Test Intelligent Systems

Number of questions: 23
Session period: 18 december 2020 12:15 - 15:00
Duration: 105 minutes

Instruction

Exam duration: 1 HOURS 45 Minutes, unless you get 30 minutes extra time.

The following tools are permitted:
- Calculator
- Scrap paper (nothing written or printed on it)
- Check here for the cheat sheet

Please remember to follow the Proctoring protocol that was shared earlier. This means that no phones are allowed to be used, no headphones, not communication with others etc. Please make sure that you have made a desk-scan (no room scan is required). If you have any doubts, use the Proctoring help.

This second partial exam consists of 23 questions of varying complexity. In total you can reach maximally: 60 Points

There are 5 topics according to the main topics of lectures 9-16 (Set Theory, Fuzzy Logic, Calculating with Probabilities, Bayesian Networks and Machine Learning). For each topic, there are some simple reproduction questions (1 or 2 points), understanding questions (2 to 4 Points) and do-questions (3 to 6 Points).

Make sure you do not lose too much time on individual questions.

Multiple *Choice* questions have only **one** correct solution. In this case, the options are given by circles. **Multiple *Answers*** questions have one or more correct solution. In this case, the options are given by squares. The score of Multiple Answer questions is calculated automatically by the system as follows: Score = proportion good * (1 - proportion wrong * (1/a + (((n-k)/n)*(1-1/a))), where n is the total number of alternatives, k the number of correct answers and a a weighting factor set to 2. The more correct answers you give the more points you get, but wrong answers also count negatively.

Proctoring

Question 1 – Multiple choice – Question-ID: 219297 (1 point)

Please make sure that you follow the Proctoring protocol. This means that at this stage

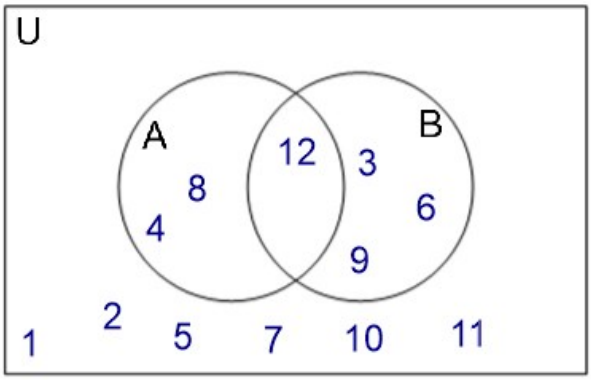
- you should have done **the desk-scan** (not a room scan, but shown your desk). You can still do it at this stage.
- **only permitted tools** are on your desk (Calculator, scrap paper).
- you are **clearly visible** (and remain so during the exam)
- you will **not communicate** with anybody during the exam (i.e. phones etc are OFF and not in proximity).
- you will have shown your **student-id**, which clearly shows your image
- you do not use **head-phones**

A I did not do it. I am aware of the risk that my exam might be declared invalid.

B I did it and fullfill all the requirements of the protocol.

Set Theory

Question 2 – Multiple choice – Question-ID: 217973 (2 points)



Given the Venn diagram above, what is the complement of A union B, (A U B)'?

A {1, 2, 5, 7, 10 ,11, 12}

B {1, 2, 5, 7, 10, 11}

C {12}

D {1,2, 3, 4, 5, 6, 7, 8, 9, 10 ,11}

Question 3 – Fill in (numerical) – Question-ID: 217107 (2 points)

Suppose we have a universal set that contains 41 elements. If |A| = 14, |B| = 19 and |A ∪ B| = 28, what is the cardinality of A intersection B, |A ∩ B|?

Uncertainty and Vagueness

Question 4 – Multiple choice – Question-ID: 205072 (1 point)

Consider the following statement: "Half of the Dutch men are married". Which of the following statements is true?

A The author of this statement describes the marital status of Dutch men with a degree of vagueness.

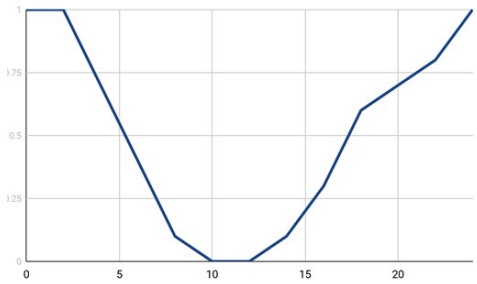
B The author of this statement describes the marital status of Dutch men with a degree of uncertainty.

