
WRITTEN EXAMINATION
“Foundations of Machine Learning and Data Mining”
SUMMER TERM 2006
October 11TH, 2006
PROF. DR. RALF MÖLLER

Name, Student Id:

Name of curriculum:

Signature:

IT IS NOT PERMITTED TO WRITE ON THESE PIECES OF PAPER BEFORE THE START SIGNAL. IN ADDITION, WRITING ON THESE PIECES OF PAPER AFTER THE END SIGNAL WILL HAVE THE EFFECT THAT YOU DO NOT PASS THIS EXAM WITHOUT FURTHER WARNINGS. THIS ALSO APPLIES IF YOU WRITE YOUR NAME AND/OR STUDENT ID AFTER THE END SIGNAL.

Please note:

- 1. Do not open the exam sheets until permitted by the supervisor!**
- 2. Put your student identification card as well as your passport on the table.**
- 3. Check if your student Id is on the list of Ids handed out by the supervisors. If your Id is not on this list, then**
 - (a) Fill out a proviso, which will be given to you by the supervisor. You may not start the exam until you have returned the proviso to the supervisor.**
 - (b) Also fill out the additional form given to you by the supervisor. Please go to the students office which is responsible for your curriculum and let them sign the form. Personally return the signed form to the STS secretary (Harburger Schloßstr. 20, 2nd floor, Frau Hantschmann); bring your passport for identification as well.**
- 4. You have 90 minutes for answering the questions. Additional resources are not allowed.**
- 5. There is sufficient space for your solutions on the examination sheets.**
- 6. If you receive additional pieces of paper from the supervisor, please write your name and student Id also on these pages, and add a page number.**
- 7. If you need to leave the examination room, silently inform the supervisors. Do not leave your table. Wait until the supervisor approaches your desk and gives you the permission to leave the room.**

1) Explain the advantages and disadvantages of using probabilities for making decisions in an uncertain environment where you don't have a complete description of the state of the world of interest (e.g. dental diagnosis world). Compare with other approaches.

2) In probabilistic theory, possible worlds are defined by assignment of values to random variables.

- a. What types of random variables exist?
- b. Describe the necessary properties of domain value for random variables.
- c. What are atomic events and what are their properties?

3) Describe the terms “probability distribution” and “full joint probability distribution” by given an example. Suppose, your world of interest has three random variables.

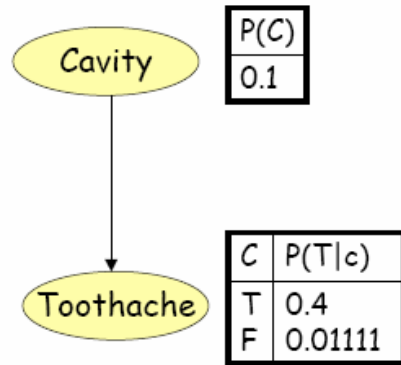
4) Suppose, you have world which is described by three random variables A, B and C, and you know the joint full distribution. How can you derive the formula for computing the distribution $\mathbf{P}(A|C=c)$?

5) Describe the terms “absolute independence” and “conditional independence” by giving the related formulas and an example.

6) A Bayesian network is a data structure that describes the dependencies among random variables.

a. Enumerate the properties of a Bayesian network?

- b. The following Bayesian network directly answers queries like $P(t|c)$. But what about queries like $P(c|t)$? Can those queries be answered? If yes, and how? If know, why not?



7) Describe the steps of learning conditional probabilities (parameter learning) in a Bayesian network where we only know the structure of the network and do have a set of examples. The Maximum Likelihood should be used in this process.

8) What are the weaknesses of decision tree learning? (3 characteristics).

9) What is a “version space” and what are the advantages in the context of learning?

10) What are the G-Boundary and S-Boundary in version spaces? Explain their role for answering queries.

11) Give a short description of the following three terms and explain the differences:

- a. Explanation-based learning
- b. Relevance-based learning
- c. Knowledge-based inductive learning

12) Which kind of learning problems cannot be solved with attribute based learning algorithms and why?