



Försättsblad till skriftlig tentamen vid Linköpings Universitet

(fylls i av ansvarig)

Datum för tentamen	2010-03-09
Sal	KÅRA
Tid	14-18
Kurskod	TDDC17
Provkod	TEN1
Kursnamn/benämning	Artificiell intelligens
Institution	IDA
Antal uppgifter som ingår i tentamen	8
Antal sidor på tentamen (inkl. försättsbladet)	9
Jour/Kursansvarig	Mariusz Wzorek
Telefon under skrivtid	070-3887122
Besöker salen ca kl.	15:30
Kursadministratör (namn + tfnr + mailadress)	Anna Grabska Eklund Ankn. 23 62, annek@ida.liu.se
Tillåtna hjälpmedel	Miniräknare
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	57 + 2

Linköpings Universitet
Institutionen för Datavetenskap
Patrick Doherty

Tentamen
TDDC17 Artificial Intelligence
09 March 2010 kl. 14-18

Points:

The exam consists of exercises worth 37 points.
To pass the exam you need 18 points.

Auxiliary help items:

Hand calculators.

Directions:

You can answer the questions in English or Swedish.
Use notations and methods that have been discussed in the course.
In particular, use the definitions, notations and methods in appendices 1-3.
Make reasonable assumptions when an exercise has been under-specified.
Begin each exercise on a new page.
Write only on one side of the paper.
Write clearly and concisely.

Jourhavande: Mariusz Wzorek, 070-3887122. Mariusz will arrive for questions around 15:30.

1. Consider the following theory (where x, y and z are variables and $nono, america$ and jim are constants):

$$\forall x[American(x) \wedge Weapon(y) \wedge Sells(x, y, z) \wedge Hostile(z) \rightarrow Criminal(x)] \quad (1)$$

$$\exists x[Owns(nono, x) \wedge Missile(x)] \quad (2)$$

$$\forall x[Missile(x) \wedge Owns(nono, x) \rightarrow Sells(jim, x, nono)] \quad (3)$$

$$\forall x[Missile(x) \rightarrow Weapon(x)] \quad (4)$$

$$\forall x[Enemy(x, america) \rightarrow Hostile(x)] \quad (5)$$

$$American(jim) \quad (6)$$

$$Enemy(nono, america) \quad (7)$$

- (a) Convert formulas (1) - (7) into clause form with the help of appendix 1. [2p]
- (b) Prove that $Criminal(jim)$ is a logical consequence of (1) - (7) using the resolution proof procedure. [3p]
- Your answer should be structured using a resolution refutation tree (as used in the book).
 - Since the unifications are trivial, it suffices to simply show the binding lists at each resolution step. Don't forget to substitute as you resolve each step.
- (c) Convert the following formula into clause form with the help of appendix 1. (x, y and z are variables). Show each step clearly. [1p]

$$\forall x[\forall y[Animal(x) \rightarrow Loves(x, y)] \rightarrow [\exists yLoves(y, x)]] \quad (8)$$

2. The following questions pertain to the course article by Newell and Simon entitled *Computer Science as an Empirical Enquiry: Symbols and Search*.

- (a) What is a *physical symbol system* (PSS) and what does it consist of? [2p]
- (b) What is the Physical Symbol System Hypothesis? [1p]
- (c) What is the Knowledge Representation Hypothesis? [1p]

3. A* search is the most widely-known form of best-first search. The following questions pertain to A* search:

- (a) Explain what an *admissible* heuristic function is using the notation and descriptions in (c). [1p]
- (b) Suppose a robot is searching for a path from one location to another in a rectangular grid of locations in which there are arcs between adjacent pairs of locations and the arcs only go in north-south (south-north) and east-west (west-east) directions. Furthermore, assume that the robot can only travel on these arcs and that some of these arcs have obstructions which prevent passage across such arcs.

The *Manhattan distance* between two locations is the shortest distance between the locations ignoring obstructions. Is the Manhattan distance in the example above an admissible heuristic? Justify your answer explicitly. [2p]

- (c) Let $h(n)$ be the estimated cost of the cheapest path from a node n to the goal. Let $g(n)$ be the path cost from the start node n_0 to n . Let $f(n) = g(n) + h(n)$ be the estimated cost of the cheapest solution through n .

Provide a general proof that A* using tree-search is optimal if $h(n)$ is admissible. If possible, use a diagram to structure the proof. [2p]

4. The following question pertains to Decision Tree Learning. Use the definitions and data table in Figure 3, Appendix 3 to answer this question. Figure 1 shows a partial decision tree for the Table in Figure 3 with target attribute *PlayTennis*.