Question 5 – Multiple choice – Question-ID: 219971 (2 points)

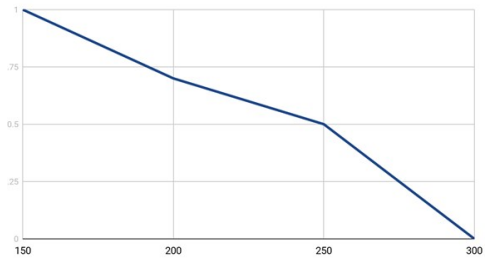
The director of our institute wants to reduce the environmental impact of the company by reducing the CO2 emissions caused by transportation. Therefore, she wants to give a devise a policy to support those who do not live very close (they can come by bicycle) or those who live very far (they should do remote work).

Since you know the director is a big fan of fuzzy set theory, you decide to model this problem with fuzzy sets. How do you model the fuzzy set of "**employees who live moderately far away**", i.e. those who do not live close, but also not far away?

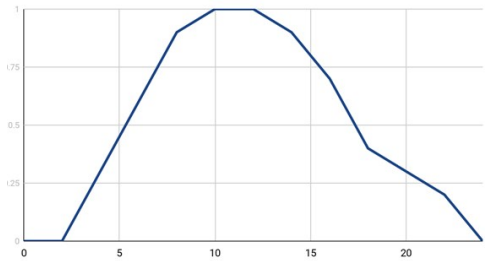
A



B



C



Question 6 – Fill in (multiple) – Question-ID: 217115 (4 points)

Given the following rules for fuzzy logic:

– f(¬A) = 1 - f(A)
– f(A v B) = max(f(A);f(B))
– f(A & B) = min(f(A);f(B))
– f(A -> B) = min(1;1 – f(A) + f(B))

In Fuzzy logic the meaning of a formula is defined as a value between 0 and 1. What is the meaning of f(((A -> B) v B) & (¬A v ¬B)) if f(A)=0.8 and f(B)=0.5?

f((A -> B) v B) = _____
f(¬A v ¬B) = _____
f(((A -> B) v B) & (¬A v ¬B)) = _____

Probability Theory

Question 7 – Multiple response – Question-ID: 217964 (3 points)

Consider the following full joint probability distribution:

P	Q	R	P(P, Q, R)
true	true	true	0.26
true	true	false	0.33
true	false	true	0.04
true	false	false	0.14
false	true	true	0.03
false	true	false	0.08
false	false	true	0.11
false	false	false	0.01

Which statements are true?

A P(P = true v Q =true v R = true) = 1

B P(R = false) = 0.56

C P(Q=true|R-true) = P(R=true|Q=true)

D P(Q = true) = 0.26

E P(P = true & Q = true) = 0.59

F P(R = false) = 0.63

Question 8 – Fill in (numerical) – Question-ID: 217977 (4 points)

You are planning a picnic today, but the morning is cloudy
- 50% of all rainy days start off cloudy.
- But cloudy mornings are common, as about 40% of days start cloudy.
- This month is usually a dry month (only 10% of the days tend to be rainy)

What is the chance of rain during the day given the morning is cloudy?

Question 9 − Multiple choice − Question-ID: 219972 (1 point)

What is the most important assumption that underlies the idea of Bayesian Networks, which allows these models to calculate probabilites for large sets of random variables efficiently?

