Test Intelligent Systems

Number of questions: 23
Session period: 16 december 2021 15:30 - 18:30
Duration: 135 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

This second partial exam consists of 23 questions of varying complexity. In total you can reach maximally: 62 Points

There are 5 topics according to the main topics of lectures 8-13 (Set Theory, Calculating with Probabilities, Bayesian Networks, Machine Learning and Philosophy of Mind). 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.

There will be no answers during the exam over the content. If you believe that something is ambigous (or even wrong), write down the question and get in contact with me after the exam. If you can make a reasonable case, you will get the points you deserve.

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.

Set Theory

Question 1 – Multiple choice – Question-ID: 291325 (1 point)

Which is an example of disjoint sets?

A C = {whole numbers} and D = {rationalnumbers}

B A={multiples of two} and B = {multiples of three}

C E = {even numbers} and F = {odd numbers}

D G = {multiples of five} and H = {multiples of ten}

Question 2 – Fill in (numerical) – Question-ID: 292179 (2 points)

Suppose we have |A| = 12, |B| = 17 and $|A \cup B| = 18$, what is the cardinality of the intersection of A and B, i.e. $|A \cap B|$?

Question 3 – Multiple choice – Question-ID: 291142 (2 points)

Let A and B be two non-empty subsets of a set X such that A is not a subset of B, then for all A and B:

A A and the complement of B are non-disjoint

B A is a subset of the complement of B

C B is s subset of A

D A and B are disjoint

Probability Theory

Question 4 – Multiple choice – Question-ID: 291333 (2 points)

Given the full joint distribution as in the below table. What is the vector of probability values for Toothache, given that cavity is true?
(Note: Normalize the values)

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

A 0.6,0.4

B 0.5,0.5

C 0.2, 0.8

D 0.8,0.2

Question 5 − Multiple choice − Question-ID: 291335 (2 points)

You throw two fair dice, one green and one red, and observe the numbers uppermost. The following pair of events are dependent.

1. the sum is 5
2. the red die shows a 2

A False

B True

Question 6 − Multiple response − Question-ID: 293850 (4 points)

Given the following information:

- The probability that someone with a cold also has fever is 60%.
- The prior probability of having a cold is 0.001.
- More or less 1 out of 100 people in the population has fever.

I have fever. You want to calculate the probability that I have a cold using Bayes' rule. Which of the following statements is true? (multiple answers possible)

A The probability cannot be calculated as P(cold|fever) is not provided in the text.

B The calculation has to use a naive version of Bayes rule and assume conditional independence of fever and cold.

C P(cold|fever) < P(fever|cold)

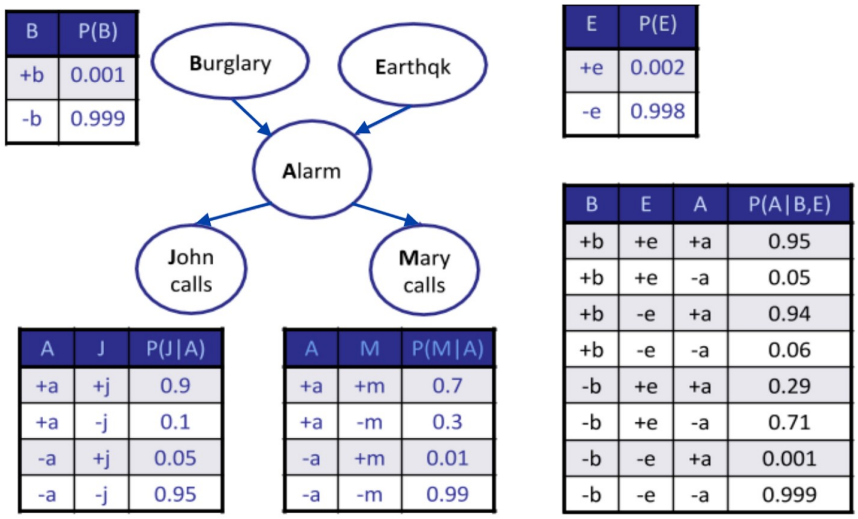
D The necessary information that P(cold) = 0.001, P(fever|cold) = 0.6 and P(fever)= 0.01 is given in the text.

E P(cold|fever) = P(fever) * P(fever|cold) / P(cold) = 0.01 * 0.6 / 0.001 = 6

F P(cold|fever) = P(fever|cold) * P(cold) / P(fever) = 0.6 * 0.001 / 0.01 = 0.06

Question 7 – Fill in (multiple) – Question-ID: 291145 (4 points)

Given the following Bayesian network, Please fill in the following blanks:



B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

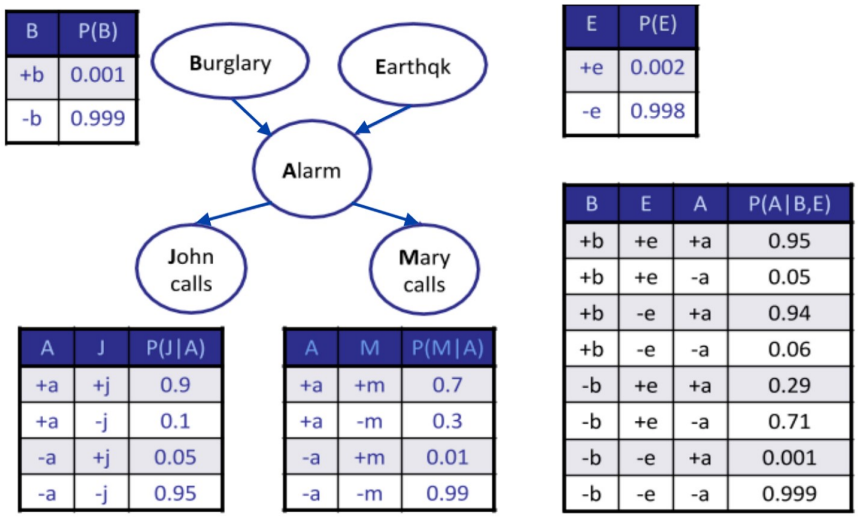
a) The probability that there has been no burglary is

b) The probability that the alarm has been activated if both earthquake and Burglary didn't occur is

c) The probability John called if the alarm is activated is

Question 8 – Fill in (multiple) – Question-ID: 292213 (4 points)

Given the following Bayesian network from the previous question, Please fill in the blank:



B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

This question requires some more calculation. Make sure you do not loose too much time on it.

What is the probability that the alarm has been activated, but neither burglary nor earthquake has occurred, and both John and Mary call?

(Hint: The full joint distribution is the product of the local conditional distributions $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$)

Machine Learning

Question 9 – Multiple choice – Question-ID: 291339 (1 point)

For a well-defined learning problem, what are important components that can influence the choice of model.

A All of the other options

B The type of feedback

C The type of available features

D The type of learning task.

Question 10 – Fill in (multiple) – Question-ID: 291569 (5 points)

What kind of model do we need for the following cases (**classification, regression, ranking, collaborative filtering, clustering**)?
First, indicate whether this would be **supervised or unsupervised** machine learning.

Example:
a) Searching for overlaps in the parts of the population affected by diabetes and arthritis.
unsupervised
clustering

a) Deciding if an MRI image shows a cancerous tumour or not.

b) Twitter is a social networking site that allows micro-blogging services where people broadcast short, public messages refer as tweets. Twitter wants to find out about similar users.

c) Model to forecast the sales number based on the history of the sales.

d) Auto-categorizing the news provided by the News channel.

e) Suggesting a meal in the cafeteria self-order counter based on ratings of meals given by patients and hospital workers.

Question 11 – Multiple choice – Question-ID: 291604 (1 point)

Find out the wrong combination.

A False positive=correctly identified

B True negative=correctly rejected

C All of the mentioned

D False negative=incorrectly rejected

Question 12 – Fill in (multiple) – Question-ID: 291613 (6 points)

A machine learning model categories given input into two classes “YES” or “NO”. The below table provides the actual class of the input and the predicted output from the model. Looking at the table solve the confusion matrix table. Calculate the true positive, false positive, true negative, false negative values, precision and recall for this classification.
Complete the below blank space with the correct value.

Actual Class	Predicted Class by model
NO	NO
YES	YES
YES	NO
YES	YES
NO	YES
YES	NO
YES	YES

		Actual Class	
Predicted Class		YES	NO
	YES	TP	FP
	NO	FN	TN

TP
FP
FN
TN
Precision
Recall

Question 13 – Multiple choice – Question-ID: 291588 (1 point)

A resampling procedure is used to evaluate machine learning models on a limited data sample for training referred to as Cross-validation. The resampling procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. If k=8, then it will be 8-fold cross-validation.

A False

B True

Question 14 – Multiple choice – Question-ID: 291607 (1 point)

Which of the following distance measure can we use in the case of categorical (having finite number of categories or distinct groups) in k-NN?

1. Hamming Distance
2. Euclidean Distance
3. Neighbourhood Distance
4. Regression Distance

A 2 and 3

B 1, 2 and 3

C 1 and 4

D 1 and 2

E all of them

F 1

Question 15 – Multiple response – Question-ID: 291616 (1 point)

Which of the following techniques can be used to pre-process textual information in feature vectors used to determine similarity between texts.

A Grounding

B Syllabisation

C Lemmatization

D Discretisaton

E Stemming

Question 16 – Multiple choice – Question-ID: 291618 (2 points)

Given the three feature vectors:

1. <1,1,1,1,1,0,0,1,0,0>
2. <0,1,1,0,0,1,1,1,0,0>
3. <0,1,0,1,1,0,0,0,1,1>

Which vector represents the nearest neighbour of the below vector.
<0,1,0,0,1,1,0,1,0,0>

If we apply Euclidean distance?

A None of them

B 2

C 1

D 3

Question 17 − Multiple response − Question-ID: 292265 (4 points)

Observations have been made about the weather, the clarity of explanation and happiness of individual students.

