



Institutt for datateknikk
og informasjonsvitenskap

Eksamensoppgave i

TDT4173 – Machine Learning and Case-Based Reasoning

Monday December 20th, 09:00 - 13:00

Contact person during the exam: Professor Agnar Aamodt, mobile 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.

Question 1 - Fundamentals

- a) Define the notion of *learning* in a computer program, either as a formal definition or in your own words. Give two examples of (well-posed) learning problems according to your definition.
- b) What is a version space? How does the candidate elimination algorithm represent the version space?
- c) 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?
- d) What is inductive bias? Describe the two main types of inductive bias. What is a method without any inductive bias able to learn?

Question 2 – Bayesian learning

- a) What does Bayes Theorem state? Describe its meaning in words.
- b) What characterizes the naive Bayes classifier? What is the main difference between the naive Bayes method and a Bayesian belief network?
- c) What does the maximum a posteriori (MAP) hypothesis express? How can it be determined based on Bayes Theorem? What does the maximum likelihood (ML) hypothesis express?
- d) Given the following joint distribution over the random variables A, B, and C:

A	B	C	P(A,B,C)
0	0	0	1/8
0	1	0	1/8
0	0	1	1/8
0	1	1	1/8
1	0	0	1/8
1	1	0	1/8
1	0	1	1/8
1	1	1	1/8

Is A conditionally independent of B given C?

If your answer is YES, change the top two rows so that the answer becomes NO.

If your answer is NO, change the top two rows so that the answer becomes YES.

Question 3 – Case-based reasoning

- a) What characterizes lazy learning, as opposed to eager learning? Give an example of both.
- b) In their paper “Footprint-based retrieval” Smyth and McKenna propose a competence model of a set of cases. Describe briefly what is expressed by the following five core concepts:

- Retrieval space
- Adaptation space
- Reachability set
- Coverage set
- Footprint set

c) In his taxonomy of similarity mechanisms, Cunningham splits similarity mechanisms into four subcategories. List them and describe their main characteristics.

d) Given the following examples, represented using three Boolean-valued features:

F1 = false F2 = false F3 = false category = +
 F1 = false F2 = false F3 = true category = -
 F1 = false F2 = true F3 = false category = +
 F1 = true F2 = true F3 = true category = +
 F1 = true F2 = false F3 = false category = -
 F1 = true F2 = false F3 = true category = +

The task is to classify an unknown example as belonging to category + or - using a case-based or instance-based approach, by identifying the 3 nearest neighbors and using the most common category among them as the estimated class.

Propose a simple similarity function for this task. Determine the category for the following test example:

F1 = false F2 = true F3 = true category = ?

Question 4 – Mixed questions

Please answer Yes or No, and add a sentence to explain your answer.

- a)
- Is cross-validation a method that uses separate sets of training examples, validation examples, and test examples?
 - Is “post-pruning” a method to avoid overfitting?
 - Is there a method that can tell us the true error of a learned hypothesis based on the error over the example data?
- b)
- Is FOIL a sequential covering method?
 - Is decision tree learning (e.g. ID3) a sequential covering method?
 - Is KBANN an explanation-based learning method?
- c)
- Is a weak classifier a classifier that performs slightly worse than random guessing?
 - Is Boosting a method for combining the votes of multiple classifiers?
 - Are ensemble methods more useful for strong classifiers than for weak ones?
- d)
- Is the SVM learning method guaranteed to find the globally optimal hypothesis with respect to its target function??
 - Is it so that while gradient descent methods have the problem of local minima, the EM method does not?
 - Is it so that no method can do better than a naïve Bayes classifier if the distribution of data is known?