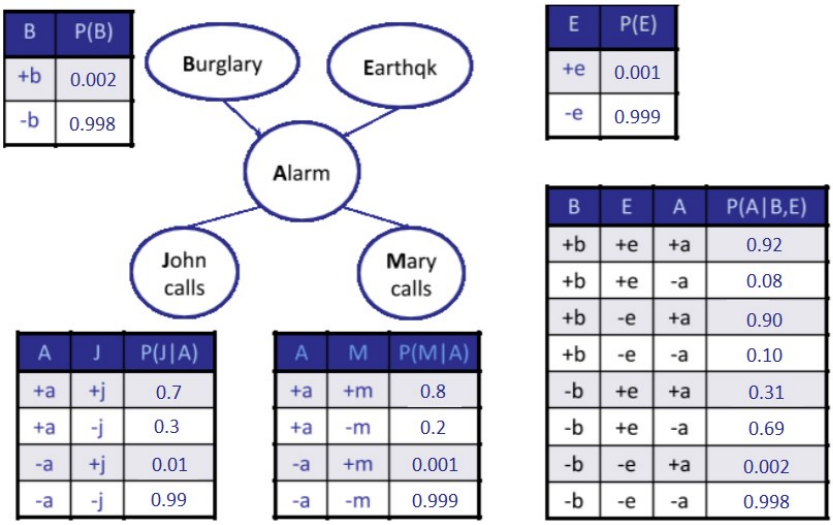
A The assumption that there are only finitely many models, and that we can therefore simply add the conditional probabilities in the Bayesian Networks to calculate the local probabilities.

B The assumption that from unsatisfiable formulas everything can be derived. This means that the probability of a child is conditional of the probability of their parents and their siblings in the network.

C The assumption that variables are independent of their siblings and other non-descendend nodes, so that their probability can be calculated only on the basis of their parents.

D The naiveness assumption that all variables are independent, and that the denominator in Bayes rule can be ignored. Therefore we only deal with estimates and not proper probabilites.

Question 10 − Fill in (numerical) − Question-ID: 217962 (4 points)



B	P(B)
+b	0.002
-b	0.998

E	P(E)
+e	0.001
-e	0.999

B	E	A	P(A|B,E)
+b	+e	+a	0.92
+b	+e	-a	0.08
+b	-e	+a	0.90
+b	-e	-a	0.10
-b	+e	+a	0.31
-b	+e	-a	0.69
-b	-e	+a	0.002
-b	-e	-a	0.998

A	J	P(J|A)
+a	+j	0.7
+a	-j	0.3
-a	+j	0.01
-a	-j	0.99

A	M	P(M|A)
+a	+m	0.8
+a	-m	0.2
-a	+m	0.001
-a	-m	0.999

Given the following Bayesian network, what is the probability that Mary calls? Round your answer to 6 decimal places.

Machine Learning

Question 11 – Fill in (multiple) – Question-ID: 205101 (3 points)

What kind of model do we need for the following cases (Classification, Regression, Ranking, Clustering)? Also indicate whether this would be supervised or unsupervised machine learning.

Please write words (from the above lists, without spaces) into the appropriate fields.

- A model that predicts the temperature based on air pressure and light intensity. We trained it with a lot of past observations.
  - Kind of model?
  - Supervised/Unsupervised?
- We want to make a system that sorts pictures according to what kind of landscape they represent (One of forest, city, dessert, coastline, and others). We used a lot of examples to train the system.
  - Kind of model?
  - Supervised/Unsupervised?
- We want to create groups of researchers working on similar topics, what ML technique do we use?
  - Kind of model?
  - Supervised/Unsupervised
- We want to order search results. We do this based on how often the search term occurs on the page.
  - Kind of model?
  - Supervised/Unsupervised

Question 12 – Multiple choice – Question-ID: 205084 (1 point)

Which of the following statements is true?

A Evaluating a Machine Learning model should always done by people who were not involved in the learning

B Evaluating on the training data will give untrustworthy results

C Evaluating with cross validation is often not applicable, as it requires too much data

D Evaluating on a specific testing set will result in overfitting

---

Question 13 – Fill in (multiple) – Question-ID: 217100 (3 points)

What is the Precision and Recall for this classification (no rounding)?

	actual +	actual -
predicted +	525	75
predicted -	225	340

Precision:

Recall:

Question 14 – Multiple choice – Question-ID: 218341 (3 points)

Consider the following training set of positive examples examples {(1,1),(1,-1),(-2,2),(-2,-2)} and negative examples {(3,3),(3,-3),(-4,4),(-4,-4)}. As seen in the last working sheet there is no linear separator (a linear function separating all the positive from the negative examples).

Consider a design matrix with two new features (replacing the old ones) $z_1 = x*y$ and $z_2 = x+y$.

Which of the following statements is true?

A There is still no linear separator for this new design matrix

B The function $z_1 = 2* z_2 + 2$ is a linear separator for this new design matrix.

C There are too many linear separators for this new design matrix to decide.

