- Case-based reasoning
- Perceptron

Dataset (iv):
Select one alternative:

- Decision-tree
- Case-based reasoning

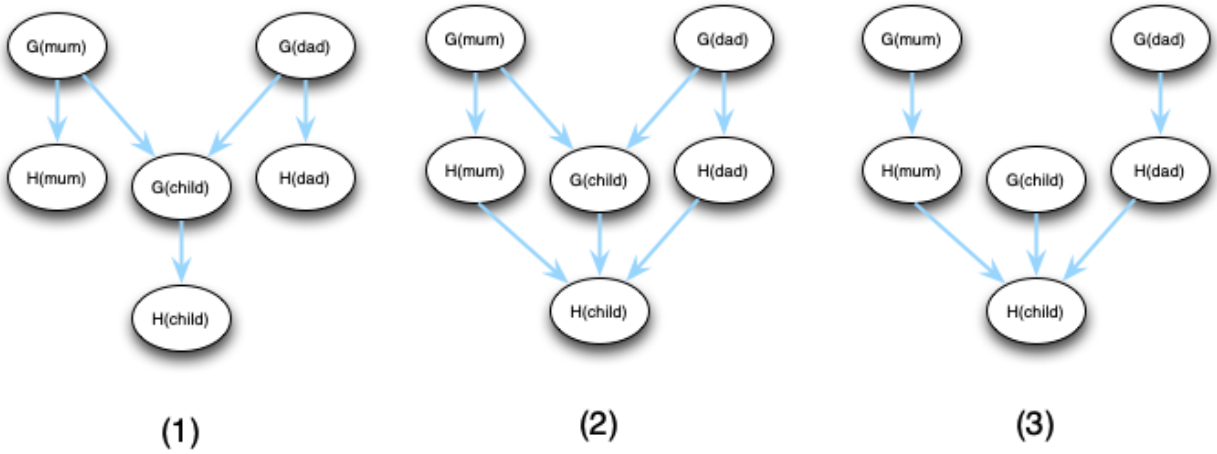
Maximum marks: 15

3(b) Note! You will get the full score for this question if you answer all sub-questions correctly. You get zero points from this question if you answer wrong (or leave your answer empty) for at least one sub-question.

Let *H(x)* be a random variable telling if person *x* is *left-handed* or *right-handed*. (Hence, *H(dad)* is for instance giving the information of the preferred hand for *dad*.)

One model to explain the "handed-ness" of a given person is that there is a gene that determines the handedness of the person with a certain probability.
Let *G(x)* represent the gene for person *x*, and assume also that this variable has the states *left-handed* and *right-handed*. Assume further that *H(x)* takes the same value as *G(x)* with a given probability. Finally, we assume that a child inherits the gene from its parents, and that it is equally probable that the gene is inherited from the father as it is from the mother.

These are three candidate models that we will consider:



Which of the model(s), one or several, in the figure above make **independence statements** that are consistent with the explanation above?
Select one or more alternatives:

- [ ] Model (2)
- [ ] Model (3)
- [ ] Model (1)

Which of the three networks in the figure is the *best* representation of the explanation above if we assume that it is only G(x) that influences H(x)?
Select one alternative:

- ( ) Model (2)
- ( ) Model (3)
- ( ) Model (1)

Which model(s), one or several, of the three networks in the figure above asserts that $\mathbf{P}(G(mum), G(dad), G(child)) = \mathbf{P}(G(mum)) \cdot \mathbf{P}(G(dad)) \cdot \mathbf{P}(G(child))$?
Select one or more alternatives:

- [ ] Model (3)
- [ ] Model (1)
- [ ] Model (2)

Maximum marks: 15

4 Note! You will get the full score for this question if you answer all sub-questions correctly. You get zero points from this question if you answer wrong (or leave your answer empty) for at least one sub-question.

Let us assume we have a dataset with *N* training-examples. Each training-example describes a *k*-dimensional input-vector **x** and a target-value *y*, meaning that our data is the collection $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$. For this question we will assume that all values are real numbers.

You train a linear regression model to find the relation between the input-vector $\boldsymbol{x}$ and target *y*. If we use $\mathbf{x}_{ij}$ to denote the value of $\mathbf{x}_i$ in dimension *j*, then this means that we assume the relationship $\hat{y}_i = \sum_{j=1}^{k} x_{i,j} \cdot w_j$. (We use $\hat{y}_i$ in the equation to make it clear that this is our guess, and not the observed value in the dataset, and let the w_j-s be weights we will use a machine learning algorithm to learn as well as we can). We will now consider this model, and see how it works.

Assume first that *N* = 1.000.000, while *k* = 10 and that you get a *very high loss on the training-data*. We measure this loss by $\mathcal{L} = \frac{1}{2} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$, and we want to make **one** change to improve the situation. Which of the suggestions below seems most promising given the description above?