- (a) What attribute should be tested in the box with the question mark on the right branch of the decision tree in Figure 1? Justify your answer by computing the information gain for the appropriate attributes in the Table in Figure 3 in Appendix 3. [3p]

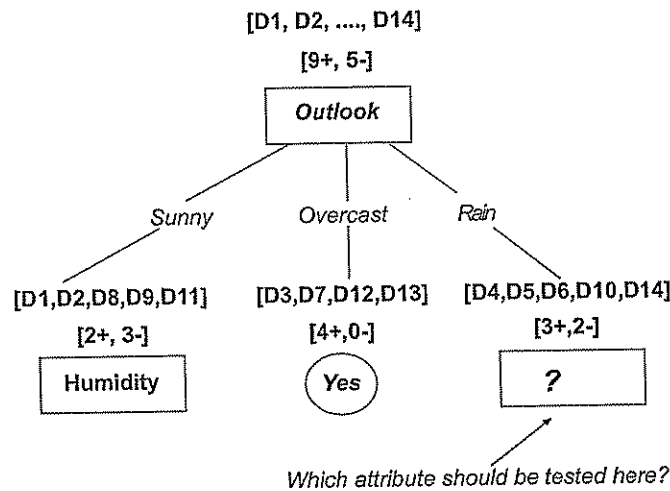


Figure 1: Partial Decision Tree for PlayTennis

5. Constraint satisfaction problems consist of a set of variables, a value domain for each variable and a set of constraints. A solution to a CS problem is a consistent set of bindings to the variables that satisfy the constraints. A standard backtracking search algorithm can be used to find solutions to CS problems. In the simplest case, the algorithm would choose variables to bind and values in the variable's domain to be bound to a variable in an arbitrary manner as the search tree is generated. This is inefficient and there are a number of strategies which can improve the search. Describe the following three strategies:

- (a) Minimum remaining value heuristic (MRV). [1p]
- (b) Degree heuristic. [1p]
- (c) Least constraining value heuristic. [1p]

Constraint propagation is the general term for propagating constraints on one variable onto other variables. Describe the following:

- (d) What is the Forward Checking technique? [1p]
- (e) What is arc consistency? [1p]

6. The following questions pertain to STRIPS-based planning. In order to use STRIPS one requires a language expressive enough to describe a wide variety of problems but restrictive enough to allow efficient algorithms to operate over it. The STRIPS language is the basic representation language for classical planners. Using terminology from logic (literals, ground terms, function free terms, etc.), describe how the following are represented in the STRIPS planning language:

- (a) State representation [1p]
- (b) Goal representation [1p]
- (c) Action representation [1p]

The following questions pertain to non-monotonic reasoning and STRIPS:

- (a) What is the Closed World Assumption? [1p]
- (b) How is the Closed World Assumption used in the context of STRIPS planning? In particular, how is it applied to state representation? [1p]

7. Modeling actions and change in incompletely represented, dynamic worlds is a central problem in knowledge representation. The following questions pertain to reasoning about action and change.

- (a) What is Temporal Action Logic? Explain by describing the ontology used in the formalism, that is, what is taken to exist, and what notation is used in the logical language to represent those things that are taken to exist. [2p]
- (b) What is the frame problem? Use the Wumpus world to provide a concrete example of the problem. Describe the problem using TAL notation (either with macros or without) or situation calculus notation. [1p]
- (c) What is the ramification problem? Use the Wumpus world to provide a concrete example of the problem. Describe the problem using TAL notation (either with macros or without) or situation calculus notation. [1p]
- (d) What is nonmonotonic logic? How can it be used to provide solutions to the frame problem? [1p]

8. Use the Bayesian network in Figure 2 together with the conditional probability tables below to answer the following questions. Appendix 2 may be helpful to use.

- (a) Write the formula for the full joint probability distribution $P(A, B, C, D, E)$ in terms of (conditional) probabilities derived from the Bayesian network below. [1p]
- (b) $P(a, \neg b, c, \neg d, e)$ [1p]
- (c) $P(e \mid a, c, \neg b)$ [2p]