D The function $z_1 = - z_2 + 4$ is a linear separator for this new design matrix

E Neither $z_1 = 2* z_2 + 2$ nor $z_1 = - z_2 + 4$ are linear separators for this design matrix, but there exist one.

Question 15 – Multiple response – Question-ID: 205096 (4 points)

Consider the trading agent trying to infer which books a user reads based on keywords supplied for earlier books he read. Suppose the learning agent has the following data:

Reading preferences S:					
Books	crime	academic	local	music	**Reads**
b1	True	false	false	true	**true**
b2	True	false	false	false	**false**
b3	false	true	false	true	**true**
b4	false	true	true	true	**false**
b5	false	false	false	true	**true**

Apply a Bayes Learner whether the following book will be read or not:

b6	true	true	false	false

We apply smoothing, which assumes a value 0.01 in case of no occurrences.

Which of the following statements is true?

A P(reads|+crime,+academic,-local,-music) is estimated through 2/5 * 1/2 * 0 * 1/2 * 1/2 = 0

B P(-reads|+crime,+academic,-local,-music) is estimated through 2/5 * 1/2 * 1/2 * 1/2 * 1/2 = 0.025

C P(-reads|+crime,+academic,-local,-music) is estimated through 3/5 * 1/3 * 1/3 * 2/3 * 0.01 = 1/2250

D P(reads|+crime,+academic,-local,-music) is estimated through 2/5 * 1/2 * 1/2 * 1/2 * 1/2 = 0.025

E The algorithm predicts that the reader will probably read the book *b6*.

F P(reads|+crime,+academic,-local,-music) is estimated through 3/5 * 1/3 * 1/3 * 1 * 0.01 = 1/1500

Question 16 – Multiple response – Question-ID: 219973 (2 points)

Consider the sentence:

In the true Computer Sciences, where science rules, a rule computes truth.

Which operations have been applied to this sentence when it has been reduced to the following set of text features:

< true, computer, science, sciences, rule, rules, computes, truth >

A Part of Speech Tagging

B Noun Tagging

C Stopword removal

D Parsing

E Stemming

F Remove Punctuation

G Sentiment Analysis

H Remove Capitalisation

Question 17 – Multiple response – Question-ID: 205107 (2 points)

For K-nearest neighbours, which of the following are true (multiple answer can be correct)?

A it is always best to choose k as big as the available dataset. This improves classification accuracy, because there will be no problems with noise.

B We do not need to define how the system measures the distance between two points because it only needs to find the nearest ones.

C We use the majority label to reduce effects of noise in the dataset.

Question 18 − Multiple choice − Question-ID: 205103 (4 points)

You are given the following words extracted from documents (one sentence is a document for this exercise), with the associated label.

1. {dog, sleep, near, field} label: passive
2. {cat, walk, field} label: active
3. {field, is, full, poppy} label: passive
4. {cat, dog, run, field} label: active
5. {cat, is, mad, dog, sleep, poppy} label: passive

Suppose you use a k-nearest neighbour algorithm (k=3) to decide whether the sentence {cat, sleep, poppy, dog} belongs to the class active or passive.

Take the following steps:
1. convert each of these sets into a feature vector (this is what we did in the lecture, but this step is not strictly necessary given the distance function we will use in 2)
2. create a k-NN classifier, with k = 3. As a distance, we use "1/(number of words in common)" --- / stands for devided by.
3. Use a majority voting to decide the label. In case of a tie, increase k, until the tie breaks.

Which of the following statements is true?

A The three nearest neighbours are **1, 2, 4** with labels **passive, active, active** so the classification is **active**

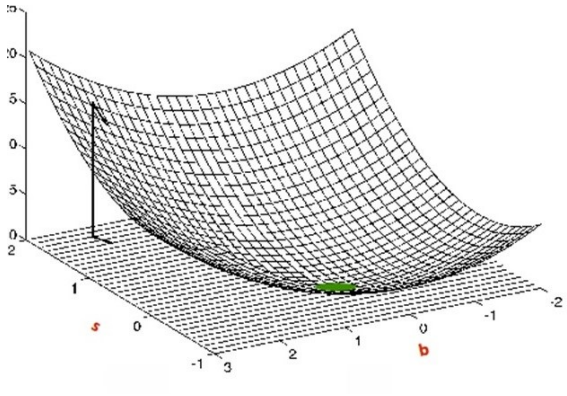
B The three nearest neighbours are **2, 4, 5** with labels **active, active, passive** so the classification is **active.**

C The three nearest neighbours are **1, 2, 5** with labels **passive, active, passive** so the classification is **passive**.

