



Institutt for datateknikk  
og informasjonsvitenskap

**Eksamensoppgave i**

## **TDT4173 – Machine Learning and Case-Based Reasoning**

**Monday December 5th 2011, 09:00 - 13:00**

*Contact person during the exam: Professor Agnar Aamodt, mobil phone 92611144.*

*Språkform: Engelsk*

*Tillatte hjelpemidler: D*

*Ingen trykte eller håndskrevne hjelpemidler tillatt.*

*Bestemt, enkel kalkulator tillatt.*

*Read the text carefully. If you find that some problems are not specified at the expected level of detail, please state explicitly what assumptions you make in your answer.*

*Sensufrist: Thursday January 5<sup>th</sup>, 2012.*

*For all questions, brief answers are just as good as longer answers, as long as they provide satisfactory answers.*

### **Question 1**

- a) How would you define the version space?  
How does the candidate elimination algorithm represent the version space?  
How does the candidate algorithm learn a concept?
- b) Under which condition is the candidate elimination algorithm able to accurately classify a given unseen instance as positive or negative?  
Under which condition is it able to accurately classify all instances as either positive or negative?
- c) What is inductive bias?  
What are the two main types of inductive bias? Give an example method for each type.  
Give an example of a method without inductive bias. What is that method able to learn?
- d) What is meant by the term overfitting in the context of inductive learning?  
How can overfitting be avoided?

### **Question 2**

- a) Give at least two reasons why Bayesian learning methods are important.
- b) What characterizes the naive Bayes classifier? What is the main difference between the naive Bayes method and a Bayesian belief network?
- c) What is the maximum a posteriori (MAP) hypothesis, and how can it be determined based on Bayes Theorem? What is the maximum likely hood (ML) hypothesis, and under which conditions is it identical to the MAP hypothesis?
- d) The EM algorithm consists of two repetitive steps. Describe what happens in these two steps.

### **Oppgave 3**

- a) Instance-based and case-based learning methods are called “lazy learners”. Why?
- b) In Aha’s paper about case-based learning algorithms 4 deficiencies of instance-based / case-based learning methods are addressed.  
Which 4 deficiencies?  
Aha argues that the 4 deficiencies can be corrected. How?
- c) Protos uses “matching knowledge” in its Retrieval step.  
How is the matching knowledge structured/represented?  
How is it used to match two cases when one or more case features are different?
- d) Schank and Leakes’ paper is about CBR for explanation and creativity.  
What is their method for achieving creative systems?

### **Question 4**

- a) KBANN combines inductive and analytic learning. Describe how, using an example if you like.
- b) How does the sequential covering method in machine learning operate?  
Give an example of a system that uses sequential covering.
- c) What is a weak classifier, and what is ensemble learning?  
Describe the most important differences between "bagging" and "boosting".
- d) Describe how SVMs work. Your description can be quite informal if you want, as long as you are able to present the main ideas. Cover the following topics:
  - When the instances from the classes are linearly separable
  - When the instances from the classes are NOT linearly separable
  - The use of "kernels"