Intelligent Interaction Design [AI&CS] 2019-2020
Test January 9th - from Remindo <u>WITH ANSWERS</u>

1) As you may know, a rational agent interacts with its environment and the more complex the environment is, the harder it is to act rationally. Which of the following environments is the most complex one?

- Partially observable, deterministic, sequential, dynamic, discrete and multi-agent.
- Fully observable, stochastic, episodic, dynamic, discrete and single agent.
- Fully observable, stochastic, episodic, dynamic, discrete and multi-agent.
- Partially observable, stochastic, sequential, dynamic, continuous and multi-agent.

2) Consider the general ***A**** Tree Search algorithm as described in the course book:

(i) ***A**** Tree search can be implemented on a Simple Reflex Agent.

(ii) ***A**** Tree search cannot be implemented on a Model-based Reflex Agent.

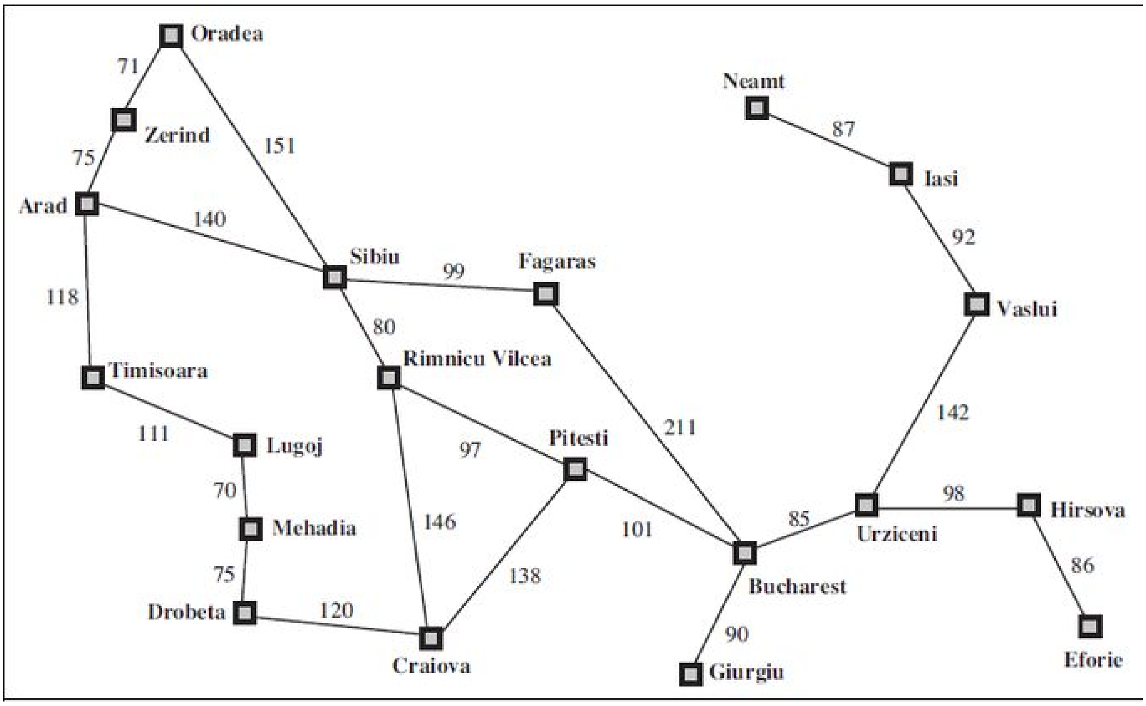
(iii) ***A**** Tree search cannot be implemented on a Goal Based Agent

- All statements (i), (ii) and (iii) are true.
- Only statement (ii) is true.
- Only statement (i) and (iii) are true.
- All statements (i), (ii) and (iii) are false.

3) Which of the following search strategies belongs tot the class of informed search?

- Depth-first Search
- Depth-limited Search
- Uniform-cost Search
- None of the above

4) Consider the simplified roadmap of Romania. The number above a road gives the road distance between the connecting cities.



Assume that a person wants to drive from Zerind to Bucharest and he/she applies a graph search algorithm to find the shortest route, with the following straight-line heuristic function:

Arad	366	Mehadia	241
Bucharest	0	Neamt	234
Craiova	160	Oradea	380
Drobeta	242	Pitesti	100
Eforie	161	Rimnicu Vilcea	193
Fagaras	176	Sibiu	253
Giurgiu	77	Timisoara	329
Hirsova	151	Urziceni	80
Iasi	226	Vaslui	199
Lugoj	244	Zerind	374

Of course the node corresponding to Zerind will be the first node expanded by this search algorithm. But what will be the third node expanded?

