Introduction to Artificial Intelligence
4005-750-01, 20081

Final Examination, November 10, 2008
Instructor: Richard Zanibbi, Duration: 120 minutes

Name: _____

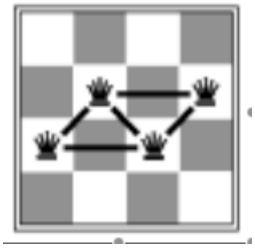
Instructions

- **The exam questions are worth a total of 100 points.**
- Hand in this sheet with your exam booklet(s).
- Remain seated in the exam room for the last 5 minutes of the exam.
- Answer all questions in the booklet provided. You may use pencil or pen, and write on the back of pages in the booklet.
- If you require clarification of a question or an additional booklet, please raise your hand.
- Write your name on all booklets that you use.

Questions

1. **True or False (2 points each (10 pts)):**
   You may provide a brief justification for your answer (for part marks)
   (a) (T / F) For first-order logic, resolution is a complete inference rule (i.e. all statements entailed by a knowledge base may be produced using resolution).
   (b) (T / F) Decision trees may be used to express *any* function of their inputs.
   (c) (T / F) Partial-order planning is a modification of forward chaining that allows multiple solution paths to be represented within a single plan.
   (d) (T / F) Constraint satisfaction problems differ from standard search problems in that the goal state contains additional information.
   (e) (T / F) In probability theory, an *atomic event* is a complete specification of the value of all state variables.

2. **Constraint Satisfaction (10 pts)**
   Consider the n-queens problem show below.



(a) (7) Briefly define the board above as a constraint satisfaction problem, and then solve it using backtracking search. Make sure to apply the most-contrainined variable, minimally constraining value, and forward checking heuristics.

(b) (3) Define the min-conflicts algorithm, and show two steps of the min-conflicts algorithm being applied to the same n-queens problem above.

3. **Logic Programming (12 pts)**

(a) (5) Convert the following situation into a Prolog program (define relationships as given!).

The law says that it is a crime for a Canadian to sell beer to allies of the nation Ohoh. All beer acquired by allies of Ohoh was sold to it by Colonel Molson, who is Canadian. A nation's allies include the *allies of its allies* (i.e. alliance is defined *transitively*). Nono is an ally of Soso, and Soso is an ally of Ohoh. The country Nono has some beer.

(b) (7) Show how Prolog would process the query `criminal(molson)` for the program you've written to represent the situation above. Be sure to show how the Prolog would apply rules and unify terms (you may use a tree to do this).

4. **Logical Inference (16 pts):**

(a) (4) First-order logic is said to be *semi-decidable*. What does this mean, and what causes first-order logic to be *semi-decidable* whereas propositional logic is *decidable*?

(b) (5) Convert the following first-order statements to conjunctive nor-

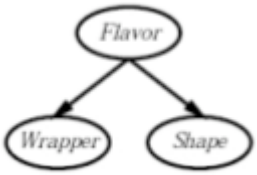
mal form.

1. $\forall x\ [\forall y\ Animal(y) \rightarrow Loves(x, y)] \rightarrow [\exists y\ Loves(y, x)]$
2. $\forall x\ [\exists y\ Animal(y) \wedge Kills(x, y)] \rightarrow [\forall z\ \neg Loves(z, x)]$
3. $\forall x\ Animal(x) \rightarrow Loves(Jack, x)$
4. $Kills(Jack, Tuna) \vee Kills(Curiosity, Tuna)$
5. $Cat(Tuna)$
6. $\forall x\ Cat(x) \rightarrow Animal(x)$

(c) (7) Using the above knowledge base and resolution, show that "Curiosity killed the cat (named Tuna)." You may use a tree to represent your proof.

5. **Planning (11 pts)**
   (a) (5) Provide an example of a *partial order plan*. Explain how ordering constraints and causal links are represented, and the structure of the plan itself.
   (b) (6) Give a simple example of a STRIPS search problem specification. Demonstrate the application of an action (and its effect on state), and also indicate what methods may be used to create plans from STRIPS specifications.

6. **Uncertainty and Bayesian Networks (17 pts)**
   (a) (4) Show the *structure* of the joint probability distribution table corresponding to the following Bayesian network containing boolean variables (you do not have to provide probability values).



   (b) (5) Show how the probability $P(Flavor = true \mid Wrapper = true)$ may be computed directly from the joint probability distribution table. Use a formula to unambiguously indicate which entries are used in the computation, but you do not have to provide or compute probability values.
   (c) (5) Give a formula representing how the probability

$$P(Flavor = true, Wrapper = false, Shape = true)$$

would be computed using the Bayesian network above. How does this differ from how this computation would be done using the full joint probability distribution and no independence assumptions?

(d) (3) In a Bayesian network, how are probabilities for continuous variables with both continuous and discrete parent variables represented? Also, provide an example.

7. **Inductive Learning and Decision Trees (16 pts)**
   In the following questions, you are welcome to use diagrams as part of your answer.
   (a) (3) In inductive learning, what is the form of a *hypothesis* that is learnt? What are the desirable properties of a learnt hypothesis?
   (b) (4) Define mathematical information (also known as *entropy*). Provide both the mathematical definition as well as a brief informal description.
   (c) (5) Provide pseudo code for the decision tree learning algorithm, and briefly summarize how it produces a tree from a set of labeled instances.
   (d) (4) Define *information gain*, and demonstrate how it may be used for decision tree induction.

8. **Neural Networks (8 pts)**
   (a) (4) Explain the construction of a basic unit ('neuron') in a neural network, and how inputs to the neuron determine its output value (use diagrams and formulae as needed).
   (b) (4) Provide pseudo code for the perceptron learning algorithm, and briefly explain how it may be used to train a single neuron. Are there any limitations to what can be learned by this algorithm?