
WRITTEN EXAMINATION
“Foundations of Machine Learning and Data Mining”
SUMMER TERM 2008
August 7TH, 2008
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. If you are not registered for the exam, then fill out a proviso, which will be given to you by the supervisor.**
- 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.**
- 7. If you need to leave the examination room, silently inform the supervisor.**
- 8. The exam is closed book. That means that the only things you are allowed to have on your desk or use during the exam are pens and the exam itself.**
- 9. All phones off. A switched on phone is considered cheating.**
- 10. Cheating will cause you to fail this exam.**

1) Machine learning has different approaches. Explain the following terms and describe the differences in the approaches.

a. Supervised Learning

(2 pts)

b. Unsupervised Learning

(2 pts)

2) Name and explain three approaches for selecting the next attribute in decision tree learning.

(3 pts)

(continue your answer on the reverse of this sheet if necessary)

- 3) What is meant by the term “noisy data” in the context of decision tree learning? Give two examples for different kinds of noise. (3 pts)

- 4) Explain how the k-nearest-neighbor classification algorithm works and how the influences of the closest neighbors can be strengthened. Also describe a possible situation where this scheme does not work well. (4 pts)

(continue your answer on the reverse of this sheet if necessary)

- 5) Explain the general structure of a version space and give an algorithmic description of updating the version space if a positive example is used for learning. (4 pts)

- 6) Suppose you have a large amount of textual description of image files that you want to classify using a fixed set of categories. You also have a training set of textual descriptions that are already classified. In this scenario we make the assumption that each word occurs independently in image file descriptions, with frequencies determined by the category of the image file description. For this task a naïve Bayes model can be used.

- a. Explain how such a model can be constructed, given as "training data" a set of images files and descriptions that have already been assigned to categories. (2 pts)

(continue your answer on the reverse of this sheet if necessary)

- b. Explain how to compute the category of a new image file given its description. (2 pts)

- 7) In the context of parameter learning in Bayesian networks describe the approaches and differences of
 - a. Full Bayesian Learning (2 pts)

- b. Maximum a posteriori (MAP) Learning (2 pts)

(continue your answer on the reverse of this sheet if necessary)

c. Maximum Likelihood Learning

(2 pts)

8) Describe the 5 basic steps of the k-means clustering algorithm (you can draw diagrams to explain the steps). (5 pts)

(continue your answer on the reverse of this sheet if necessary)

StudentId:

Name:

page 7 of 7

- 9) Describe the main differences of hierarchical and partitional clustering algorithms. (2 pts)

(continue your answer on the reverse of this sheet if necessary)