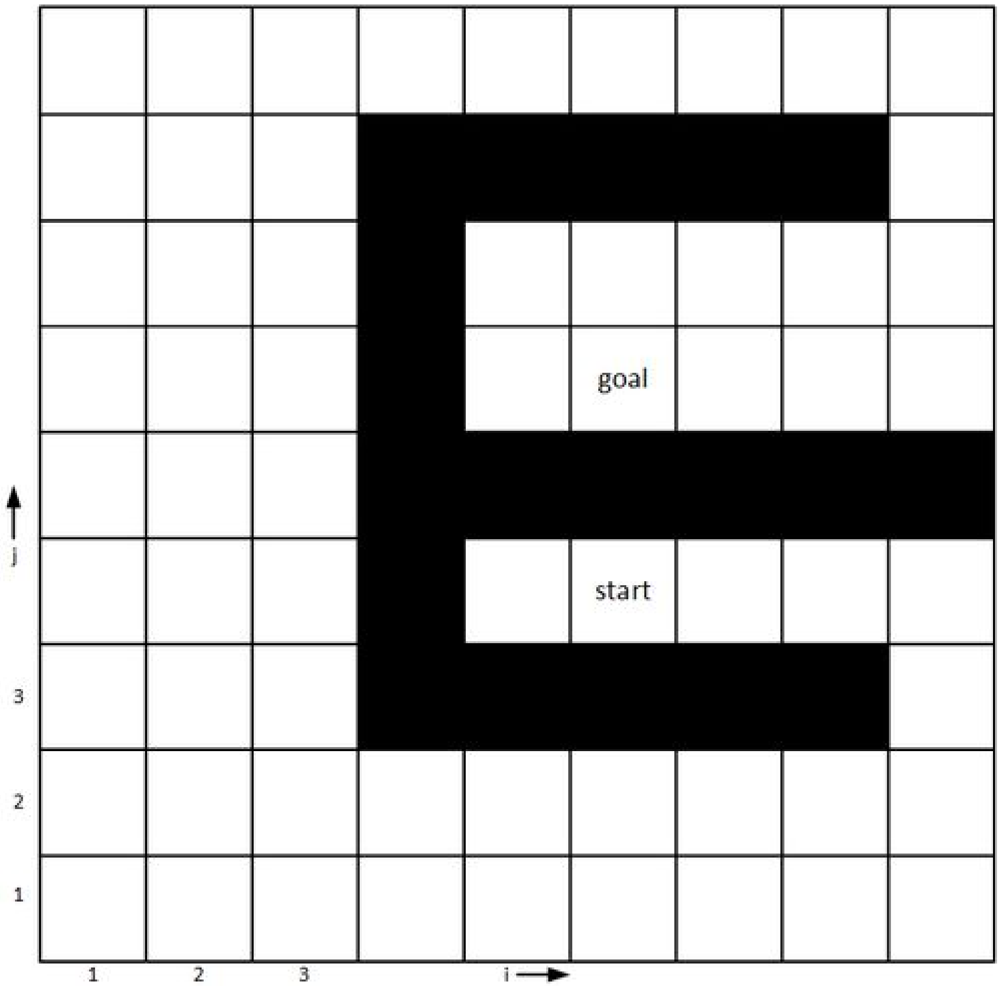
- The node corresponding to Timisoara.
- The node corresponding to Arad
- The node corresponding to Sibiu.
- None of the above.

5) Consider the following path finding problem in which an agent wants to go from the start cell (6,4) (denoted by start, first index is horizontal, second index vertical) to the goal cell (6,6). The agent can only make the following moves:

(i) The horizontal and vertical moves *one cell up, down, left,* or *right* and each theses moves have a cost of 1.

(ii) The diagonal moves *left diagonal up, left diagonal down, right diagonal up* and *right diagonal down.* For instance if the agent is in cell (3,2) and does the action *left diagonal up* then it will end up in cell (2,3). Each of these moves have a cost of 2.

The black cells form a barrier which the agent cannot pass. Assume the agent applies ***A* Graph Search*** with heuristic function ***h*** the Manhattan distance.



Which of the following nodes will **never** be in the list of open nodes?

- node corresponding to cell (state) (1,8)
- node corresponding to cell (state) (5,4)
- node corresponding to cell (state) (1,1)
- None of the above

Consider the above grid based search problem with heuristic function ***h*** the Manhattan distance and the following two statements about ***h***:

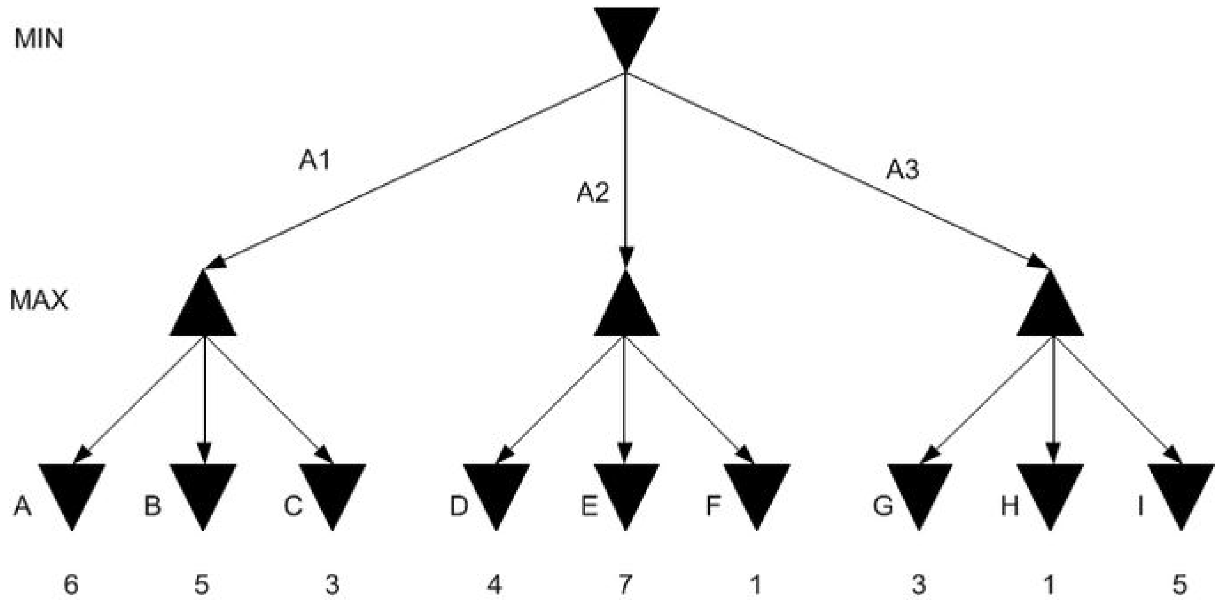
(i) ***h*** is admissible.

(ii) ***h*** is consistent.

Which of the above statements are true?

- Both statements (i) and (ii) are true
- Both statements (i) and (ii) are false.
- Only statement (ii) is true.
- Only statement (i) is true.

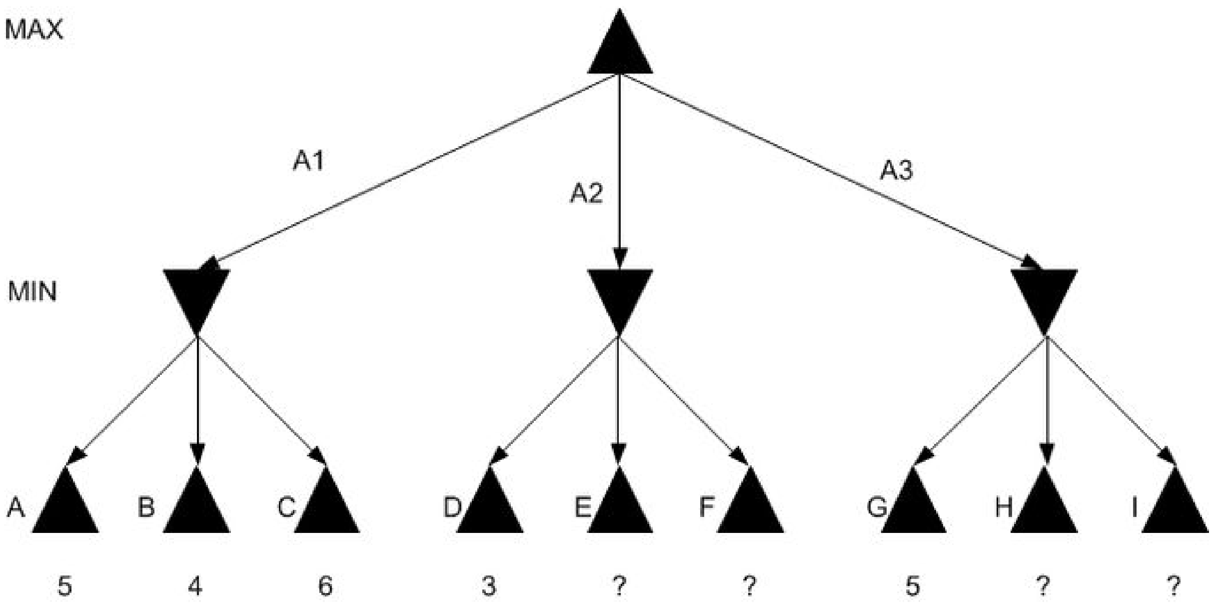
6) Consider the following part of a two-player game tree.



What will be the value of the top node?

- 5
- 4
- 3
- 1

7) Consider the following two-player game tree. A ? below a node means that the node is not explored yet, so value is unknown.