	Sunny	Rainy	Windy	Comprehensive expl.	Label - happy
D1	Y	N	N	Y	Y
D2	N	Y	Y	N	Y
D3	N	Y	N	Y	Y
D4	N	Y	Y	N	N
D5	Y	Y	Y	Y	N

Use Naive Bayes classification to estimate whether the student will feel happy or not on a day with no sun and no rain, although windy and attending a good
explanation? Which of the statements is true?

A Naive Bayes (with smoothing of adding a value of 0.02) will predict that the student will feel NOT happy.

B The prior probability P(Label-happy=Y) is 3/5

C Naive Bayes (with smoothing of adding a value of 0.02) will predict that the student will feel happy.

D With Naive Bayes we predict that Label-happy=Y if P(-sun,-rain,+w,+e | Label-happy=Y) > P(-sun,-rain,+w,+e | Label-happy=N)

E The conditional probablity P(Sunny|Label-happy=Y) is 2/3

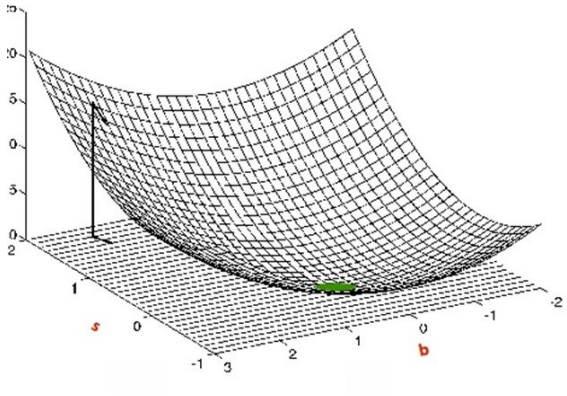
F With Naive Bayes we predict that Label-happy=Y if P(Label-happy=Y|-sun,-rain,+w,+e) > P(Label-happy=N|-sun,-rain,+w,+e)

G The prior probability P(Label-happy=Y) is 2/3

H The conditional probablity P(Sunny|Label-happy=Y) is 1/3

Question 18 – 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 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

B 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)

C 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.

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

G 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)

Question 19 – Multiple choice – Question-ID: 292316 (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.

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 (so minimises the error rate)?

A There is no optimal fitting function

B f(x)= x - 5/3

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

D f(x)= x - 3/5

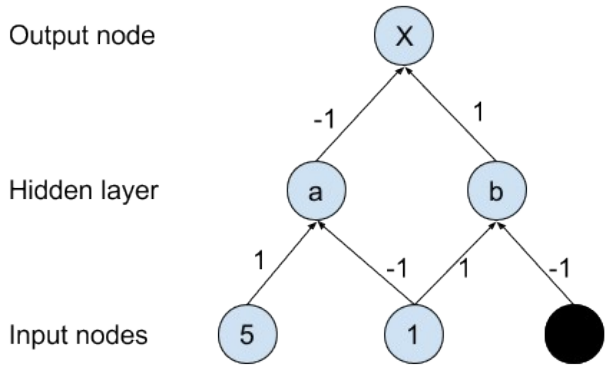
E f(x)= x + 3/5

F f(x)= x + 5/3

Question 20 – Fill in (numerical) – Question-ID: 85389 (4 points)

Here, we see a basic multilayer feedforward neural network, with weights and values on the input nodes. It has no activation function (all nodes are *linear*). The black node is a bias node with value 1.

What is the value on the output node (marked X)?

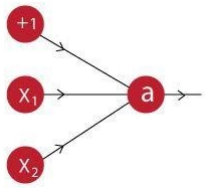


Question 21 – Multiple choice – Question-ID: 291623 (4 points)

(Advanced) Do not loose to much time of this question.

Let us assume we implement an AND function to a single neuron. Below is a tabular representation of an AND function:

X1	X2	X1 AND X2
0	0	0
0	1	0
1	0	0
1	1	1



The activation function of our neuron is denoted as:

$$f(x) = \begin{cases} 0, & for\ x < 0 \\ 1, & for\ x \geq 0 \end{cases}$$

This just means that node a returns a 0, if its value is smaller than 0, and a 1 otherwise.

Which of the following weights of this network would implement such a logical AND function as defined in the table?

A b = 1.5, w1 = 2, w2 = 2

B None of these

C b = 1, w1 = 1.5, w2 = 1.5

D b = -1.5, w1 = 1, w2 = 1

Philosophy of Mind

There will be two questions related to Philosophy of Mind.

Question 22 – Multiple choice – Question-ID: 291785 (2 points)

Libbet's experiment (about the timing of the intention for an action and the corresponding brain/muscle activity for the action) shows that (one answer is correct)

A we adjust our intentions to what we are instructed to do

B we don't really have intentions, we just ascribe them to each other

C our conscious intention for an action is subjective

D our conscious intention for an action is not the cause of that action

---

Question 23 – Multiple response – Question-ID: 291779 (2 points)

Which of these are reasonable objections to the Turing test (multiple answers possible)

A it degrades humans by comparing them to machines

B it imposes a white male dominant view on the notion of intelligence

C it is 70 years old, so it no longer applies to modern AI

D it uses human intelligence as the standard for intelligence

---

End of test Intelligent Systems