Examination in Machine Intelligence

Thomas Dyhre Nielsen

January 21, 2019

On the next six pages you will find six questions covering different aspects of the course. The questions differ in their level of difficulty, and for each correctly answered question you will get a certain amount of points as indicated by each question. When solving the questions you are allowed to use all available material such as books, pocket calculator, etc., however, laptops/tablets and other networking devices are *not* allowed.

Before you answer a question make sure that you have read the question carefully. Moreover, make sure that you argue for your answers (e.g. include intermediate results) so that it is possible to follow your line of thought. Finally, it is important that your solutions are presented in a readable form. The answers to the questions should be written in English.

In addition to the six pages with questions, you are also provided with 10 pages that you can use when writing your answers to the questions.

- For each sheet of paper containing your response to the questions, please include your name, study number, current page number, and the total number of pages.
- If you need more paper, simply raise your hand to contact one of the guards in the examination room.

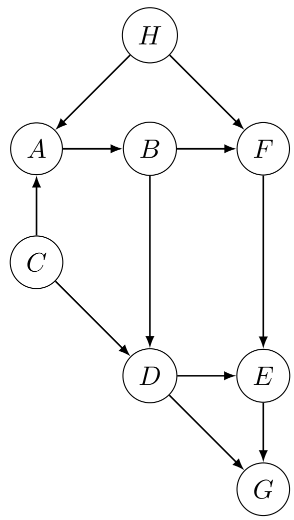
Lastly, please specify your name, study number, and study program here:

Full name	
Study number	
Study program	

Good luck with the questions
Thomas Dyhre Nielsen

Question 1 - 10 points

Consider the graph below:



1. List all probability distributions (on the form $P(X|Y_1, \ldots, Y_n)$) that should be specified in order to obtain a Bayesian network from the graph.

2. Assume that the variable D has three states labeled $\{a, b, c\}$ and that the remaining variables (i.e., $\{A, B, C, E, F, G, H\}$) all have two states labeled t and f. Give an example of a table (containing probability values) representing a valid conditional probability distribution for variable D.

3. Which variables are d-separated from A given hard evidence on B?

4. Which variables are d-separated from B given hard evidence on A?

5. Which variables are d-separated from F given hard evidence on A?

Solution:

1. $P(A|C, H)$, $P(B|A)$, $P(C)$, $P(D|B, C)$, $P(E|D, F)$, $P(F|B, H)$, $P(G|D, E)$, $P(H)$

2. An example of a valid conditional probability distribution could be

	$B = t$	$B = f$
$C = t$	(0.1, 0.3, 0.6)	(0.3, 0.3, 0.4)
$C = f$	(0.05, 0.85, 0.1)	(0.7, 0.1, 0.2)

where the 3-tuples define a probability distribution for each configuration of the parent variables B and C.

3. $\emptyset$
4. $\{C, H\}$
5. $\emptyset$

Question 2 - 20 points

Consider the variables A_1, A_2, A_3, and A_4, each of which is given the state space $\{1, 2, \ldots, 8\}$. The variables are related by the following constraints

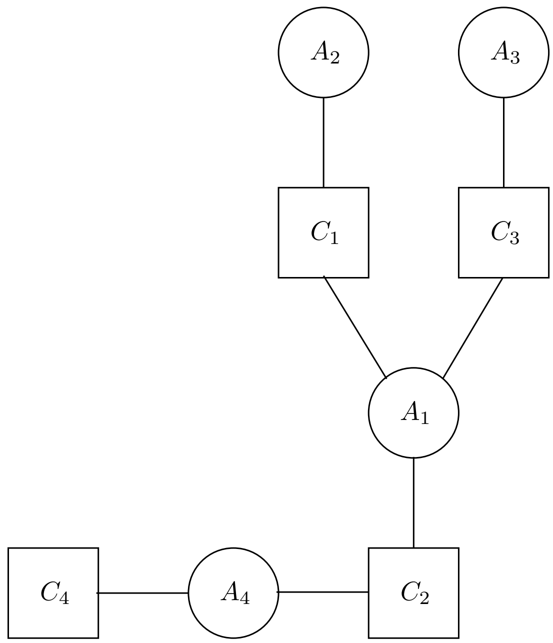
1. $A_1 > A_2$
2. $A_1 < A_4$
3. $A_3 = A_1 + 5$
4. $A_4 < 6$.