Assume one applies alpha-beta pruning and consider the following statements:

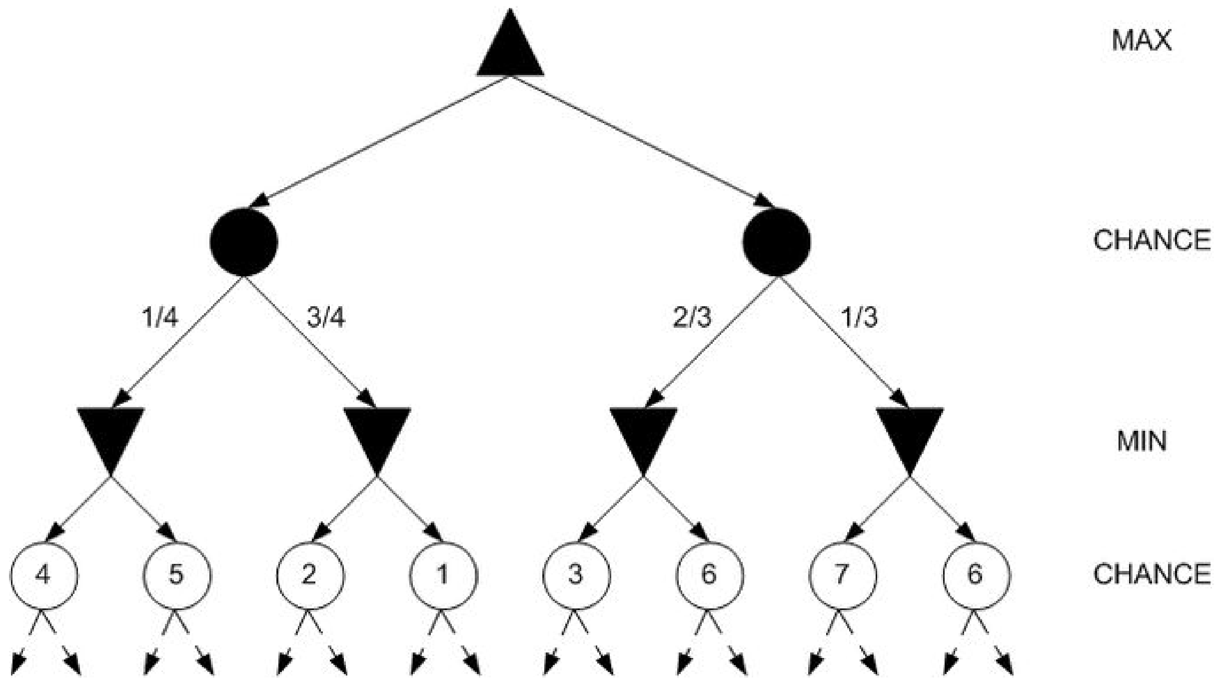
(i) Alpha-beta pruning will **not** calculate the value of node E.

(ii) Alpha-beta pruning will calculate the value of node H.

Which of the above statements are true?

- Both statements (i) and (ii) are true
- Both statements (i) and (ii) are false.
- Only statement (ii) is true.
- Only statement (i) is true.

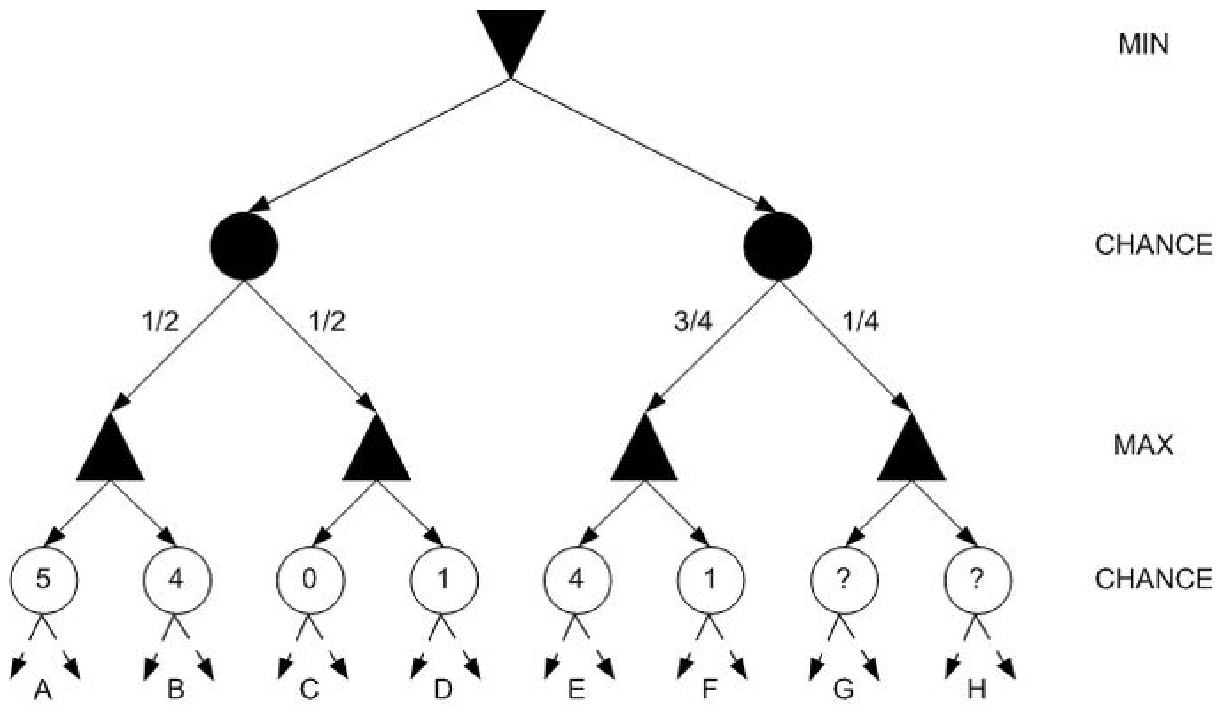
8) Consider the following two-player game tree in which the game has an element of chance, which is shown by the so-called probability (chance) nodes in the game tree.



