Group A Final Exam LV 185.A83– Class of 2020- Machine Learning for Health Informatics

Vor- und Zuname →		
Matrikelnummer →		
Studienkennzahl →		
Semesteranzahl →		
Prüfungsdatum →		⬅

Welcome to the final exam of the lecture LV 185.A83

This exam consists of a sum of questions in different question blocks. You can reach a maximum of 100 credit points, which will be used for calculation of your final grade:

Fail	Poor	Average	Good	Excellent!
Nicht genügend	Genügend	Befriedigend	Gut	Sehr gut
5	4	3	2	1
0-50	51-69	70-79	80-89	90-100

Before you start, please answer three general questions.

I feel well at the moment.

Yes, I fully agree *No, I fully disagree*

①----------②----------③----------④----------⑤

Health AI is a difficult subject.

Yes, I fully agree *No, I fully disagree*

①----------②----------③----------④----------⑤

Health AI is very interesting.

Yes, I fully agree *No, I fully disagree*

①----------②----------③----------④----------⑤

Good luck! ☺ ...

A) Yes/No decision question block

Please check the following sentences and decide if the sentence is true = YES; or false = NO; for each correct answer you will be awarded 2 credit points.

01	The major problem for Machine Learning in Health Informatics (for health AI generally) is that we do not have independent, identically and equally distributed data.	❒ Yes ❒ No	2 total
02	The term probability mass function (PMF) is used for continues variables and the term probability density function is used for discrete variables.	❒ Yes ❒ No	2 total
03	Causality in the sense of Judea Pearl is the science of cause and effect, whereas Causability is a measurement to what extent an machine explanation maps to human understanding .	❒ Yes ❒ No	2 total
04	According to the Expected Utility Theory of von Neumann & Morgenstern an optimal single decision is the decision whose expected utility is arg min E(U|d) .	❒ Yes ❒ No	2 total
05	The concordant partial AUC developed by Carrington et al. combines vertical and horizontal perspectives and equals the partial c statistics.	❒ Yes ❒ No	2 total
06	Markov decision processes (MDP) are important because they are random processes in which the future, given the present is dependent of the past.	❒ Yes ❒ No	2 total
07	Modelling “gsunder Hausverstand” (common sense) can be done by a Bayesian model and is an important aspect for concept learning.	❒ Yes ❒ No	2 total
08	Class imbalance occurs in classification problems, where you have one class with very few samples compared to other classes in your dataset, this results in bias and unfair learning models.	❒ Yes ❒ No	2 total
09	Knowledge modeling is a process of creating a computer interpretable model of standard specifications about a kind of process and/or about a kind of facility or product.	❒ Yes ❒ No	2 total
10	Layer-Wise Relevance Propagation is a general solution for understanding classification decisions by pixel-by-pixel (or layer-by-layer) decomposition of linear classifiers.	❒ Yes ❒ No	2 total

Sum of Question Block A (max. **20 points)**	

B) Multiple choice question block (MCQ)

Note: The following questions are composed of two parts: the stem, which identifies the question or problem and a set of alternatives which can contain 0, 1, 2, 3 or 4 correct answers to the question, along with a number of distractors that might be plausible – but are incorrect. Please **select the correct answers** by ticking ☒ - and do not forget that every question can have 0, 1, 2, 3 or 4 correct answers. Each question will be awarded 4 points ***only if everything is correct.***

01	To reach a level of usable intelligence we need to … ❒ a) ... collect large amounts of top-quality data. ❒ b) ... disentangle underlying explanatory factors of the data. ❒ c) ... extract knowledge and generalize. ❒ d) ... understand the data in the context of a problem domain.
02	Automatic Machine Learning fails when … ❒ a) ... having large amounts of top-quality data. ❒ b) ... dealing with rare events. ❒ c) ... dealing with NP-hard problems. ❒ d) ... confronted with Salt-and-Pepper-Noise.
03	Causality is …. ❒ a) … not so important, because correlation is easier to measure. ❒ b) … important, because it allows to answer “why” questions. ❒ c) … the relationship that you observe between two or more variables ❒ d) … aka Causation.
04	Abductive Reasoning … ❒ a) … is used to get best explanation from incomplete set of preconditions. ❒ b) … seeks to find the simplest and most likely explanation. ❒ c) … solutions are aka "best available" or "most likely. ❒ d) ... is used often in diagnostic expert systems.
04	Standardized medical data … ❒ a) … is the majority of all data in the hospital. ❒ b) … is the basis for accurate communication. ❒ c) … contains tags or markers to separate structure and content. ❒ d) ... ensures that information is interpreted by all subsequent medical professionals with the same understanding.
06	Information retrieval models in the health care domain, which are following the concept of “reasoning with uncertainty” apply … ❒ … Probability theories. ❒ … Graph theories. ❒ … Algebra calculus. ❒ … Fuzzy set theories.

07	MCMC is important because … ❒ a) ... it uses the concept of randomness to solve problems with uncertainty. ❒ b) ... it allows to solve multidimensional integrals – otherwise intractable. ❒ c) ... it can be used for systems with many dof. ❒ d) ... it allows to calculate characteristics of high-dimensional distributions.	4 total
08	The ethics commission at a standard medical research institution checks ❒ a) … the output of a study in terms of citations. ❒ b) … the social value. ❒ c) … fair selection of study population (inclusion – exclusion criteria). ❒ d) ... risk-benefit ratio and the minimization of risks.	4 total
09	A determination of more/less, a transformation $x \mapsto f(x)$ in $\mathbb{R}$ and the basic statistics median and percentiles are applicable on … ❒ a) ... nominal data structures. ❒ b) ... ordinal data structures. ❒ c) ... interval data structures. ❒ d) ... rational data structures.	4 total
10	Superintelligence … ❒ a) ... is inevitable ❒ b) ... is impossible ❒ c) … we do not know it ❒ d) ... does not need a body – it just needs an Internet connection	4 total

Sum of Question Block B (max. 40 points)	

C) Free recall blocks – please follow the instructions below. At each question you will be assigned the credit points indicated if your option is correct (partial points may be given).

01	Probability. 0,8% of the patients have cancer. Cancer is deteced with a probability of 98% and the chance that a negative test-result is correct is 97 %. a) Write down the equation for the posterior probability b) Calculate **all** probabilities c) Calculate P(Cancer | Test=true) Solutions:	4 each 16 total

02	Decisions: a) Explain in detail the two types of decision and b) provide two examples for each of them:	8 total
03	Shannon-Entropy: a) Explain in your own words what entropy is and why it is important for us b) Write down the equation for the Shannon-Entropy c) Which entropy is higher: A) 123456789 or B) 12344679 d) Calculate the Entropy for 1100101 (b=2)	2 each 8 total

04	Problem Solving: Human vs. Computers In which problem solving tasks do humans perform better and in which problem solving tasks do computers perform better? Provide three different examples of each and explain them and define what "better" means!	4 each 8 total

Sum of Question Block C (max. 40 points)	

Thank you very much for taking the exam!

Our best wishes for your further studies,

Andreas Holzinger and the Human-Centered AI Group