You should

- Give a constraint network representation of the domain above.
- Make the network arc-consistent.
- Find all satisfying solutions to the problem (if any exist) using variable elimination with the elimination ordering A_2, A_1, A_3, and A_4. Show the operations involved when eliminating each of the variables.
- Find another elimination ordering that is more efficient than the elimination ordering above. Show the operations involved when eliminating the variables according to your suggested elimination ordering. What are the satisfying solutions?

Solution:

Subproblem 1



Subproblem 2

- $sp(A_1) = \{2, 3\}$
- $sp(A_2) = \{1, 2\}$
- $sp(A_3) = \{7, 8\}$
- $sp(A_4) = \{3, 4, 5\}$

Subproblem 3

Denote the constraints in the problem formulation as C_1, C_2, C_3, and C_4, where the indexes correspond to the enumeration used in the problem formulation.

Eliminating A_2:

Only C_1 is involved in the elimination:

$$C_1 = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline 2 & 1 \\ 3 & 1 \\ 3 & 2 \end{array}\right)$$

Hence,

$$C_5 = C_1^{\downarrow\{A_1\}} = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline 2 & 1 \\ 3 & 1 \\ 3 & 2 \end{array}\right)^{\downarrow\{A_1\}} = \left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array}\right)$$

Eliminating A_1:

$$\begin{aligned}
C_6 &= (C_5 \bowtie C_3 \bowtie C_2)^{\downarrow\{A_3,A_4\}} \\
&= \left(\left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array}\right) \bowtie \left(\begin{array}{c|c} A_1 & A_3 \\ \hline 2 & 7 \\ 3 & 8 \end{array}\right) \bowtie \left(\begin{array}{c|c} A_1 & A_4 \\ \hline 2 & 3 \\ 2 & 4 \\ 2 & 5 \\ 3 & 4 \\ 3 & 5 \end{array}\right)\right)^{\downarrow\{A_3,A_4\}} \\
&= \left(\begin{array}{c|c|c} A_1 & A_3 & A_4 \\ \hline 2 & 7 & 3 \\ 2 & 7 & 4 \\ 2 & 7 & 5 \\ 3 & 8 & 4 \\ 3 & 8 & 5 \end{array}\right)^{\downarrow\{A_3,A_4\}} = \left(\begin{array}{c|c} A_3 & A_4 \\ \hline 7 & 3 \\ 7 & 4 \\ 7 & 5 \\ 8 & 4 \\ 8 & 5 \end{array}\right)
\end{aligned}$$

Eliminating A_3:

$$\begin{aligned}
C_7 = C_6^{\downarrow A_4} &= \left(\begin{array}{c|c} A_3 & A_4 \\ \hline 7 & 3 \\ 7 & 4 \\ 7 & 5 \\ 8 & 4 \\ 8 & 5 \end{array}\right)^{\downarrow A_4} \\
&= \left(\begin{array}{c} A_4 \\ \hline 3 \\ 4 \\ 5 \end{array}\right)
\end{aligned}$$

By backtracking we find the satisfying solutions to be

$$\left(\begin{array}{c|c|c|c} A_1 & A_2 & A_3 & A_4 \\ \hline 3 & 2 & 8 & 5 \\ 3 & 1 & 8 & 5 \\ 2 & 1 & 7 & 5 \\ 3 & 2 & 8 & 4 \\ 3 & 1 & 8 & 4 \\ 2 & 1 & 7 & 4 \\ 2 & 1 & 7 & 3 \end{array}\right)$$

Subproblem 4

A better elimination ordering is A_2, A_3, A_1, A_4.

Eliminating A_2:

As above, only C_1 is involved in the elimination:

$$C_1 = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline 2 & 1 \\ 3 & 1 \\ 3 & 2 \end{array}\right)$$

Hence,

$$C_5 = C_1^{\downarrow\{A_1\}} = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline 2 & 1 \\ 3 & 1 \\ 3 & 2 \end{array}\right)^{\downarrow\{A_1\}} = \left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array}\right)$$

Eliminating A_3:

$$C_6 = C_3^{\downarrow A_1} = \left(\begin{array}{c|c} A_1 & A_3 \\ \hline 2 & 7 \\ 3 & 8 \end{array}\right)^{\downarrow A_1} = \left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array}\right)$$

Eliminating A_1:

$$C_7 = (C_2 \bowtie C_5 \bowtie C_6)^{\downarrow A_4} = \left(\left(\begin{array}{c|c} A_1 & A_4 \\ \hline 2 & 3 \\ 2 & 4 \\ 2 & 5 \\ 3 & 4 \\ 3 & 5 \end{array} \right) \bowtie \left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array} \right) \bowtie \left(\begin{array}{c} A_1 \\ \hline 2 \\ 3 \end{array} \right) \right)^{\downarrow A_4}$$

$$= \left(\begin{array}{c} A_4 \\ \hline 3 \\ 4 \\ 5 \end{array} \right)$$

We end up with the same set of satisfying solutions as above.