D The three nearest neighbours are **1, 4, 5** with labels **passive, active, passive** so the classification is **passive**.

Question 19 – Multiple response – Question-ID: 205100 (3 points)

In the following image, also shown in the lecture, the green point is identified as the optimal solution to a learning problem.



Which of the following statements are true? (multiple answers are possible).

A With a Neural Network we can predict the value of the 3rd dimension (orthogonal to s and b on the x and y axis) given the representation of the features according to variables s and b.

B The green point is the point where f(x) = sx+b reaches a minimum, ie where the linear separator reaches the lowest point (in other words, f(x)=0

C There are two ways to find the best machine learning model in this scenario: either we analytically find the green point as the point where the derivation of the error function is 0, or we apply gradient descent as a search method to move step by step to the green point. (Gradient descent is a local search similar to hill-climbing)

D The green point refers refers to the linear function with the lowest error rate (the one that fits a particular learning model best according to the training data)

E The third dimension (orthogonal to s and b) enumerates all possible linear separators, so that the green point is the optimal one (as it is the lowest).

F The green point refers to the point in the binary vectorspace with the best training example (the one that fits a particular learning model best)

G The third dimension (orthogonal to s and b) shows the error rate for a linear separator function, eg. f(x) = sx+b.

Question 20 – Multiple choice – Question-ID: 219974 (4 points)

This exercise is about Regression.

Let us try to fit a linear function to a number of training sets where the slope s is known to be -1 (like in the practise exam).

Suppose we have three pairs of values (1,0), (2,0) and (3,-1) for two features. What is the optimal fitting function in this model class?

A f(x)= -x + 14/6

B f(x)= -x + 10/3

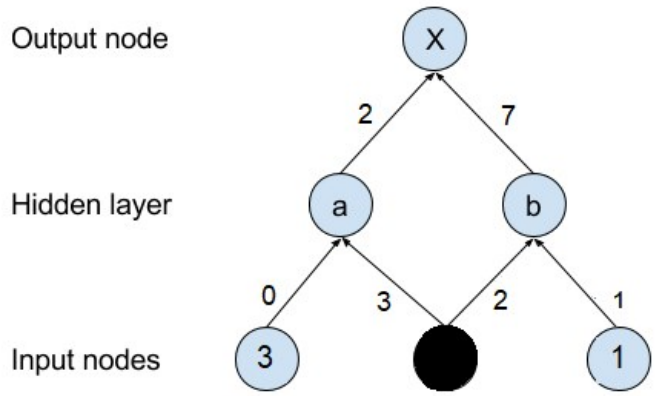
C Neither of the functions in the other options is the optimal fitting function, but there exist one.

D f(x)= -x + 1

E f(x)= -x + 10/6

F There is no optimal fitting function

Question 21 − Fill in (numerical) − Question-ID: 217098 (3 points)



In the above image, a basic multilayer feedforward neural network, with weights and values on the input nodes is shown. It has no activation function (all nodes are linear). The black node is a bias node. What is the value on the output node (marked X)?

Philosophy of Mind

There will be a single question related to Philosophy of Mind.

Question 22 – Multiple choice – Question-ID: 218343 (2 points)

What is meant by "symbol grounding"

A Symbol grounding is the idea that robots can never be truly intelligent, becaue they are not as grounded as humans are

B Symbol grounding is the idea that the Chinese characters in Searle's Chinese Room experiment are mapped to English words, so that the person in the box can understand them

C Symbol grounding is the idea that all intelligent behaviour is grounded in symbol manipulation

D Symbol grounding is the idea that a true understanding of abstract concepts requires a mapping to sensory experience

---

Question 23 – Multiple choice – Question-ID: 218342 (2 points)

Which one of the below is a reasonable objection to the Turing test

A The Turing Test dates from the '50s and computers are very different now

B The Turing Test incorrectly assumes that all intelligence will be similar to human intelligence

C The Turing Test is limited to narrow domains of expertise

D The Turing Test only tests for intelligence, and not for consciousness

---

End of test Intelligent Systems