What is the correct value for the top node if one applies the expectiminimax algorithm?

- 4
- 7
- 5
- 6

9) Consider the following game tree with an element of chance.



The letters under the bottom row of chance nodes are labels for the nodes just above the letter. One can also apply (alpha-beta) pruning on this game tree. The numbers inside the chance nodes on the bottom row are the computed values of these chance nodes indicates not computed yet. Which of the following statements is true?

- If the value of node G is larger or equal than 0 then pruning will never calculate the value of node H.
- If the value of node G is smaller than 0 then pruning will always calculate the value of node H.
- If the value of node G is larger or equal than 0 then pruning will always always calculate the value of node H.
- None of the above.

10) Consider the proposition:

$$(\neg A \Rightarrow B) \Rightarrow \neg C$$

How many models are there for this proposition?

- 3
- 5
- 1
- 7

11) We are given the following premisses:

(i) $P \vee Q$

(ii) $P \Rightarrow (T \vee R)$

(iii) $\neg S \vee \neg P$

(iv) $\neg Q$

The question is whether we can conclude (i.e. prove) ***T*** from these premisses. Which of the following statements is correct?

- No, the conclusion T does not follow, but if you add the premiss $\neg S$ the conclusion can be derived.
- No, the conclusion T does not follow, but if you add the premiss $\neg R$ the conclusion can be derived.
- Yes, the conclusion T follows, but we can't prove $\neg T$.
- Yes, we can conclude (i.e. prove) T because we can prove everything: the premisses are inconsistent.

12) Transforming the logical formula

$$[\neg(W \wedge R) \Rightarrow (S \vee T)]$$

into conjunctive normal will result in?

- $\neg(W \wedge R) \vee (S \vee R)$
- None of the above
- $(W \vee S \vee T) \wedge (R \vee S \vee T)$
- $\neg W \vee \neg R \vee S \vee R$

13) Let Person(x) denote x is a person and let HasSocNr(x,n) mean x has social security number n. Let Digitis(n,k) mean number n has k digits.

Consider the statement "Everyone's social security number has nine digits".

Here are two formulas in first order logic.

(i) $\forall x, n \ Person(x) \Rightarrow (HasSocNr(x, n) \wedge Digits(n, 9))$

(ii) $\forall x, n \ (Person(x) \wedge HasSocNr(x, n)) \Rightarrow Digits(n, 9)$

Which one is a correct representation of the statement?

- None is correct
- Only (i) is correct
- Only (ii) is correct
- Both are correct

14) We want to Skolemise the following sentence in first-order logic:

$$\forall x \; \exists y \; \forall z \; A(y, z) \;\Rightarrow\; [B(y) \;\wedge\; (C(x, y) \;\vee\; D(y, z))]$$

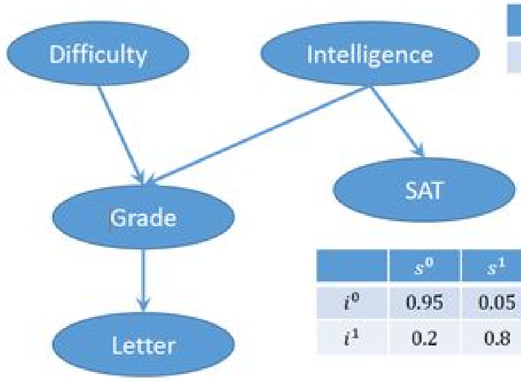
Only one the following four substitutions, produces a correct Skolemisation. Which one?