Question 3 - 20 points

Consider the three variables A, B, and C with state spaces $sp(A) = \{\text{red}, \text{green}, \text{blue}\}$, $sp(B) = \{\text{pos}, \text{neg}\}$, and $sp(C) = \{\text{true}, \text{false}\}$. Assume that the following data set with 20 cases has been collected for the three variables.

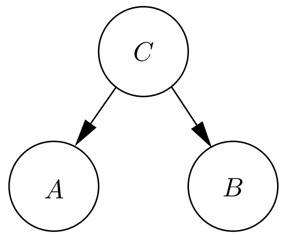
A	B	C
blue	neg	true
red	neg	false
blue	pos	true
blue	neg	false
red	neg	false
blue	neg	true
red	neg	false
red	neg	false
blue	pos	true
green	neg	false
blue	pos	true
blue	pos	true
red	neg	false
red	neg	false
red	pos	true
green	neg	false
green	pos	true
green	neg	true
red	neg	false
red	neg	false

You should:

1. Let C be the class variable. Show the structure of the naive Bayes classifier model for the three variables above.
2. Estimate the probabilities (empirical frequencies) for the naive Bayes model using the data set above.
3. Calculate the conditional probability distribution $P(C \mid A = \text{blue}, B = \textit{neg})$.
4. Use the naive Bayes model to classify a new data instance defined by $A = \text{blue}$ and $B = \textit{neg}$.
5. Calculate the conditional probability $P(B|A)$.
6. Describe at least one other machine learning task that one might perform on the data set above besides doing classification wrt. C.

Solution:

Subproblem 1



Subproblem 2

$$P(C) = (9/20_{true}, 11/20_{false})$$

$$P(A|C) = \begin{array}{cc|cc} & & \multicolumn{2}{c}{C} \\ & & true & false \\ \hline & \text{red} & 1/9 & 8/11 \\ A & \text{green} & 2/9 & 2/11 \\ & \text{blue} & 6/9 & 1/11 \end{array}$$

$$P(B|C) = \begin{array}{cc|cc} & & \multicolumn{2}{c}{C} \\ & & true & false \\ \hline B & pos & 6/9 & 0/11 \\ & neg & 3/9 & 11/11 \end{array}$$

Subproblem 3

$$P(C \mid \text{blue}, neg) = \frac{P(C, \text{blue}, neg)}{P(\text{blue}, neg)} = \frac{P(C, \text{blue}, neg)}{\sum_C P(C, \text{blue}, neg)}$$

By inserting the values from the conditional probability tables above, we find:

$$P(C, \text{blue}, neg) = (9/20 \cdot 6/9 \cdot 3/9, 11/20 \cdot 1/11 \cdot 11/11) = (0.1, 0.05)$$
$$P(\text{blue}, neg) = 0.15.$$

Hence,

$$P(C \mid \text{blue}, neg) = \frac{(0.1, 0.05)}{0.15} = \left(\frac{2}{3}, \frac{1}{3}\right).$$

Subproblem 4

According to $P(C \mid \text{blue}, neg)$ we classify the instance as *true*.

Subproblem 5

$$P(B \mid A) = \frac{\sum_C P(C)P(A \mid C)P(B \mid C)}{P(A)} = \sum_C P(C \mid A)P(B \mid C)$$

We find that

$$P(A, C) = \begin{array}{cc|cc} & & \multicolumn{2}{c}{C} \\ & & \textit{true} & \textit{false} \\ \hline A & \text{red} & 1/20 & 8/20 \\ & \text{green} & 2/20 & 2/20 \\ & \text{blue} & 6/20 & 1/20 \end{array}$$

and

$$P(A) = \sum_C P(A, C) = (9/20, 4/20, 7/20) = (0.45, 0.2, 0.35).$$

Thus,

$$P(C \mid A) = \begin{array}{cc|cc} & & \multicolumn{2}{c}{C} \\ & & \textit{true} & \textit{false} \\ \hline A & \text{red} & 1/9 & 8/9 \\ & \text{green} & 2/4 & 2/4 \\ & \text{blue} & 6/7 & 1/7 \end{array}$$