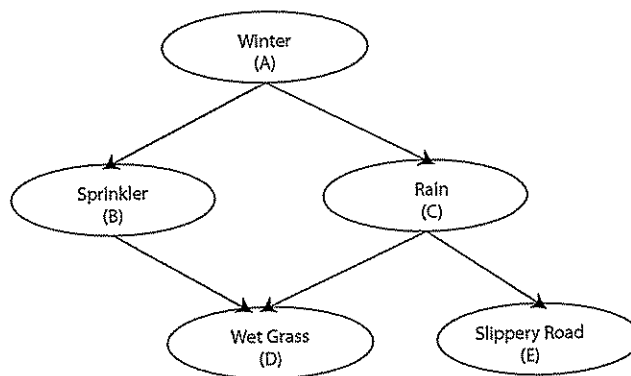


Figure 2: Bayesian Network Example

A	P(A)
T	.6
F	.4

A	B	P(B A)
T	T	.2
T	F	.8
F	T	.75
F	F	.25

A	C	P(C A)
T	T	.8
T	F	.2
F	T	.1
F	F	.9

B	C	D	P(D B, C)
T	T	T	.95
T	T	F	.05
T	F	T	.9
T	F	F	.1
F	T	T	.8
F	T	F	.2
F	F	T	0
F	F	F	1

C	E	P(E C)
T	T	.7
T	F	.3
F	T	0
F	F	1

Appendix 1

Converting arbitrary wffs to clause form:

1. Eliminate implication signs.
2. Reduce scopes of negation signs (move \neg inwards).
3. Standardize variables within the scopes of quantifiers (Each quantifier should have its own unique variable).
4. Eliminate existential quantifiers (Skolemization). This may involve introduction of Skolem constants or functions. Remember that:
 - $\exists xP(x)$ after Skolemization is $P(c)$ for some new unused constant c .
 - $\forall y\exists zQ(y, z)$ after Skolemization is $\forall yQ(y, f(y))$ for some new unused function constant f .
5. Convert to prenex form by moving all remaining universal quantifiers to the front of the formula.
6. Remove the universal quantifiers from the front of the formula since all remaining variables are universally quantified.
7. Put the quantifier free formula into conjunctive normal form (CNF). Two useful rules are:
 - $\omega_1 \vee (\omega_2 \wedge \omega_3) \equiv (\omega_1 \vee \omega_2) \wedge (\omega_1 \vee \omega_3)$
 - $\omega_1 \wedge (\omega_2 \vee \omega_3) \equiv (\omega_1 \wedge \omega_2) \vee (\omega_1 \wedge \omega_3)$
8. Eliminate the \wedge symbols.
9. Rename the variable symbols so that no variable symbol appears in more than one clause.

Appendix 2

A generic entry in a joint probability distribution is the probability of a conjunction of particular assignments to each variable, such as $P(X_1 = x_1 \wedge \dots \wedge X_n = x_n)$. The notation $P(x_1, \dots, x_n)$ can be used as an abbreviation for this.

The chain rule states that any entry in the full joint distribution can be represented as a product of conditional probabilities:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1) \quad (9)$$

Given the independence assumptions implicit in a Bayesian network a more efficient representation of entries in the full joint distribution may be defined as

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i)), \quad (10)$$

where $\text{parents}(X_i)$ denotes the specific values of the variables in $\text{Parents}(X_i)$.

Recall the following definition of a conditional probability:

$$\mathbf{P}(X | Y) = \frac{\mathbf{P}(X \wedge Y)}{\mathbf{P}(Y)} \quad (11)$$

The following is a useful general inference procedure:

Let X be the query variable, let \mathbf{E} be the set of evidence variables, let \mathbf{e} be the observed values for them, let \mathbf{Y} be the remaining unobserved variables and let α be the normalization constant:

$$\mathbf{P}(X | \mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \quad (12)$$

where the summation is over all possible \mathbf{y} 's (i.e. all possible combinations of values of the unobserved variables \mathbf{Y}).

Equivalently, without the normalization constant:

$$\mathbf{P}(X | \mathbf{e}) = \frac{\mathbf{P}(X, \mathbf{e})}{\mathbf{P}(\mathbf{e})} = \frac{\sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})}{\sum_{\mathbf{x}} \sum_{\mathbf{y}} \mathbf{P}(\mathbf{x}, \mathbf{e}, \mathbf{y})} \quad (13)$$

Appendix 3

Definition 1

Given a collection S , containing positive and negative examples of some target concept, the entropy of S relative to this boolean classification is

$$Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus},$$

where p_{\oplus} is the proportion of positive examples in S and p_{\ominus} is the proportion of negative examples in S .

Definition 2

Given a collection S , containing positive and negative examples of some target concept, and an attribute A , the information gain, $Gain(S, A)$, of A relative to S is defined as

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v),$$

where $values(A)$ is the set of all possible values for attribute A and S_v is the subset of S for which the attribute A has value v (i.e., $S_v = \{s \in S \mid A(s) = v\}$).

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Figure 3: Sample Table for Target Attribute PlayTennis

For help in converting from one logarithm base to another (if needed):

$$\log_a x = \frac{\log_b x}{\log_b a}$$

$$\ln x = 2.303 \log_{10} x$$

Note also that for the example, we define $0 \log 0$ to be 0.