- $\{y/C\}$ with C a skolem constant
- $\{y/S(z)\}$
- $(y/S(x)\}$
- $\{y/S(z, x)\}$

15) Consider the Bayesian Recommendation Letter Network depicted below. This Bayesian Network has nodes for corresponding random variables (1) Difficulty, (2) Intelligence, (3) Grade, (4) SAT and (5) Letter. The network describes a model which relates the intelligence of a student, the difficulty of a course he takes, the grade obtained by the student, the SAT score of the student, and the strength of the recommendation letter obtained by the student from the professor. We denote these random variables with their first letters D, I, G, S and L respectively. Each of the random variables are discrete random variables and take values from a finite domain. The conditional probability tables (CPT) are shown in the figure. For example, the CPT associated with the node for the random variable G has 2 columns associated with the 2 possible values of G namely g0 and g1, and 4 rows corresponding to the combinations of values possible for the direct dependencies of the node. For example, according to the CPT of variable G: $P(g0 \mid i0,\ d0) \;=\; 0.3$.

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^0	g^1
i^0, d^0	0.3	0.7
i^0, d^1	0.05	0.95
i^1, d^0	0.9	0.1
i^1, d^1	0.5	0.5

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^0	0.9	0.1
g^1	0.4	0.6

What is the value of the full joint probability ?

- 0.118
- 0.014
- 0.006
- 0.212

Once again consider the Bayesian Recommendation Letter Network of the previous question. Consider the following two statements about conditional independencies encoded by the network structure.

[I] Variable L is dependent of variable S given the value of the variable G

[II] Variable D is independent of variable I if the values for variables G and L are unknown

Which statements are true?.

- Only I is true
- Both I and II are true
- Neither I nor II is true
- Only II is true

16) Given an urn ***H*** containing 5 white balls, 2 red balls and 3 blue balls. Assume that one draws 3 balls with replacement from ***H***. Let the outcome of this chance experiment be $X = (Xw, Xr, Xb) = (1, 1, 1)$ i.e. 1 white ball, 1 red ball and 1 blue ball (order not relevant). What is the probability P(X)?

- 0.05
- 0.10
- 0.18
- 0.20

17) One witness of a nighttime accident involving a taxi declares that the taxi was green. Another witness of the same nighttime accident also declares independently that the taxi was green. All taxis in town are blue or green. It is known that under dim lighting conditions discrimination between blue and green is 70% reliable. Which means that $P(W = b \mid C = b)$ as well as $P(W = g \mid C = g)$ are 0.70, where C is the two valued variable with values g and b indicating the color of the taxi, C = b means *the taxi is blue* and W is the two valued variable indicating the declaration of the witness, W = b means *the witness says the taxi is blue*. Suppose that 7 out of 10 taxis are actually blue. What is the probability that the taxi is blue (and thus not green).

- 0.70
- 0.30
- 0.06
- 0.54

18) This is an adapted version of the brewery example discussed in the lecture. Now we assume that the database of the brewery consists of the following samples in which the column A is the class label:

Ex.	U	H	I	T	S	A
1	Y	M	N	P	M	N
2	N	S	N	P	L	N
3	Y	M	N	A	M	N
4	N	M	N	P	S	N
5	N	M	Y	P	M	Y
6	Y	N	N	A	S	N
7	N	N	N	A	S	Y
8	N	S	N	A	M	Y
9	N	L	Y	P	S	Y
10	N	M	N	P	S	Y

What is the entropy of this dataset with respect to the class label A?

- 1
- 0
- 0.5
- None of the above.

Once again consider the dataset above. What will be the best approximation of the information gain for feature with respect to class label ? You can find a table of values at the last page of this exam.

- 0.40
- 0.60
- 0.57
- 0.43

Once again consider the brewery example above. Assume that the attribute has the highest information gain w.r.t. to the class label . What will be the attribute for the decision node below the branch if ones uses information gain as selection criteria?

- I
- H
- S
- T

19) Consider the following part of the decision tree, with two leaf nodes and one parent node which splits on attribute A. The notation P:x N:y means that the node has x positive examples and y negative examples.

In order to apply X^2 pruning one has to calculate the value of Δ given by

$$\Delta = \sum_{i=1}^{i=d}[\frac{(p_i - \widehat{p_i})^2}{\widehat{p_i}} + \frac{(n_i - \widehat{n_i})^2}{\widehat{n_i}}]$$

What is the value of in this case?

- 10.0
- 26.0
- 34.0
- 18.0

20) Consider the Neural Network (NN) depicted below, where the dummy (``bias'') inputs and corresponding weights are not shown!

We assume the following weights:

$$w_{0,3} = 1,\ w_{1,3} = -2,\ w_{2,3} = -1, w_{0,4} = -1,\ w_{1,4} = -1,\ w_{2,4} = -1.$$

