English

Norges teknisk-naturvitenskapelige universitet
Institutt for datateknikk og informasjonsvitenskap

Examination in	TDT4171 Artificial Intelligence Methods
Date	May 29th
Antall timer	4
Antall studiepoeng	7,5
Antall sider:	3
Tillatte hjelpemidler	D
Sensurdato	June 19th

All 5 questions, including sub-questions, shall be answered. Each question is weighted as shown.

Contact during the examination: Assoc. Prof. Helge Langseth, IDI, tlf. (735)96488

Question 1 – Bayesian Networks (25%)

a) Describe the syntax and semantics of a Bayesian Network.

b) Model the following problem using a Bayesian network (only the *structure* is required, the quantitative part should not be given). The model should be as simple and easy to understand as possible:

There is a higher fraction of tuberculosis patients among those who have recently visited Asia than there is amongst those who have not. There is a larger fraction of bronchitis among smokers than there is among non-smokers. Both bronchitis as well as tuberculosis can lead to being short-winded (problems breathing). Both tuberculosis and lung cancer can be detected from X-ray scanning, but bronchitis cannot. The X-ray cannot be used to distinguish between lung cancer and tuberculosis. There is a higher fraction of lung cancer patients among smokers than among non-smokers.

c) A medical doctor gets your Bayesian network (including a quantitative part) and uses it on a patient who is short-winded and has just visited Asia. The medic wants to use the Bayesian network for decision support. To what extent can the model help the decision-making?

d) Would you say that a Bayesian network is a *natural* modeling tool for this problem? What are the characteristics of a problem domain where Bayesian networks can be used successfully? Can you give examples of situations where Bayesian networks are *not* suitable?

Question 2 – Instance-based and Case-based reasoning (15%)

a) What are the characteristics of Instance-based learning, as opposed to other types of machine learning methods?

b) How are problems solved in a *k*-nearest-neighbor method?

c) Describe the four main steps in the CBR-cycle. Explain briefly what happens in each step.

Question 3 - Neural Networks (25%)

a) Research on Artificial Neural Networks is partly motivated by knowledge about how the brain works. Explain this relation.

b) *Gradient descent* is a powerful, general-purpose algorithm, which amongst other things can be used for learning the weights $\boldsymbol{w}$ of a neural network. Give a reason why it is useful to update the weight w_i by using the formula $w_i \leftarrow w_i - \eta \cdot \frac{\partial E(\vec{w})}{\partial w_i}$, where η is a positive constant. Explain the symbols in the formula.

c) What are the strengths and weaknesses of Gradient Descent?

d) Create a multilayer artificial neural network, and find weights to ensure that the network represents the function x_1 XOR x_2, where x_1 and x_2 are binary inputs; that is, the output is true if exactly one of x_1 or x_2 is true, and false otherwise. You should make the network as simple as possible (i.e., containing as few nodes as possible).

Can the function x_1 XOR x_2 be represented exactly by a perceptron (a neural network *without* a hidden layer)? Explain your answer.

***Hint*: In question (d) you can let the transfer function *g( )* in the nodes be the *step-function*.**

Question 4 – Probabilistic models over time (20%)

a) Explain the *Markov Assumption* using your own words. Give an example where the Markov assumption is reasonable, and one where it is not.

b) What is *filtering* and *smoothing*? Explain the difference between the two.
(You may want to use a graph to explain, formulas are not required here.)

c) What are *Kalman Filters*? What assumptions are underlying a Kalman Filter model?

Question 5 – Mixed questions (15%)

a) What is the *Strong AI hypothesis*? What is *The Chinese Room*, and how is The Chinese Room used to argue against the Strong AI hypothesis?

b) What are *unigram*, *bigram* and *trigram* models for strings of words? Many probabilistic language-processing systems use *bigram* models (and not unigram or trigram). Give at least one argument for bigram models over unigram models and at least one argument for bigram models over trigram models.

c) What is the *maximum expected utility principle*, and why is this principle important when creating rational agents?