Final Exam: Th. Dec.08, 1.30-3.30pm

Intro AI ICS 171 2005
Instructor: Max Welling

- *This exam is closed book*
- *You may use a calculator*
- *Always explain your answer*
- *Spend your time wisely: get a shot at each question*
- *You can get a total of 56 points*
- *Good Luck !*

1.(12 pts) **First Order Logic**
Consider the following sentence in FOL:

$$\forall x \text{ Married(Father(x), Mother(x))} \Rightarrow$$
$$\exists y \text{ Certificate}(y) \wedge \text{Names}(y, \text{Father}(x), \text{Mother}(x))$$

In English: *For every person who has a father and a mother that are married, there exists a paper which is a wedding certificate and which contains the names of both the father and the mother of this person.*

a.(4 pts) Identify the *functions, properties, binary relations, relations with arity 3, quantifiers and connectives* in this sentence.

a) answer: Function: Father, Mother, Property: Certificate, Binary Relation.: Married, Tertiary Relation: Names, Quantifiers: $\forall, \exists$, Connectives: $\rightarrow, \wedge$.

b.(4 pts) Consider the sentence: $\exists x\ [p(x) \Rightarrow q(x)]$. Assume we know that there is no value for x for which $q(x)$ is true. Is it still possible that the above sentence is true?

b) answer: Yes, because $p(x)$ can still be false.

c.(4 pts) Provide the truth-table for $(P \vee Q) \Rightarrow (P \wedge Q)$ where P and Q are two propositions.

b) answer: P:T,T,F,F, Q:T,F,T,F, $(P \vee Q) \Rightarrow (P \wedge Q)$:T,F,F,T.

2.(12 pts) **Inference**
Consider the following logical sentence in CNF form:
$(\neg A \vee \neg B \vee C) \wedge (\neg A \vee B) \wedge (A) \wedge (\neg A \vee \neg B \vee \neg C)$.

a.(2 pts) Is this sentence in Horn form?
Are all clauses definite clauses?

a) answer: Yes, No (last clause).

b.(4 pts) Use either resolution or forward chaining to show that this sentence is unsatisfiable. In other words, there is no assignment of False/True values for the variables A,B,C that will render the sentence True.

b) answer: FC: First rewrite al clauses into implications, and then derive A,B,C,False in that order. With resolution you use derive A,B,C in the same order and then show that you get an empty clause.

c.(2 pts) Next consider the following sentence:
$(\neg A \vee \neg B \vee C) \wedge (\neg A \vee B) \wedge (\neg A \vee \neg B \vee \neg C) \wedge (A \vee B) \wedge (A \vee \neg C)$.
Is this sentence in Horn form?

c) answer: No (fourth clause).

d.(4 pts) We will solve this problem using backtracking search. We use the following heuristics: 1) order the variables according to the number of clauses they are involved in (the variable involved in the largest number is first), 2) choose a value for a variable that satisfies the largest number of clauses.
Provide a solution (i.e. an assignment of values True/False to the variables such that all clauses are satisfied).

d) answer: A=F, B=T, C=F (no backtracking needed).

3.(20 pts) **Learning**
Consider the following dataset of 10 data-items.

A	B	+	-
T	T	3	1
T	F	1	2
F	T	0	0
F	F	0	3

where the column "A" represents the value for the first attribute, column "B" the value for the second attribute, the column "+" the number of data-cases with a positive label and the attribute values in the corresponding row and similarly for "–". For instance: there are 3 data-cases with a label "+" that have $A = T$ and $B = T$. In the following we will learn a very simple decision tree on this data-set.
You may use the fact that,

$$\text{Entropy}(p) = -p(T)\log(p(T)) - p(F)\log(p(F)) \quad (1)$$

where $p(T)$ can be estimated by the fraction of samples with value "True". Also remember that in computing the information gain, the entropy for each branch in the tree needs to be weighted by the fraction of the data-cases that follow that branch.

a.(4 pts) Compute the amount of information in the dataset, i.e. what is the negative entropy of the data-set?

a) answer: $P_T = 2/5, P_F = 3/3, I = -0.67$.

b.(6 pts) Compute the information gain if you split on attribute A. Similarly, compute the information gain if you split on attribute B. Which attribute is therefore preferred as the root node of the decision tree?

b) answer: Split on A. For the information after the split we need: $P(A = T) = 7/10, P(A = F) = 3/10, I(A = F) = -0.683, I(A = F) = 0, IG_A = 0.2$. Similarly for B: $IG_B = 0.18$. We prefer to split on A.

c.(4 pts) Draw the best two level decision tree, using the result from item b). What is the classification error on the training set?

b) answer: First split on A. If $A = F->[-]$. If A=T, split on B. If $B = T->[+]$, if $B = F->[-]$. There will be a total of 2 errors on the training data.

d.(2 pts) If you fit a very complex model (with many parameters) to a small dataset, and you observe that all the training cases have been correctly classified, can we conclude that the error on the testing data is also (close) to zero?

b) answer: No, one can over-fit on the training data.

e.(4 pts) Explain in words how a "1-nearest-neighbor" algorithm classifies new test data, given some training set.

b) answer: Classify a test case with the same label as the nearest training case.

4.(12 pts) **Uncertainty**

Joe needs to go to the doctor to check if he has "monkey pox" (MP). The doctor asks him 2 questions: 1) "Do you have red bumps (RB) on your body" and 2) "Do you have a fever (FR)". Joe's answers are "yes, I have red bumps" (RB=T), and "yes, I have a fever" (FR=T). The doctor (now worried) has the following joint probability table at his disposal:

	MP=T	MP=T	MP=F	MP=F
	RB=T	RB=F	RB=T	RB=F
FR = T	7.2E-7	1.8E-7	0.71999928	0.17999982
FR = F	8E-8	2E-8	0.07999992	0.01999998

a.(4 pts) Compute the prior probability of getting monkey pox $P(MP = T)$ from the joint table.

a) answer: $P(MP = T) = 7.2e - 7 + 1.8e - 7 + 8e - 8 + 2e - 8 = 1e - 6$.

b.(4 pts) Compute the conditional probability $P(FR = T, RB = T|MP = T)$.

b) answer: $P(FR = T, RB = T|MP = T) = P(FR = T, RB = T, MP = T)/P(MP = T) = 0.72$

c.(4 pts) Use Bayes' rule to compute what the doctor needs to know: $P(MP = T|FR = T, RB = T)$. Explain why this probability is actually very small, even though all the symptoms for Monkey Pox are present.

b) answer: $P(MP = T|FR = T, RB = T) = P(FR = T, RB = T, MP = T)/P(FR = T, RB = T) = 7.2e - 7/(7.2e - 7 + 0.71999928) = 1e - 6$ This is small because the prior probability on MP is very small, and the symptoms FR and RB did not add any information to the prior probability.