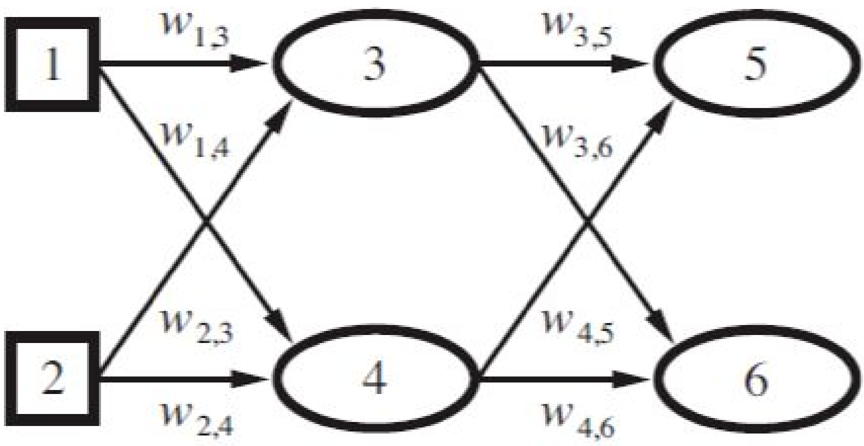
The activation function of the neurons 3 and 4 is the ***Logistic*** function (called the ***sigmoid*** function in the slides). What is the output of this NN on the input x = (1, -1)? First component is the output of neuron 3, second component the output of neuron 4. You can find a table values for the ***Logistic*** function in the formula sheet on your table.

- (-0.50, -0.27)
- (0.50, 0.27)
- (0.01, 0.02)
- (0.00, -1.00)

Once again consider the above NN and weights. Assume that for a given input x the output is $(0.4,\ 0.6)$ (first component is the output of neuron 3) and the target output is $(0, 1)$ Moreover we assume a L_2 loss function given by $1/2[(y3 - a3)^2 + (y4 - a4)^2]$. What is the value for the delta of neuron 4: $\Delta[4]$
(formula book: $\Delta[j] \leftarrow g'(in_j) * (y_j - a_j)$)

- 0.4
- -0.1
- -0.4
- 0.1

21) Consider the Neural Network (NN) depicted below, where the dummy ("bias") inputs and corresponding weights are not shown!



The activation function is the ***Logistic*** function and the weights are given by: $w_{0,3} = 1,\ w_{1,3} = -2,\ w_{2,3} = -1, w_{0,4} = 2,\ w_{1,4} = -1,\ w_{2,4} = -2,$. $w_{0,5} = 1,\ w_{3,5} = -2,\ w_{4,5} = -1, w_{0,6} = 2,\ w_{3,6} = 1,\ w_{4,6} = -2$. Moreover assume that $\Delta[5] = 0.2$ and $\Delta[6] = 0.7$ and the outputs of neuron 3 and 4 are given by $a_3 = 0.5$ and $a_4 = 0.6$. Moreover we assume a learning rate of $\alpha = 0.5$. What will be the value of $\Delta[3]$?

- 0.150
- 0.300
- 0.038
- 0.075

22) A certain classifier was tested on a test set, resulting in the following confusion matrix:

		Predicted class		
		C_1	C_2	C_3
Actual Class	C_1	160	20	31
	C_2	11	120	9
	C_3	15	10	120

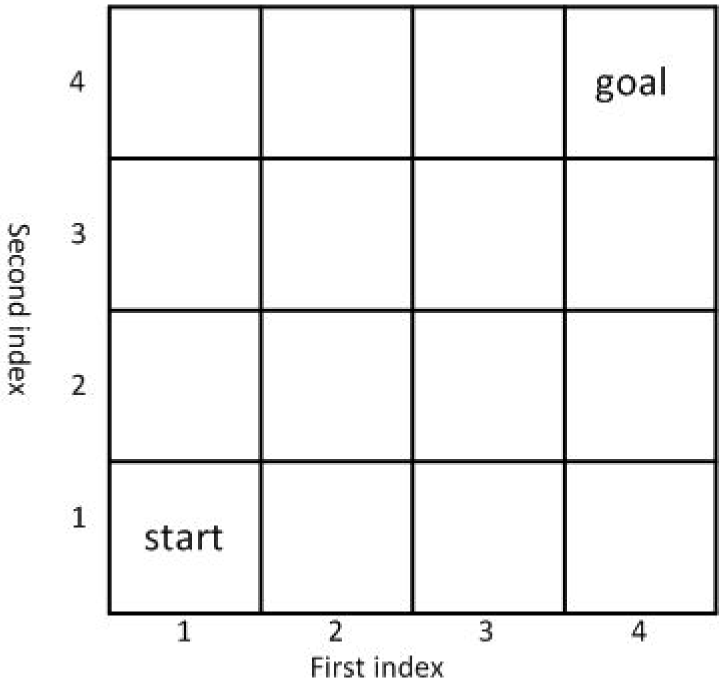
What is the recall for class C2?

- 0.30
- 0.80
- 0.71
- 0.86

Once again consider the confusion matrix in the above question. What is the error rate of this classifier?

- 0.14
- 0.19
- 0.30
- 0.24

23) Consider a grid of with start state in the corner left down and goal state right up.