Select one alternative:

- Try to obtain more data to get a more complete description of the domain (**increase** *N*)
- Try to extend the object-description, so that each $\mathbf{x}_i$ gets a "richer" representation (**increase** *k*)
- Do feature-selection to avoid overfitting caused by irrelevant dimensions interfering (**reduce** *k*)

Next we decide to learn the weights by implementing gradient descent ourselves to ensure that the weights are learned as well as possible. Which of the equations below is the correct one to make the iterative update? Note that $\eta > 0$ is the *learning-rate* and that the equation only gives the update for a single weight w_j.

Select one alternative:

- $w_j \leftarrow w_j + \eta \cdot \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial x_i}$
- $w_j \leftarrow w_j - \eta \cdot \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial x_i}$
- $w_j \leftarrow w_j - \eta \cdot \frac{\partial \mathcal{L}}{\partial w_j}$
- $w_j \leftarrow w_j + \eta \cdot \frac{\partial \mathcal{L}}{\partial w_j}$

Finally, assume that we now have *k=1*, meaning that our model simplifies to $\hat{y}_i = w \cdot x_i$. We have a small dataset with *N=3* and values {(1, 1), (2, 2), (3, 3)}, where the first number in a tuple is the value for x_i, the second is the value of y_i. Let us assume we start from the initial weight-value 0, and that the learning-rate is $\eta = 0.1$.

What is the updated weight-value after a single step of gradient descent? Use the loss-function defined previously. Give your answer with one digit after the separator (e.g. 7.3 or -0.9).

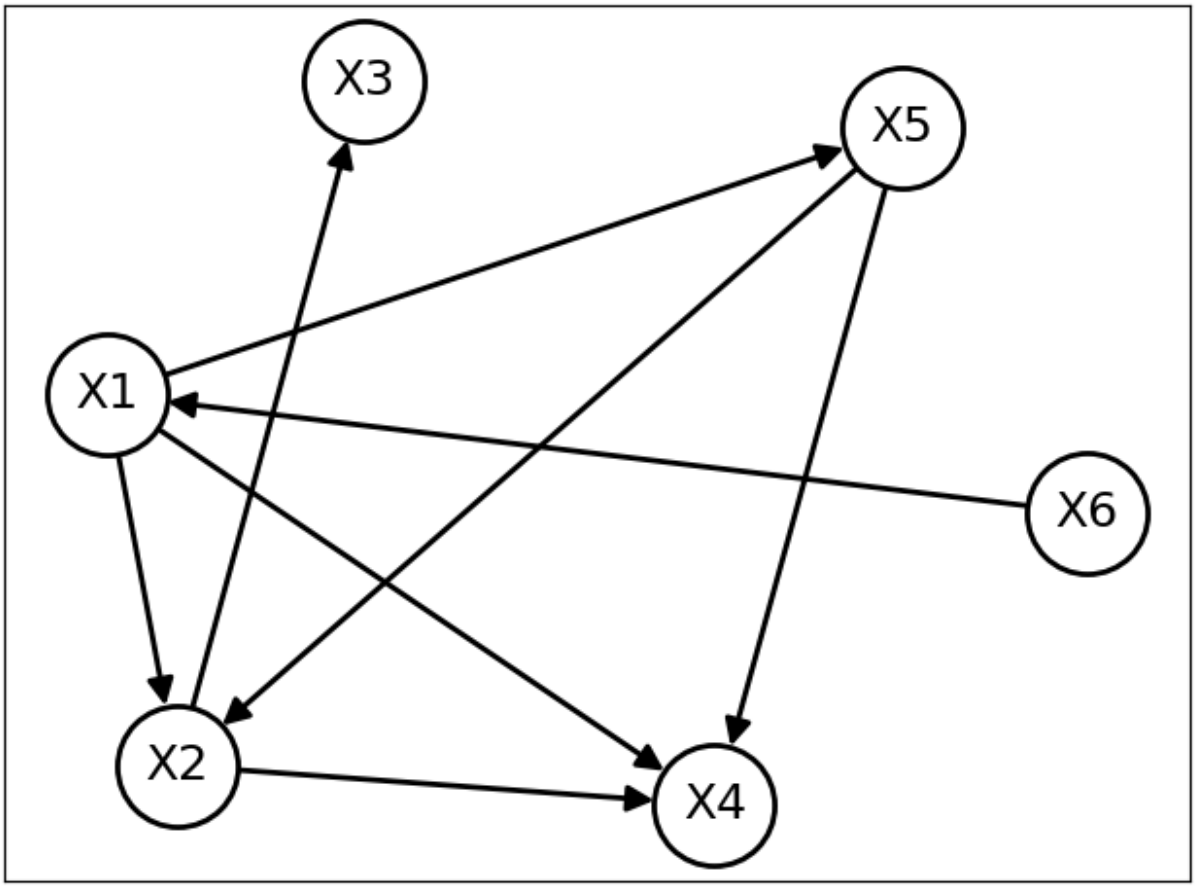
Update:

$w \leftarrow$ [ ]

Maximum marks: 15

5(a) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X3 is independent of X6

Select one alternative:

- False
- True

X1 is conditionally independent of X3 given X2

Select one alternative:

- False
- True

X4 is conditionally independent of X6 given {X2, X3}

Select one alternative:

- True
- False

X2 is conditionally independent of X6 given {X3, X4, X5}
Select one alternative:

- True
- False

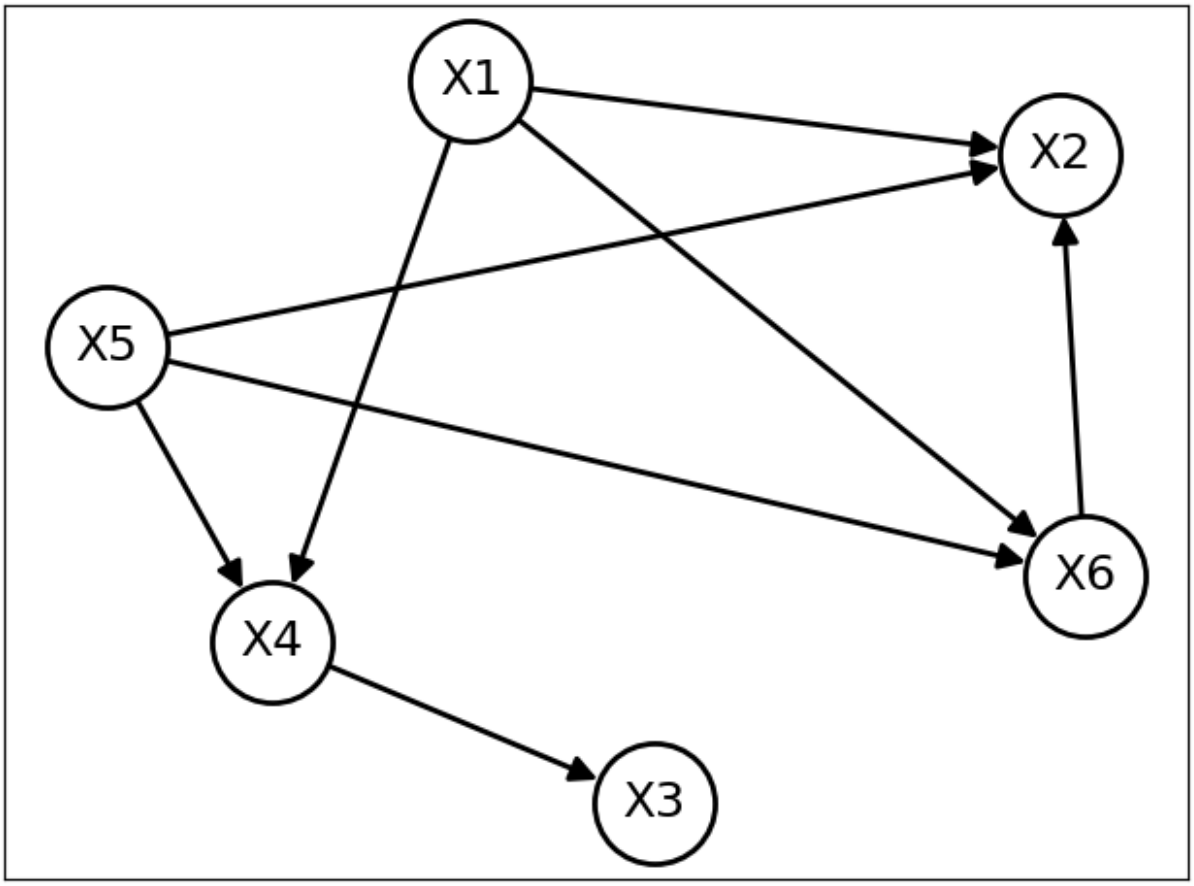
X1 is conditionally independent of X3 given {X2, X4, X5, X6}
Select one alternative:

- False
- True

Maximum marks: 2.5

5(b) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X1 is independent of X3

Select one alternative:

- False
- True

X2 is conditionally independent of X4 given X1

Select one alternative:

- True
- False

X3 is conditionally independent of X5 given {X1, X4}

Select one alternative:

- True
- False

X1 is conditionally independent of X3 given {X4, X5, X6}
Select one alternative:

- True
- False

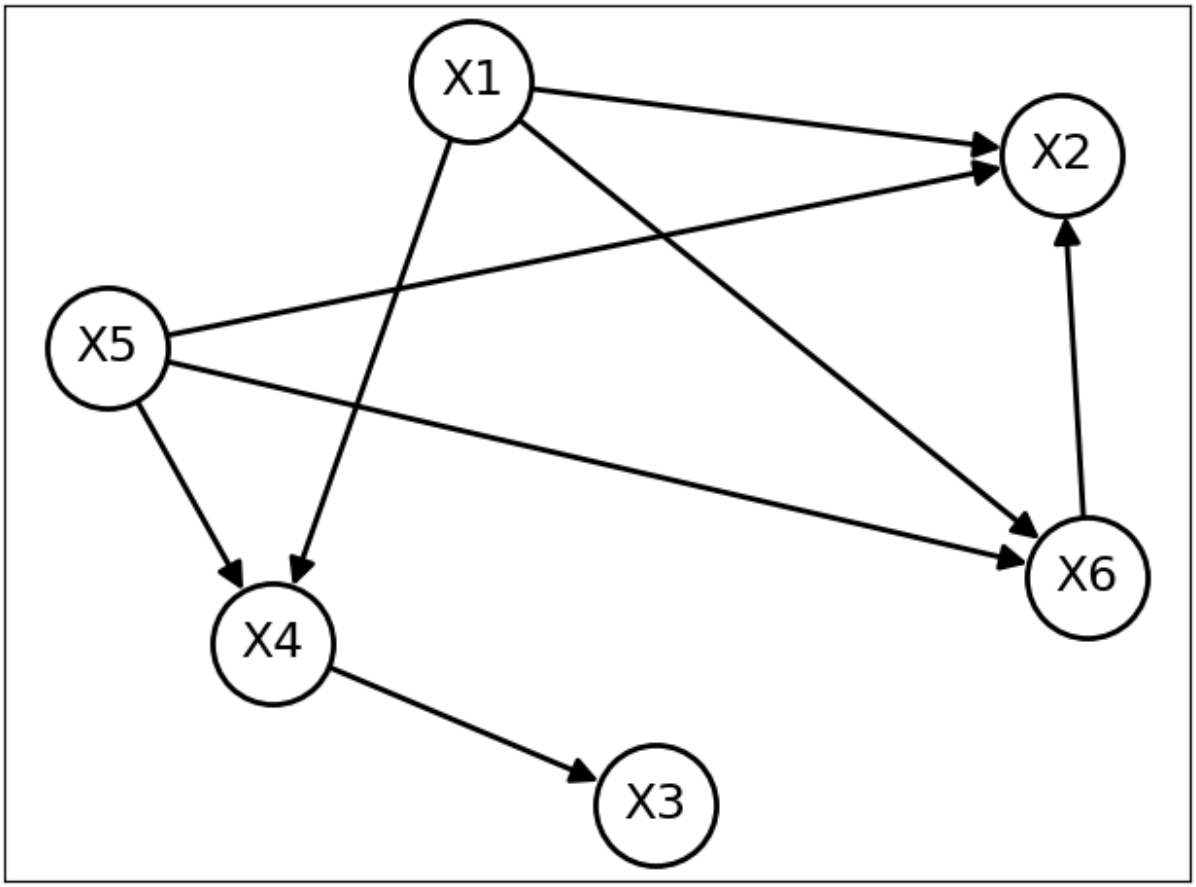
X3 is conditionally independent of X5 given {X1, X2, X4, X6}
Select one alternative:

- False
- True

Maximum marks: 2.5

5(c) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X2 is independent of X4

Select one alternative:

- True
- False

X4 is conditionally independent of X6 given X3

Select one alternative:

- True
- False

X1 is conditionally independent of X3 given {X4, X6}

Select one alternative:

- False
- True

X3 is conditionally independent of X6 given {X1, X4, X5}
Select one alternative:

- True
- False

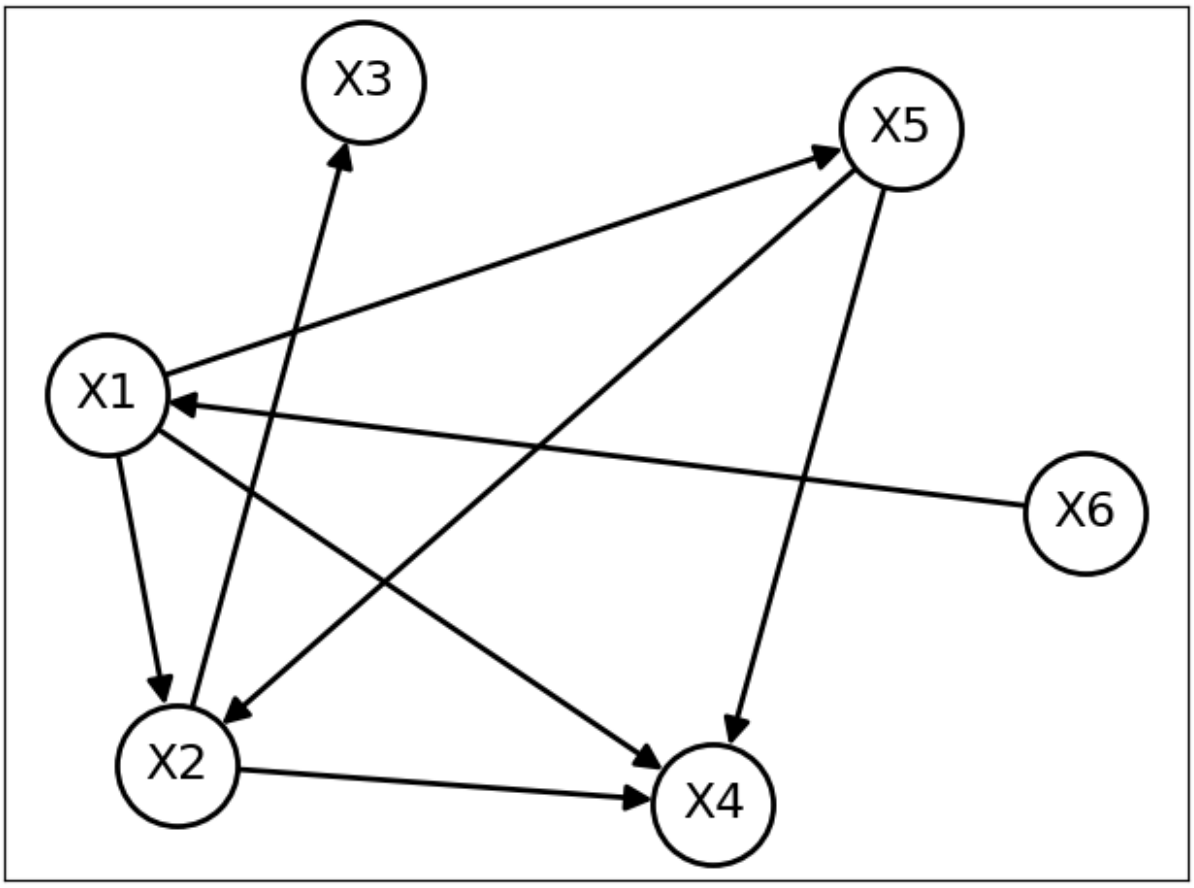
X3 is conditionally independent of X6 given {X1, X2, X4, X5}
Select one alternative:

- True
- False

Maximum marks: 2.5

5(d) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X3 is independent of X5

Select one alternative:

- True
- False

X4 is conditionally independent of X6 given X2

Select one alternative:

- False
- True

X5 is conditionally independent of X6 given {X3, X4}

Select one alternative:

- False
- True

X2 is conditionally independent of X6 given {X1, X3, X4}
Select one alternative:

- True
- False

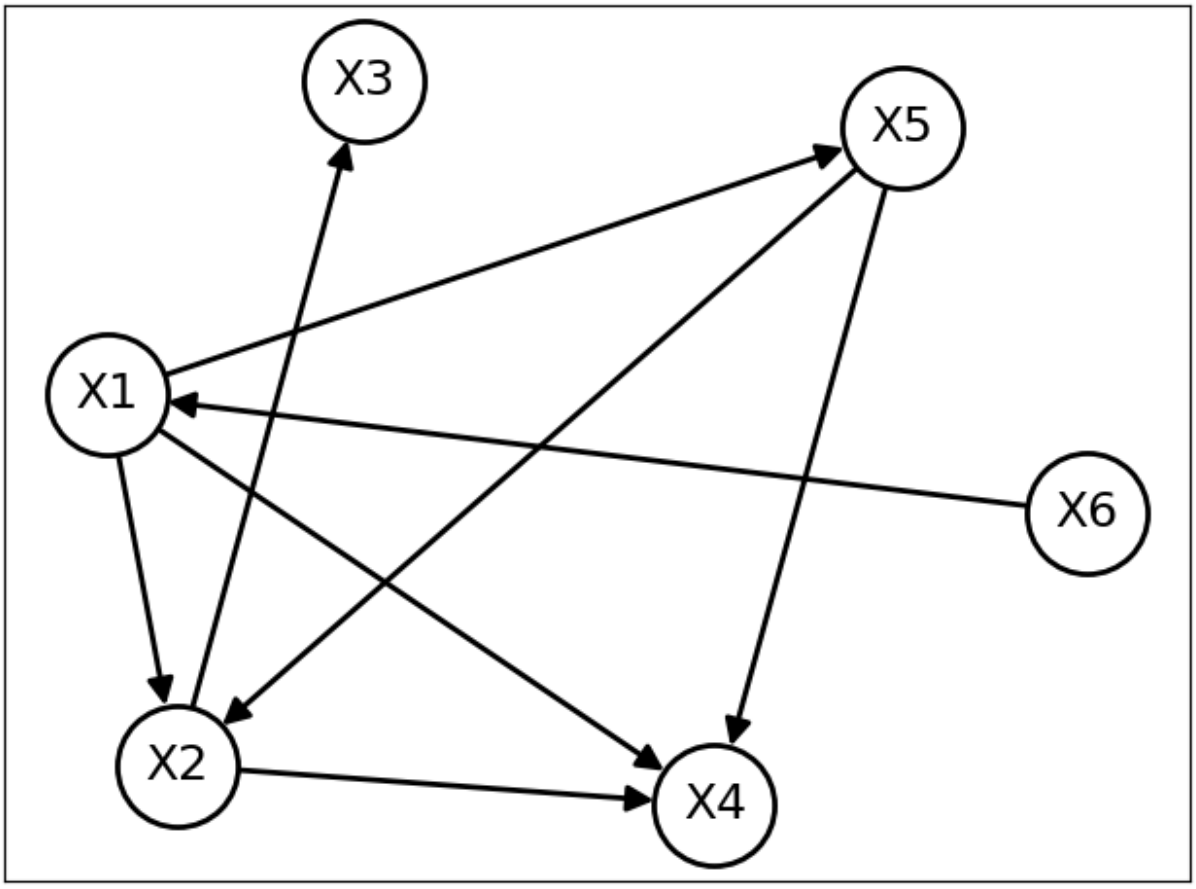
X3 is conditionally independent of X5 given {X1, X2, X4, X6}
Select one alternative:

- True
- False

Maximum marks: 2.5

5(e) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X1 is independent of X3

Select one alternative:

- False
- True

X4 is conditionally independent of X6 given X1

Select one alternative:

- True
- False

X5 is conditionally independent of X6 given {X1, X4}

Select one alternative:

- True
- False

X4 is conditionally independent of X6 given {X1, X2, X5}
Select one alternative:

- True
- False

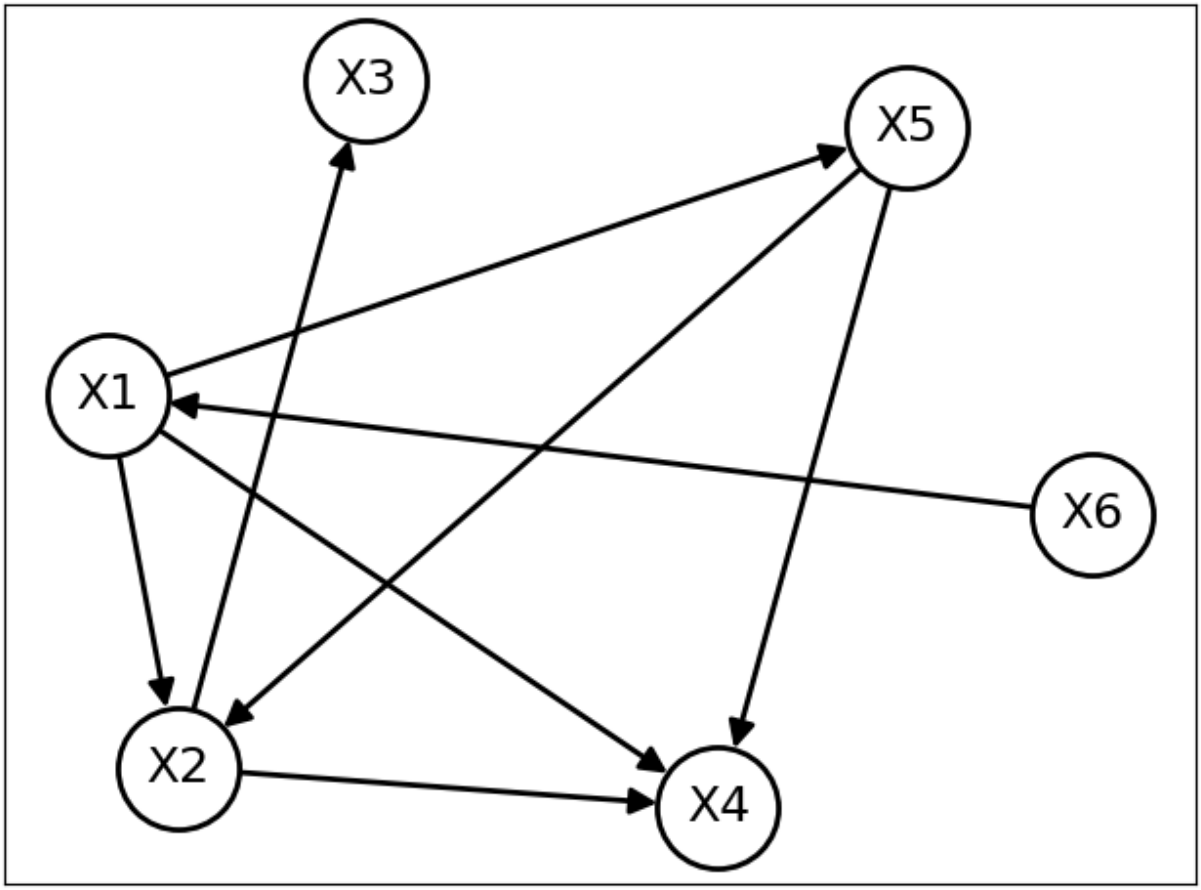
X2 is conditionally independent of X6 given {X1, X3, X4, X5}
Select one alternative:

- True
- False

Maximum marks: 2.5

5(f) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X4 is independent of X6

Select one alternative:

- False
- True

X2 is conditionally independent of X6 given X4

Select one alternative:

- False
- True

X3 is conditionally independent of X6 given {X1, X2}

Select one alternative:

- True
- False

X5 is conditionally independent of X6 given {X2, X3, X4}
Select one alternative:

- True
- False

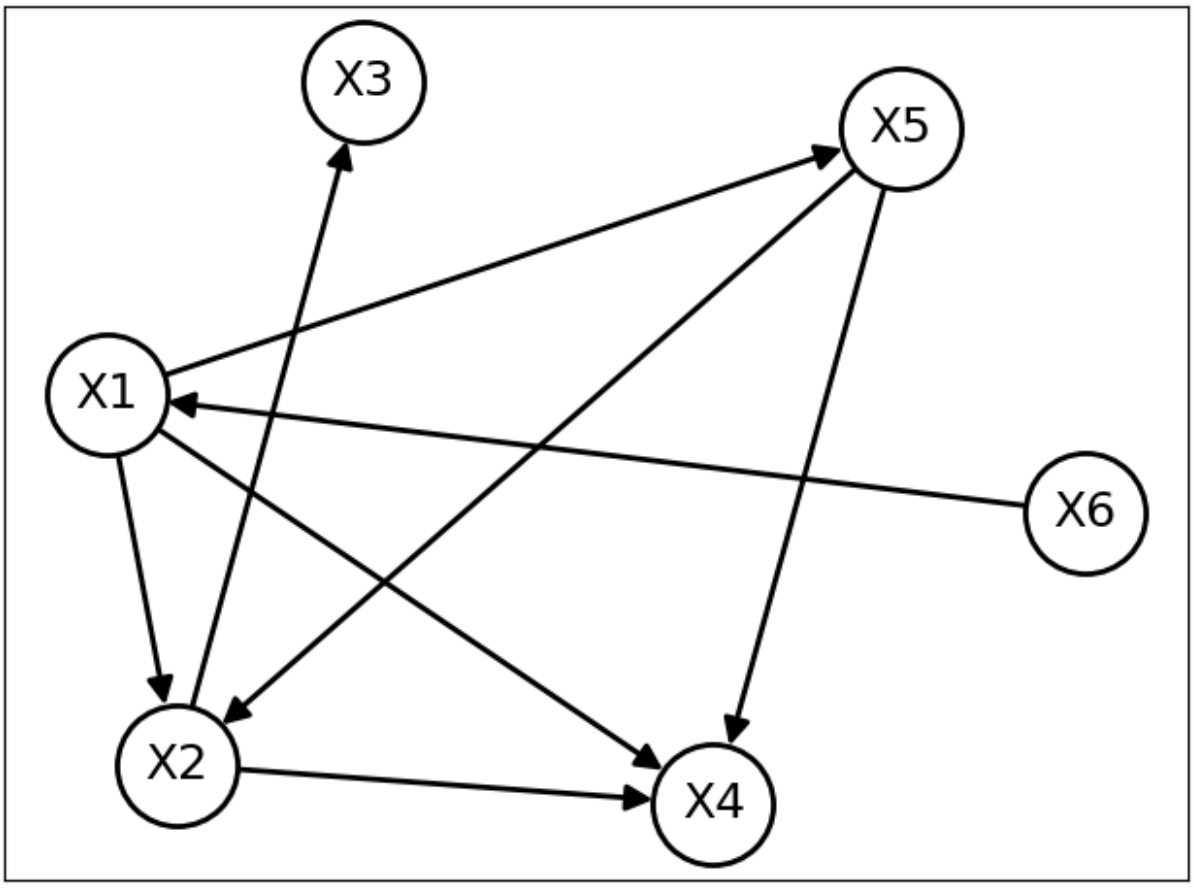
X5 is conditionally independent of X6 given {X1, X2, X3, X4}
Select one alternative:

- True
- False

Maximum marks: 2.5

5(g) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X3 is independent of X5

Select one alternative:

- True
- False

X3 is conditionally independent of X5 given X4

Select one alternative:

- True
- False

X5 is conditionally independent of X6 given {X1, X3}

Select one alternative:

- True
- False

X2 is conditionally independent of X6 given {X1, X4, X5}
Select one alternative:

- True
- False

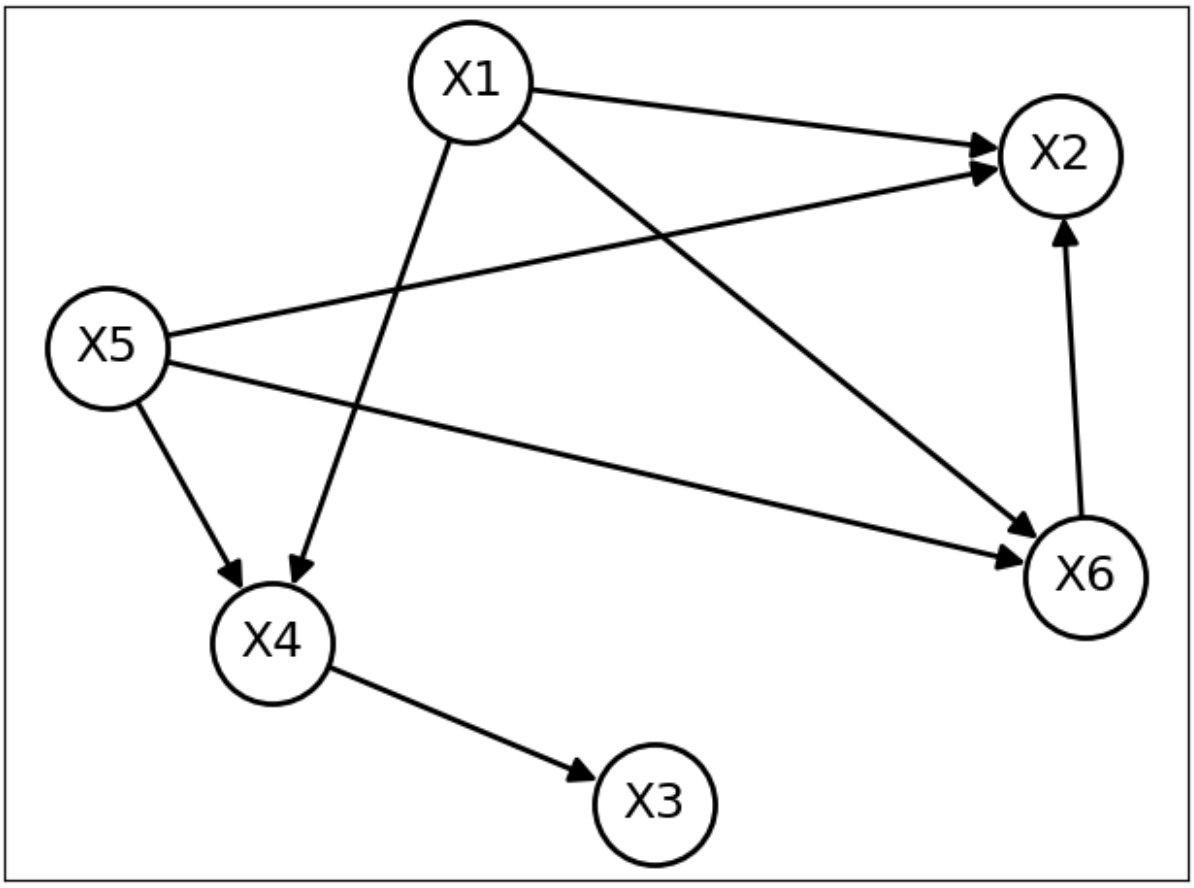
X2 is conditionally independent of X6 given {X1, X3, X4, X5}
Select one alternative:

- True
- False

Maximum marks: 2.5

5(h) Choose one alternative for each of the five following claims. You are given 0.5 point for for each correct answer, and 0 points for each wrong answer.

All statements relate to the following Bayesian network that consists of the 6 variables (X1, X2, X3, X4, X5 og X6):



X2 is independent of X3

Select one alternative:

- False
- True

X2 is conditionally independent of X4 given X5

Select one alternative:

- True
- False

X4 is conditionally independent of X6 given {X2, X3}

Select one alternative:

- False
- True

X4 is conditionally independent of X6 given {X1, X3, X5}
Select one alternative:

- True
- False

X3 is conditionally independent of X6 given {X1, X2, X4, X5}
Select one alternative:

- False
- True

Maximum marks: 2.5

6 Describe the four steps of the CBR-cycle using your own words.

Fill in your answer here

Maximum marks: 5

7 AI-technology finds new application areas all the time, for instance *autonomous weapon systems* and *advanced surveillance systems*. Discuss in light of these application areas if there should be international regulations against research on AI technology in general, against work on specific applications, or if it is better to let the AI research continue without any regulations.

Your answer can maximally be 1000 words long.

Fill in your answer here

Format

Words: 0/1000

Maximum marks: 10