Combining with $P(B \mid C)$:

$$P(C, B \mid A) = \begin{array}{cc|cc|cc|cc} & & \multicolumn{6}{c}{A} \\ & & \multicolumn{2}{c|}{\text{red}} & \multicolumn{2}{c|}{\text{green}} & \multicolumn{2}{c}{\text{blue}} \\ \hline B & \textit{pos} & 1/81 & 0 & 12/36 & 0 & 36/63 & 0 \\ & \textit{neg} & 3/81 & 88/99 & 6/36 & 22/44 & 18/63 & 11/77 \end{array}$$

and by marginalizing out C we get

$$P(B \mid A) = \begin{array}{cc|ccc} & & \multicolumn{3}{c}{A} \\ & & \text{red} & \text{green} & \text{blue} \\ \hline B & \textit{pos} & 6/81 & 12/36 & 36/63 \\ & \textit{neg} & 75/81 & 24/36 & 27/63 \end{array} = \begin{array}{cc|ccc} & & \multicolumn{3}{c}{A} \\ & & \text{red} & \text{green} & \text{blue} \\ \hline B & \textit{pos} & 0.074 & 0.33 & 0.571 \\ & \textit{neg} & 0.926 & 0.67 & 0.429 \end{array}.$$

Question 4 - 15 points

Consider the influence diagram defined by the graphical structure below together with the following conditional probability and utility tables:

A	
a_1	a_2
0.1	0.9

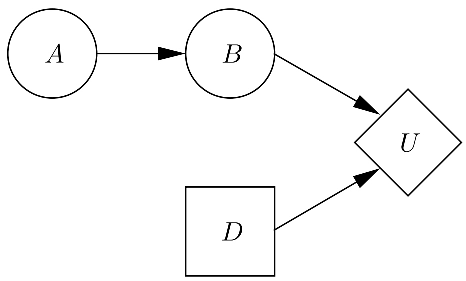
$P(A)$

		B		
		b_1	b_2	b_3
A	a_1	0.7	0.2	0.1
	a_2	0.1	0.4	0.5

$P(B|A)$

		B		
		b_1	b_2	b_3
D	d_1	100	10	0
	d_2	50	100	50
	d_3	0	10	100

$U(B, D)$



1. According to the influence diagram, specify the temporal order in which a decision maker observes and decides upon the variables of the model.
2. Calculate the expected utility of each of the three decision options for D. What is the best decision option according to your calculations?
3. Assume that A can be observed before deciding on D. Perform a value of information analysis to determine the maximum amount you should pay for observing A prior to deciding on D.

Solution:

Sub-problem 1

The sequence in which the variables are observed and decided upon is: $D \prec \{A, B\}$.

Sub-problem 2

First we calculate $P(B)$:

$$P(B) = \sum_A P(A)P(B|A)$$

$$= \sum_A \left(\begin{array}{|cc|} \hline \multicolumn{2}{|c|}{\text{A}} \\ a_1 & a_2 \\ \hline 0.1 & 0.9 \\ \hline \end{array} \cdot \begin{array}{|cc|ccc|} \hline & & \multicolumn{3}{c|}{\text{B}} \\ & & b_1 & b_2 & b_3 \\ \hline \text{A} & a_1 & 0.7 & 0.2 & 0.1 \\ & a_2 & 0.1 & 0.4 & 0.5 \\ \hline \end{array} \right)$$

$$= \begin{array}{|ccc|} \hline \multicolumn{3}{|c|}{\text{B}} \\ b_1 & b_2 & b_3 \\ \hline 0.16 & 0.38 & 0.46 \\ \hline \end{array}$$

Next we take the expectations:

$$EU(D) = \sum_B P(B)U(B, D)$$

$$= \sum_B \left(\begin{array}{|ccc|} \hline \multicolumn{3}{|c|}{\text{B}} \\ b_1 & b_2 & b_3 \\ \hline 0.16 & 0.38 & 0.46 \\ \hline \end{array} \cdot \begin{array}{|cc|ccc|} \hline & & \multicolumn{3}{c|}{\text{B}} \\ & & b_1 & b_2 & b_3 \\ \hline & d_1 & 100 & 10 & 0 \\ \text{D} & d_2 & 50 & 100 & 50 \\ & d_3 & 0 & 10 & 100 \\ \hline \end{array} \right)$$