Assume that the agent can move in four directions and these actions are deterministic. If it bounces to a wall it will stay in the same cell. Moreover the agent cannot leave the goal state, it is a terminal state. This means no action can be taken in the goal state. The over an empty set is 0 by definition. Assume that every state has reward of 0 except for the goal state, which as reward 5.00. Assume that the agent applies the following value iteration algorithm, where is the deterministic transition function.

[1] Initialize $U_{old}(s)$ to 1 for all s.

[2] Repeat

[3] $\quad For\ all\ s \in S$

[4] $\quad\quad U_{new}(s) = R(s) + max_a[\gamma U_{old}(\delta(s,a))]$

[5] $\quad$ End (for all)

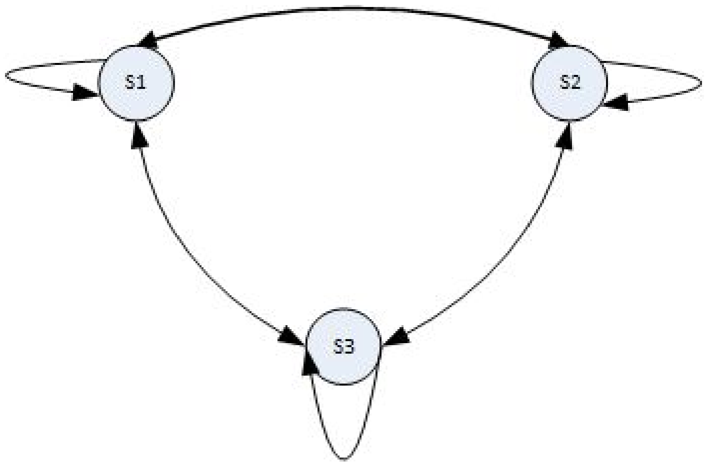
[6] $\quad U_{old} = U_{new}$

[7] Until $U_{new}(s)$ does not change for all s

Assume that the value for discount factor γ is 0.6. What will after 4 iterations of the ``Repeat'' loop (statements [2] -- [7]) be the value of $Unew(s)$ for $s = (3,3)$? Select the alternative which is closest to your answer.

- 5.00
- 1.08
- 0.65
- 1.80

24) Consider the following 3 state MDP, with states s1, s2 and s3.



The rewards are given by $R(s1) = -0.4$ and $R(s3) = -0.6$. In this simple state space the agent can only do two actions *leave* and *stay* with the following transition probabilities:

	s_1	s_2	s_3
$P(s\|s_1, stay)$	0.8	0.1	0.1
$P(s\|s_1, leave)$	0.1	0.8	0.1
$P(s\|s_2, stay)$	0.1	0.8	0.1
$P(s\|s_2, leave)$	0.2	0.6	0.2
$P(s\|s_3, stay)$	0.1	0.1	0.8
$P(s\|s_3, leave)$	0.1	0.8	0.1

Assume that the agent computed the following utilities for these states: $U(s1) = 3$, $U(s2) = 4$ and $U(s3) = 2$. What is the policy π for the states $s1$ and $s2$ corresponding to the these utilities U?

- $\pi(s1) = stay \text{ and } \pi(s2) = stay$
- $\pi(s1) = leave \text{ and } \pi(s2) = stay$
- $\pi(s1) = leave \text{ and } \pi(s2) = leave$
- $\pi(s1) = stay \text{ and } \pi(s2) = leave$

25) We use the following terminology:

Manhattan-norm (l_1-norm): $||x||_1 = \sum_{i=1}^{i=n} |x_i|$

Euclidian-norm (l_2-norm): $||x||_2 = \sqrt{\sum_{i=1}^{i=n} x_i^2}$

Which of the following statements regarding kernel functions is true?

- Decreasing the value for σ in a Gaussian kernel $K(X,Z) = e^{-\frac{||X-Z||^2}{2\sigma^2}}$ decreases the value of the kernel function.
- Using the Manhattan-norm (l_1 -norm) instead of the Euclidean-norm (l_2 -norm) in the Gaussian kernel decreases the value of the kernel function.
- A linear kernel $K(X,Z) = <X,Z>$ cannot be used if one of the support vectors X has a negative feature value.
- A linear kernel $K(X,Z) = <X,Z>$ cannot be used if one of the data points Z has a negative feature value.

26) We use the following terminology:

- True Positive: An actual anomaly detected as anomaly.
- True Negative: An actual benign sample detected as benign.
- False Positive: An actual benign sample detected as anomaly.
- False Negative: An actual anomaly detected as benign.

Consider a dataset *D* that has been split into *Dtrain* and *Dtest* using a k-fold split. We train a NIDS based on a one-class SVM using *Dtrain*. Suppose that the training dataset *Dtrain* also includes malicious datapoints. How will this affect the performance of the NIDS when predicting *Dtest*, compared to training data that only contains benign datapoints?

- The False Negative rate of your NIDS will increase.
- The True Positive rate of your NIDS will increase.
- The accuracy of your NIDS will increase.
- The False Positive rate of your NIDS will increase.