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

Select one alternative:

- False ✔
- True

Maximum marks: 1

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

Select one alternative:

- True ✔
- False

Maximum marks: 1

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

Select one alternative:

- True ✔
- False

Maximum marks: 1

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

Select one alternative:

- Prediction ✔
- Smoothing
- Filtering

Maximum marks: 1

2(e) 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 $P(x_1, x_2, \ldots x_n) = P(x_1) \cdot P(x_2) \cdot P(x_3) \cdot \ \ldots \ \cdot P(x_n)$.

Select one alternative:

- False ✔
- True

Maximum marks: 1

2(f) 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(g) 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(h) 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:

- True
- False ✔

Maximum marks: 1

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

Select one alternative:

- False
- True ✔

Maximum marks: 1

2(j) 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:

- True
- False ✔

Maximum marks: 1

2(k) 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:

- True
- False ✔

Maximum marks: 1

2(l) 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:

- False ✔
- True

Maximum marks: 1

2(m) 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(n) 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:

- True ✔
- False

Maximum marks: 1

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

Select one alternative:

- Prediction
- Smoothing
- Filtering ✔

Maximum marks: 1

2(r) 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(s) If a system passes the Turing-test, it proves that the strong AI hypothesis is true

Select one alternative:

- False ✔
- True

Maximum marks: 1

2(t) 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(u) Among the types of inference that can be done with a *Hidden Markov Model*, we find *filtering* and *smoothing*.

Select one alternative:

- True ✔
- False

Maximum marks: 1

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

Select one alternative:

- True ✔
- False

Maximum marks: 1

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

Select one alternative:

- False ✔
- True

Maximum marks: 1

2(x) 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:

- False
- True ✔

Maximum marks: 1

2(y) 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:

- True ✔
- False

Maximum marks: 1

2(z) 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 ✔

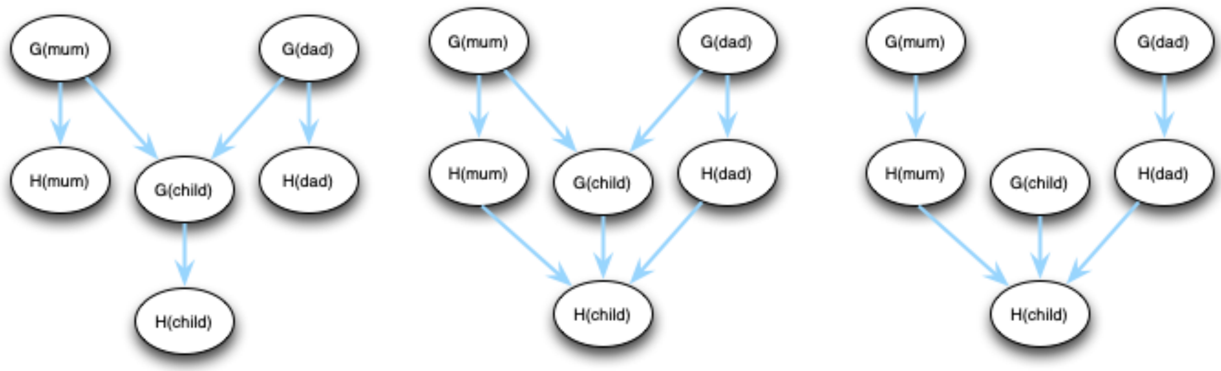
Maximum marks: 1

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.

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:



(1) (2) (3)

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

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

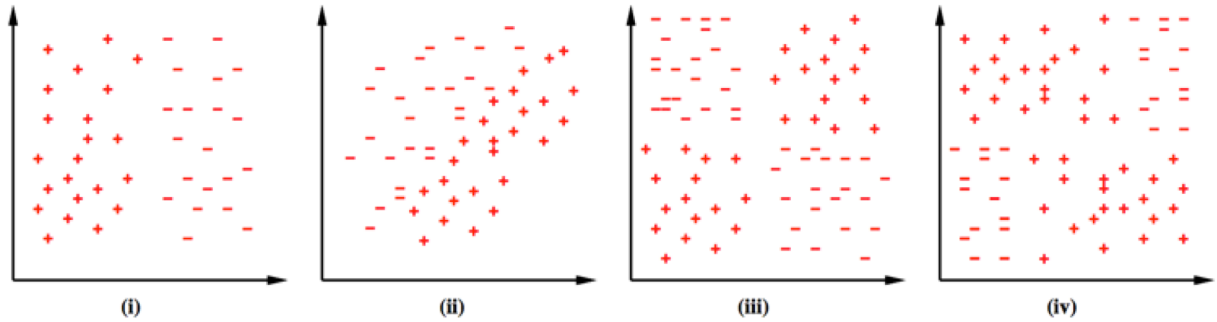
Select one or more alternatives:

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

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.**

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:

- Deep neural network using convolutions
- Perceptron

Dataset (ii):
Select one alternative:

- Decision-tree
- Perceptron

Dataset (iii):
Select one alternative:

- Perceptron
- Case-based reasoning

Dataset (iv):
Select one alternative:

- ○ Decision-tree ✔
- ○ Case-based reasoning

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 ***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 $\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2$, 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:

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

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).

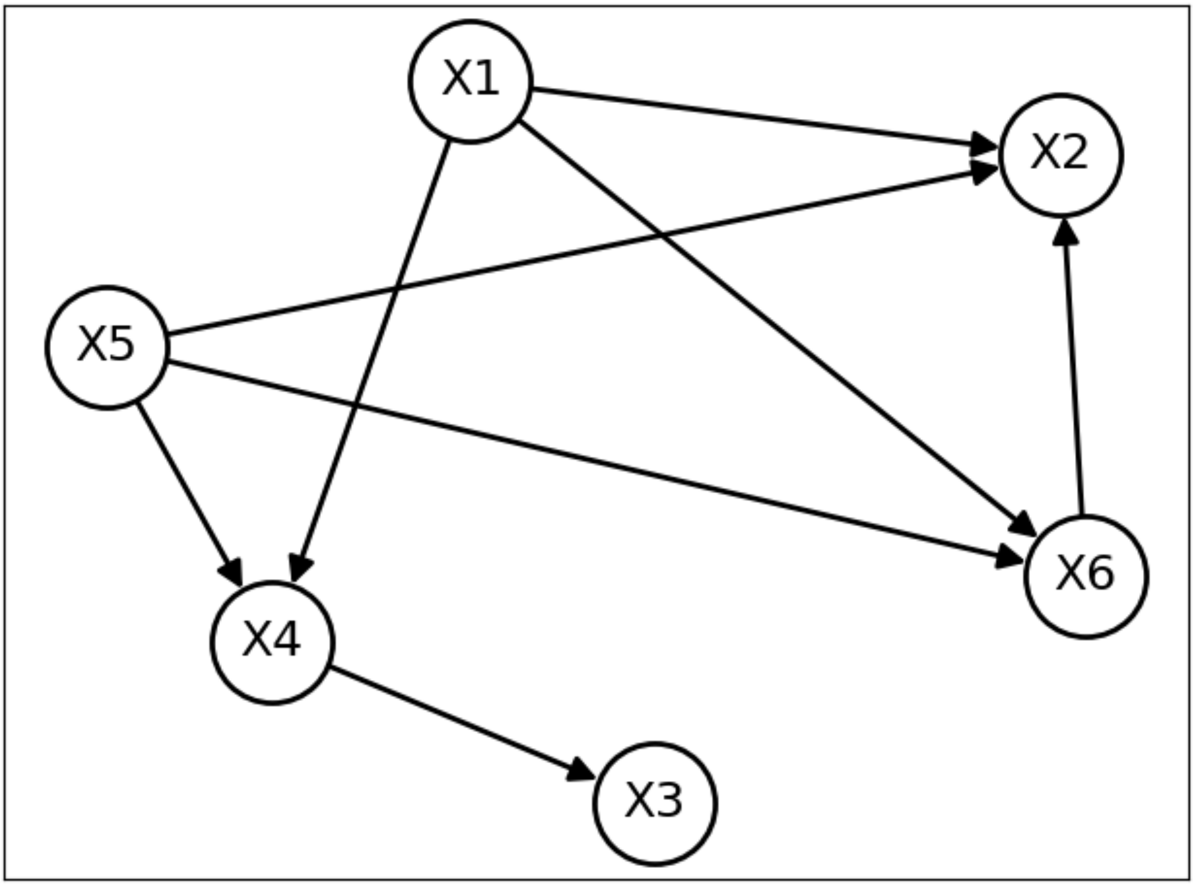
Update:

$w \leftarrow$ ☐ (1.4)

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):



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:

- False
- True ✔

X1 is conditionally independent of X3 given {X4, X5, X6}

Select one alternative:

- False
- True ✔

X3 is conditionally independent of X5 given {X1, X2, X4, X6}

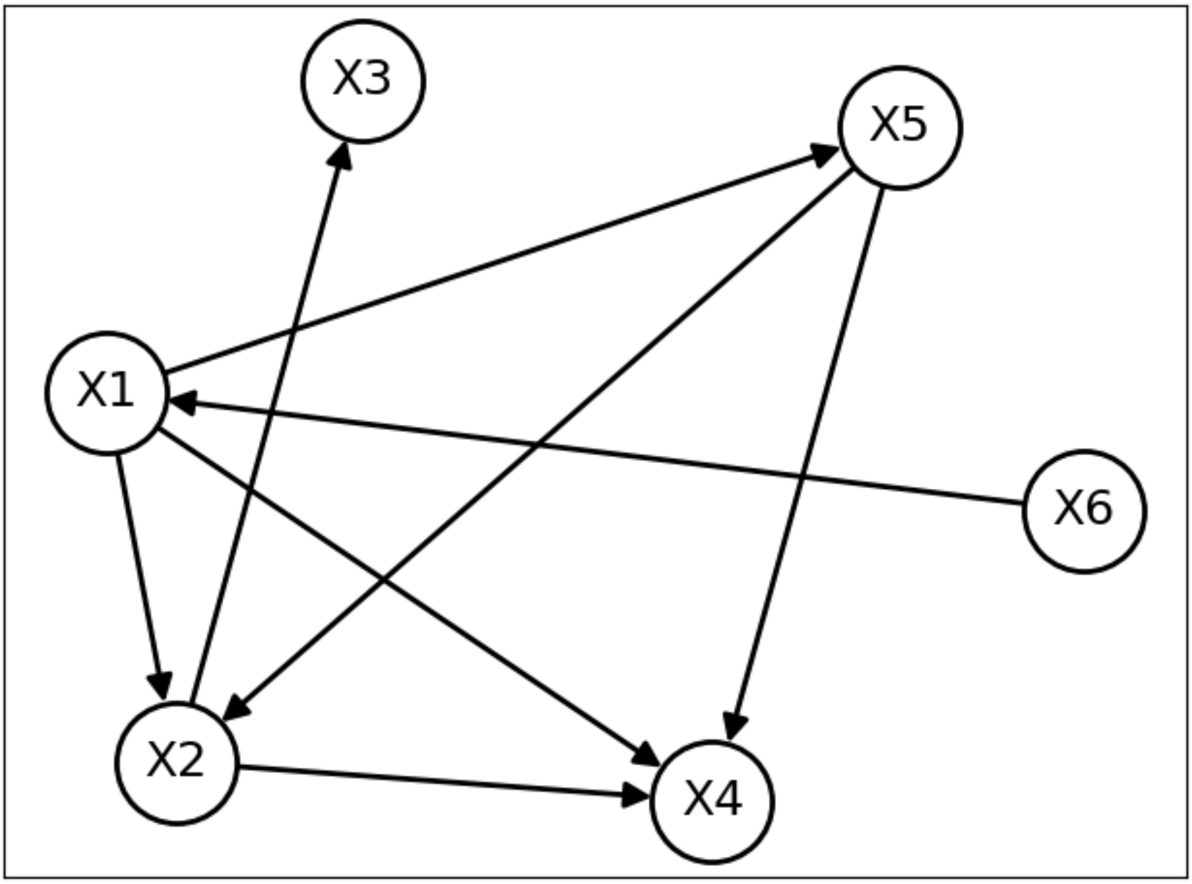
Select one alternative:

- 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):



X3 is independent of X6

Select one alternative:

- True
- False ✔

X1 is conditionally independent of X3 given X2

Select one alternative:

- True ✔
- False

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

Select one alternative:

- False ✔
- True

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

- False ✔
- True

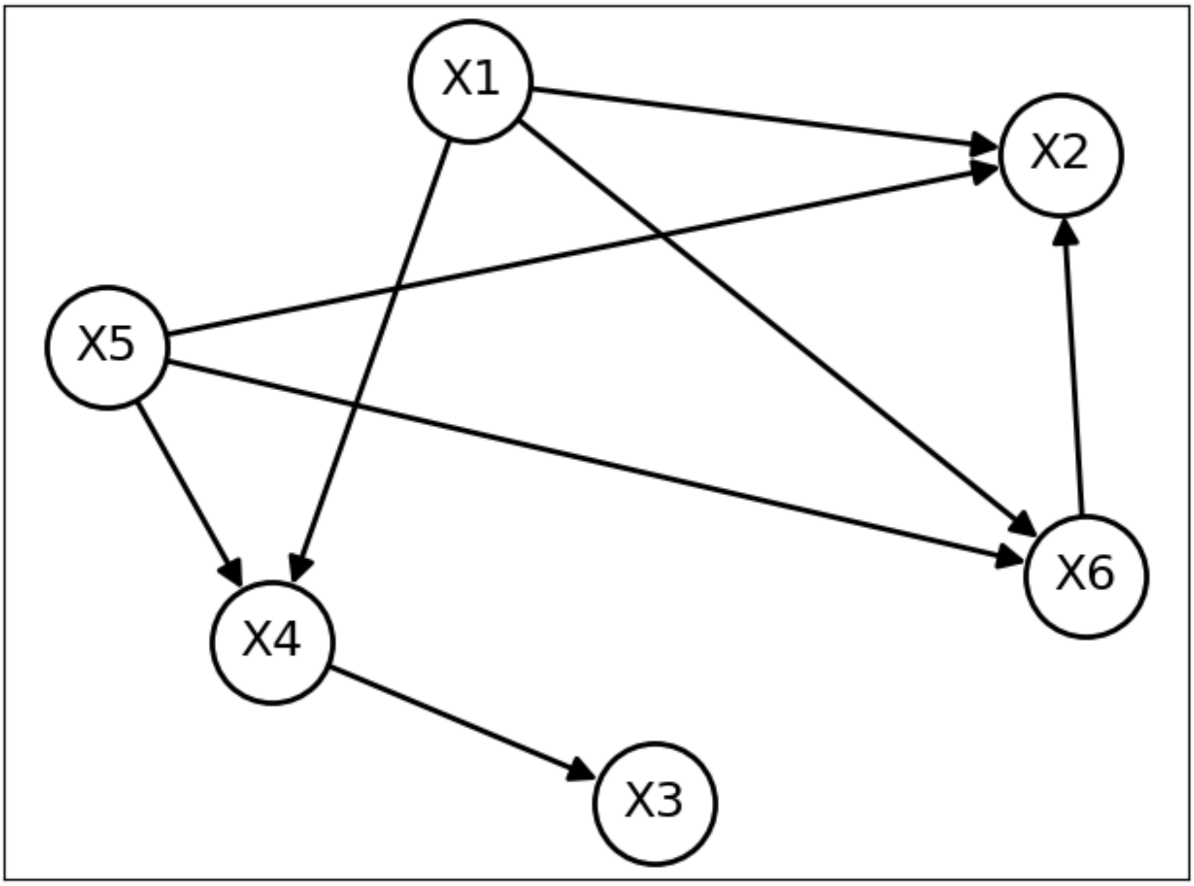
X1 is conditionally independent of X3 given {X2, X4, X5, X6}
Select one alternative:

- True ✔
- False

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:

- True ✔
- False

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

- True ✔
- False

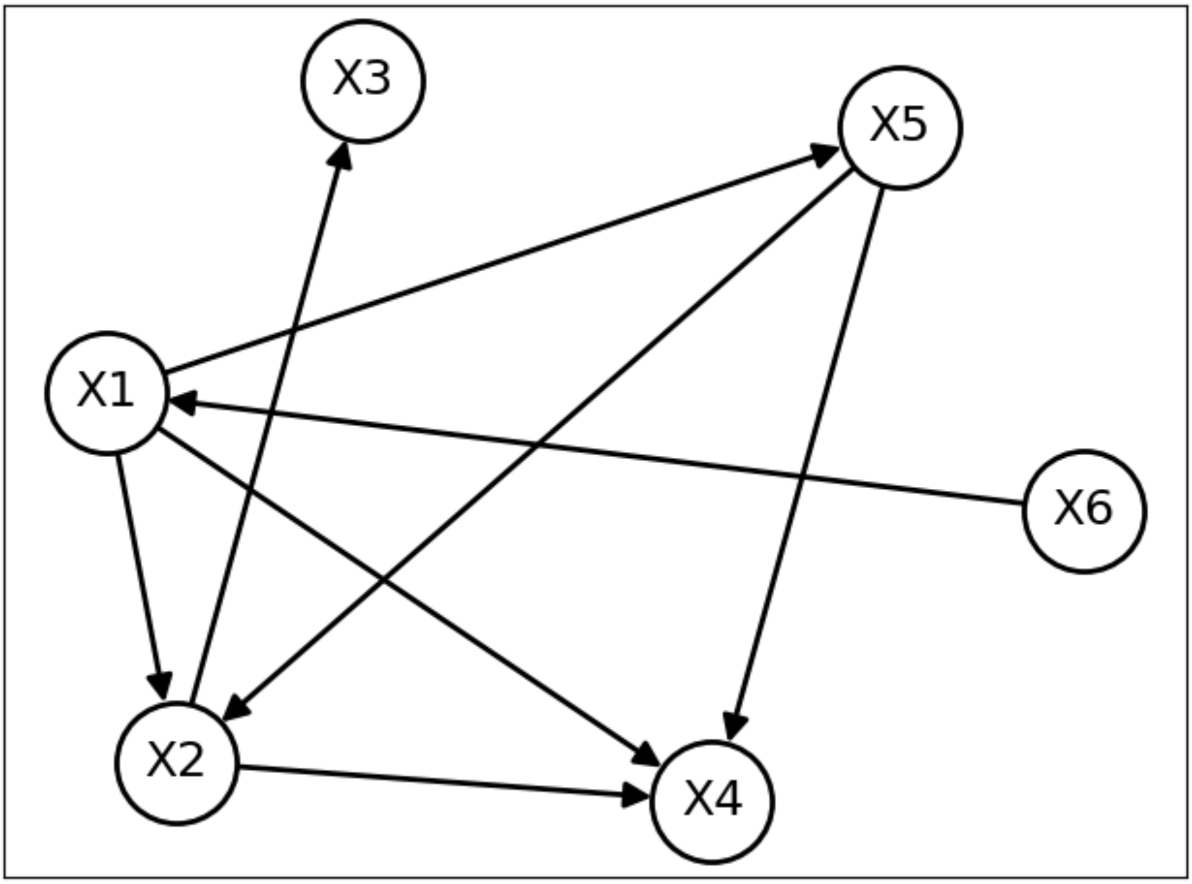
X3 is conditionally independent of X6 given {X1, X2, X4, X5}
Select one alternative:

- 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):



X4 is independent of X6

Select one alternative:

- False ✔
- True

X2 is conditionally independent of X6 given X4

Select one alternative:

- True
- False ✔

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

Select one alternative:

- False
- True ✔

X5 is conditionally independent of X6 given {X2, X3, X4}

Select one alternative:

- False ✔
- True

X5 is conditionally independent of X6 given {X1, X2, X3, X4}

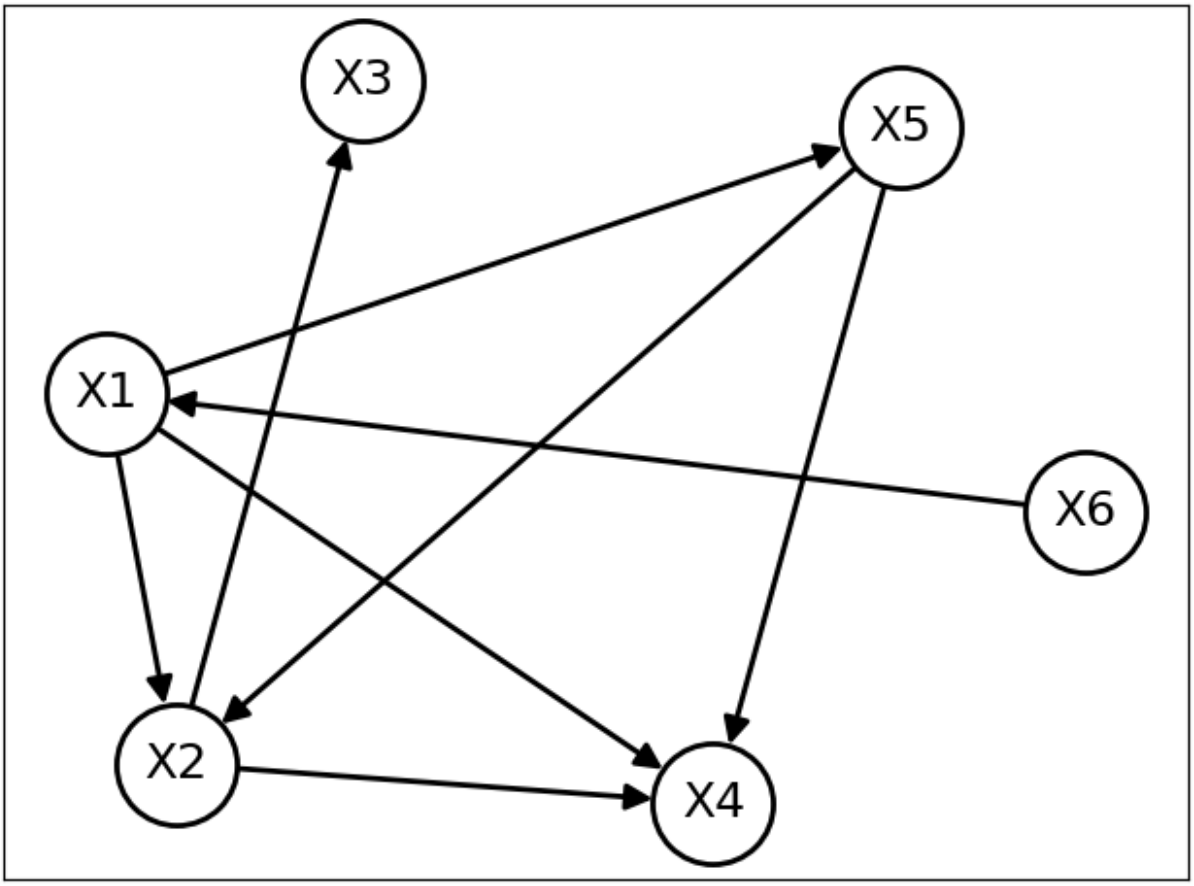
Select one alternative:

- 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:

- True
- False ✔

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:

- False
- True ✔

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}

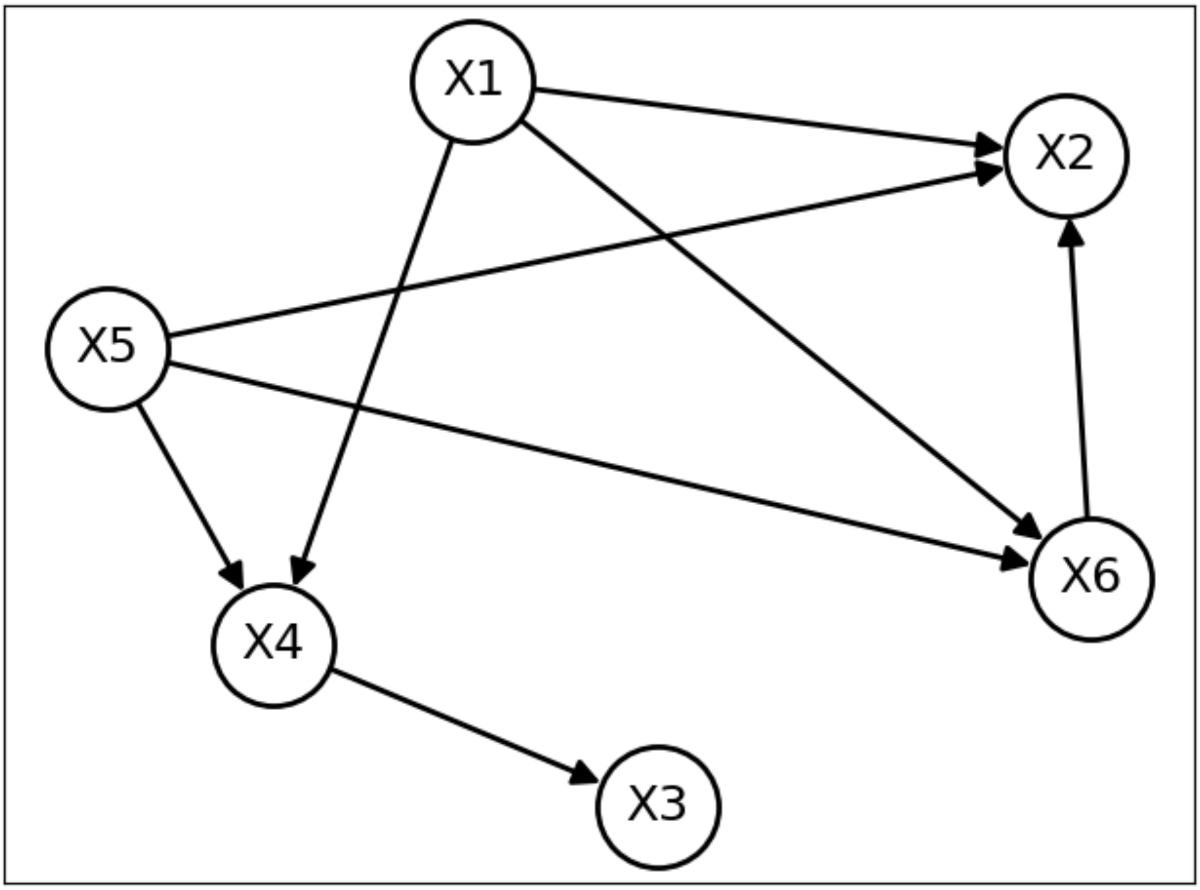
Select one alternative:

- False
- True ✔

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):



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:

- False
- True ✔

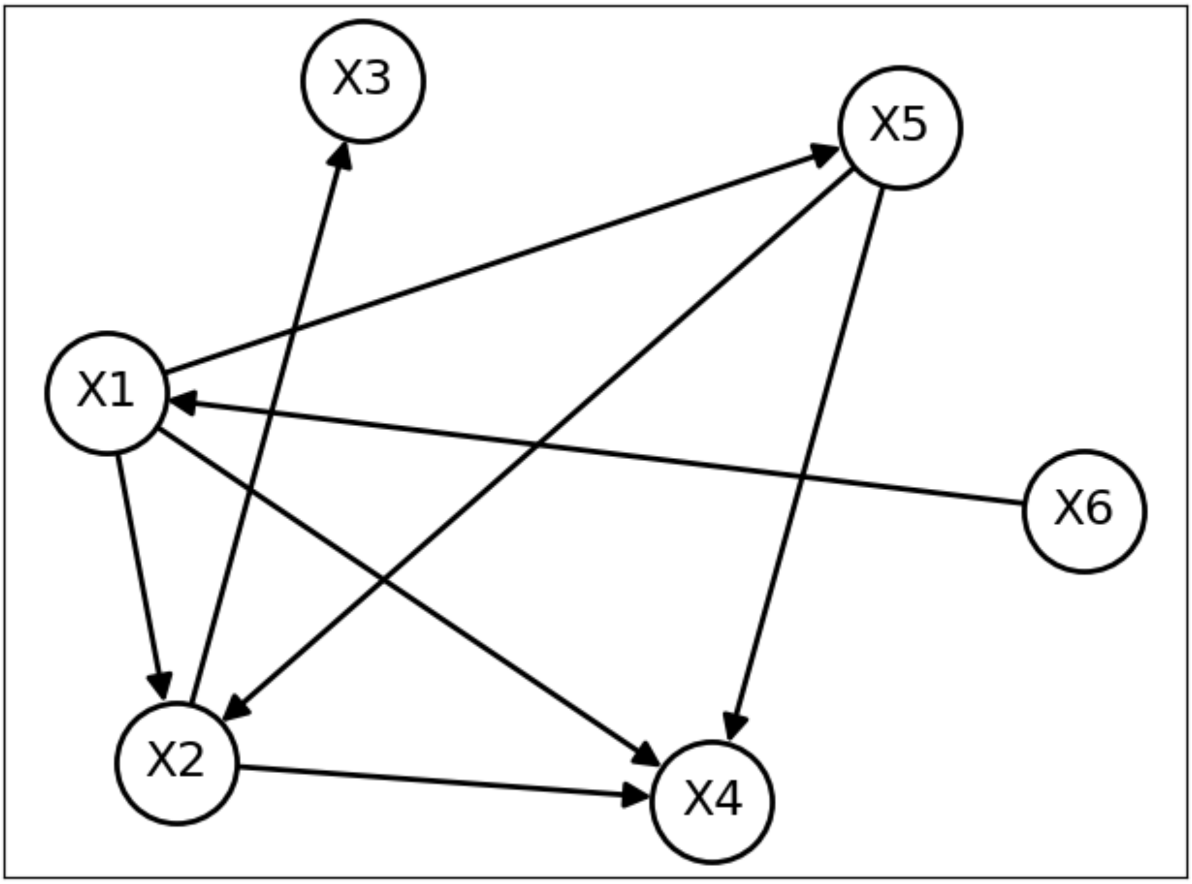
X3 is conditionally independent of X6 given {X1, X2, X4, X5}
Select one alternative:

- False
- True ✔

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:

- False ✔
- True

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}

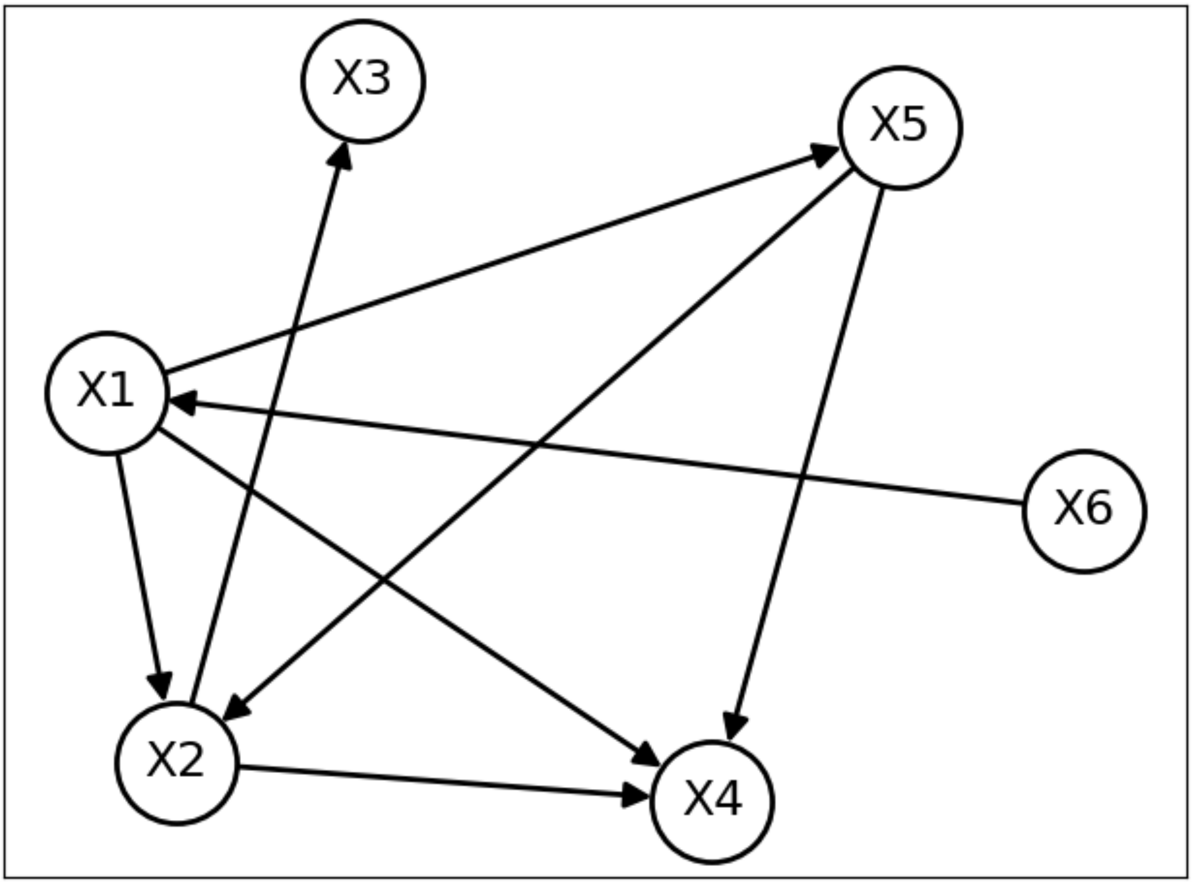
Select one alternative:

- False
- True ✔

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):



X3 is independent of X5
Select one alternative:

- False ✔
- True

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:

- True
- False ✔

X2 is conditionally independent of X6 given {X1, X3, X4}

Select one alternative:

- False
- True ✔

X3 is conditionally independent of X5 given {X1, X2, X4, X6}

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