$$= (19.8, \mathbf{69}, 49.8)$$

Sub-problem 2

For performing the VoI analysis we start by calculating the expected utilities conditioned on the two possible outcomes of A:

$$EU(D|A = a_i) = \sum_B P(B|A = a_i)U(B, D)$$

Thus:

$$EU(D|A=a_1) = \sum_B \left(\begin{array}{|c|ccc|} \hline & & \text{B} & \\ & b_1 & b_2 & b_3 \\ \hline A=a_1 & 0.7 & 0.2 & 0.1 \\ \hline \end{array} \quad \begin{array}{|cc|ccc|} \hline & & & \text{B} & \\ & & b_1 & b_2 & b_3 \\ \hline & d_1 & 100 & 10 & 0 \\ \text{D} & d_2 & 50 & 100 & 50 \\ & d_3 & 0 & 10 & 100 \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|} \hline D=d_1 & \mathbf{72} \\ D=d_2 & 60 \\ D=d_3 & 12 \\ \hline \end{array}$$

$$EU(D|A=a_1) = \sum_B \left(\begin{array}{|c|ccc|} \hline & & \text{B} & \\ & b_1 & b_2 & b_3 \\ \hline A=a_2 & 0.1 & 0.4 & 0.5 \\ \hline \end{array} \quad \begin{array}{|cc|ccc|} \hline & & & \text{B} & \\ & & b_1 & b_2 & b_3 \\ \hline & d_1 & 100 & 10 & 0 \\ \text{D} & d_2 & 50 & 100 & 50 \\ & d_3 & 0 & 10 & 100 \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|} \hline D=d_1 & 14 \\ D=d_2 & \mathbf{70} \\ D=d_3 & 54 \\ \hline \end{array}$$

Thus, the difference in expected utilities is

$$VOI = (72 \cdot 0.1 + 70 \cdot 0.9) - 69 = 71 - 69 = 1.2,$$

and you should therefore pay at most 1.2 for observing A before deciding on D.

Question 5 - 20 points

Consider the following six data points living in $\mathbb{R}^2$:

d_1	$(4, 12)$
d_2	$(3, 51)$
d_3	$(9, 80)$
d_4	$(11, 90)$
d_5	$(8, 100)$
d_6	$(3, 20)$

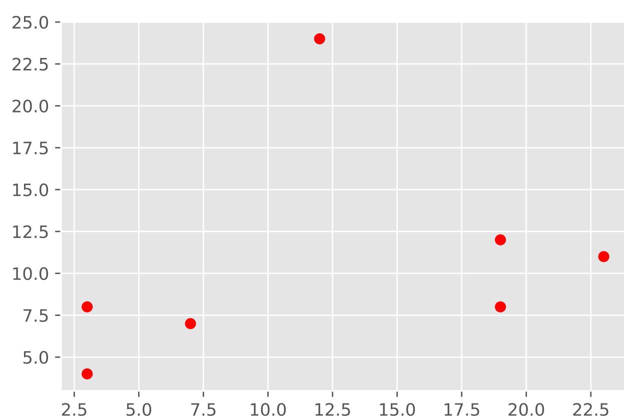
Using the Euclidean distance to measure distances:

1. Calculate the distances between data point d_2 and the other data points. Which data point is closest to d_2?
2. Let the data points $(3, 51)$ and $(9, 80)$ be initial cluster centers for the k-means algorithm. Perform one more k-means iteration by updating these cluster centers using the data set above.
3. Do you see any potential problems in doing k-means clustering and directly using the Euclidean distance with the data set above? What could you do to mitigate such problems?

Solution:

Sub-problem 1

The distances to the other points are $(39.01_{d_1}, 29.61_{d_3}, 39.81_{d_4}, 49.25_{d_5}, 31.00_{d_6})$, hence d_3 is closest.



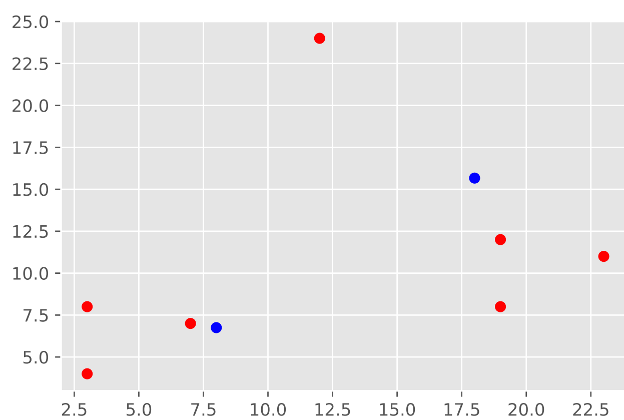
Sub-problem 2

First we find the points that belong to the two clusters defined by the initially chosen cluster centers:

dist	d_1	d_2	d_3	d_4	d_5	d_6
d_2	39.01	0.0	29.61	39.81	49.25	31.00
d_3	68.18	29.61	0.0	10.20	20.02	60.30

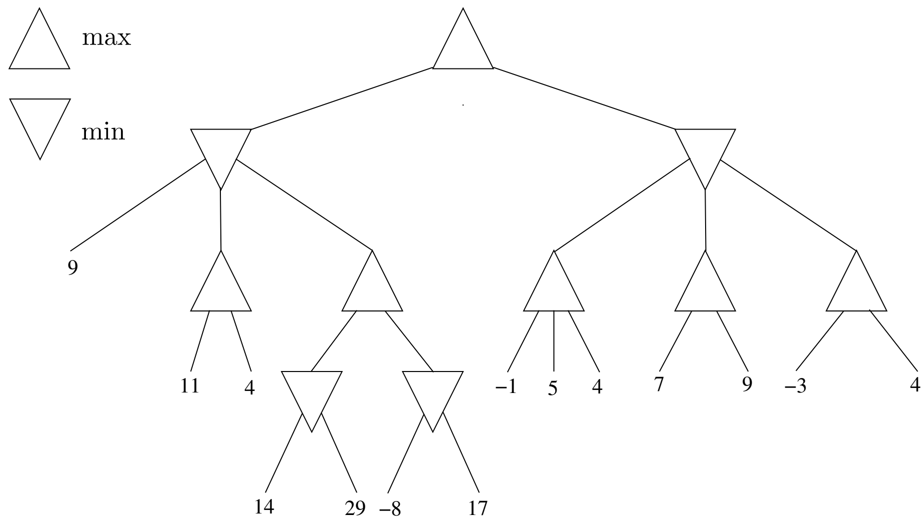
Thus, the data points d_1 and d_6 move to the cluster defined by d_2, whereas d_4 and d_5 go to the cluster defined by d_3.

The cluster center defined by $\{d_1, d_2, d_6\}$ becomes $(3.33, 27.67)$ whereas the cluster center defined by $\{d_3, d_4, d_5\}$ is $(9.33, 90.0)$.



Question 6 - 15 points

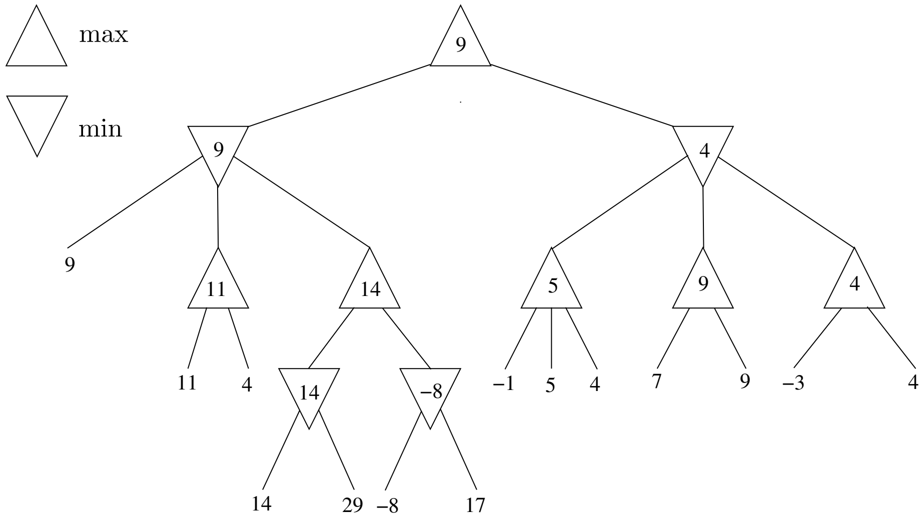
Consider the following zero-sum game tree:



1. Compute the utility values for all the nodes.
2. Assume that the utility values are calculated in a depth-first order that always considers the branches in a left to right order. Mark the nodes that will not be visited when employing $(\alpha - \beta)$ pruning.

Solution:

Sub-problem 1



Sub-problem 2