Foundations of Intelligent Systems
CSCI-630-01 (Fall 2016)

Final Examination, Fri. Dec 16, 2016
Instructor: Richard Zanibbi, Duration: 120 Minutes

Name: ____

Instructions

- **The exam questions are worth a total of 100 points.**
- The exam is closed book and notes.
- Remain in the exam room if you finish during the final five minutes of the exam, and close the door behind you quietly if you leave the exam early.
- After the exam has started, once a student leaves the exam room, they may not return to the exam room until the exam has finished.
- Place any coats or bags at the front of the exam room.
- If you require clarification of a question, please raise your hand, and remain seated.
- You may use pencil or pen, and write on the backs of pages in the booklets.
- **Additional pages are provided at the back of the exam - make sure you clearly indicate where answers to each question may be found if you use these pages.**

Questions

1. **True/False (10 points)**

   (a) (T / F) In a proof by contradiction (e.g., using resolution), we prove α by demonstrating that for knowledge base KB, no model satisfies both KB and $\neg\alpha$.

   (b) (T / F) Predicate logic is *decidable*.

   (c) (T / F) The MLP backpropogation learning algorithm is based upon simulated annealing.

   (d) (T / F) It is possible to combine Decision Trees with other machine learning techniques by embedding these other machine learning algorithms (e.g., MLP) at leaves in the tree.

   (e) (T / F) The two main we tasks we used machine learning for this semester were 1) defining parameters for functions that select a choice, and 2) defining parameters for functions that score/rank alternatives.

   (f) (T / F) Noam Chomsky created a computer program for playing checkers that included a variant of the minimax algorithm, an end-game database and machine learning for the board evaluation function, publishing a paper on the topic in 1959.

   (g) (T / F) AdaBoost can be used to create an ensemble using any type of binary classifier.

   (h) (T / F) Alan Turing proposed that the question "Can machines think?" was vacuous (i.e., devoid of interest or meaning).

   (i) (T / F) Progress in AI applications often comes from simplifying problems, thereby making search spaces smaller and/or less uncertain.

   (j) (T / F) In practice, Artificial Intelligence algorithms are always to be preferred over brute force algorithms when solving a problem by computer.

2. **Miscellaneous Topics (20 points)**

   (a) (4) List the four parts of an *incremental* search problem definition.

   (b) (2) List the three parts of a *local* search problem definition.

(c) (2) How do we compute the number of independent entries (or *degrees of freedom*) in a probability distribution?

(d) (4) For what problems are knowledge-based systems (i.e., using logical inference) more appropriate than machine learning, and why?

(e) (4) For what problems is machine learning more appropriate than knowledge-based systems, and why?

(f) (4) Why are *independence assumptions* important for AI systems that utilize probabilities?

3. **Logic (24 points)**

(a) (4) Name an example for each of the following.

i. A *sound* inference rule.
ii. A *complete* inference algorithm for propositional logic.

(b) (10) Convert the knowledge base below to Conjunctive-Normal Form (CNF), and then show that R is entailed by the knowledge base using *resolution*.

$$P$$
$$(P \wedge Q) \rightarrow R$$
$$(S \vee T) \rightarrow Q$$
$$T$$

(c) (10) Translate each of the following sentences into predicate logic, and then prove that '*John likes peanuts*' using *backward chaining*.

i. John likes all kinds of food.
ii. Apples are food.
iii. Chicken is food.
iv. Anything anyone eats and isn't killed by is food.
v. Bill eats peanuts and is still alive.
vi. Sue eats everything Bill eats.

4. **Decision Trees and AdaBoost (20 points)**

(a) (4) Provide an example of the simplest decision tree possible for a two-class classification problem. Which probability is used to select the output class in this case?

(b) (4) How do we define possible splits for a continuous (i.e., real-valued) attribute, and how do we identify the best split?

(c) (4) *Chi-squared pruning* is used to prevent over-fitting in decision trees. What does the *Chi-square test* compare between a node and its children, and when does pruning occur?

(d) (2) What common problem with machine learning algorithms does AdaBoost often avoid?

(e) (2) What is different about how AdaBoost handles training samples versus decision trees or Multi-Layer Perceptrons?

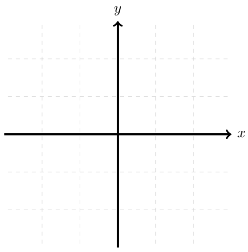
(f) (2) Which error function is the decision tree induction algorithm designed to minimize (Note: entropy is not an error function)?

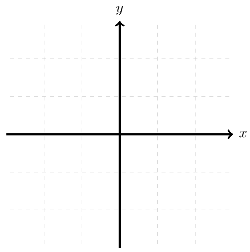
(g) (2) Which error function is AdaBoost designed to minimize at each iteration?

5. **Neural Networks (26 points)**

(a) (6) Provide the standard sigmoid (logistic) and threshold (step) activation functions below, and sketch the functions, labeling the y-axis so that minimum and maximum output values are clear.

$sigmoid(x) =$ $threshold(x) =$





(b) (2) In a neural network, what does x represent in the functions above, and how is the value of x computed?

(c) (2) What is the purpose of the *bias* input for a neuron?

(d) (5) Draw a single layer perceptron network (one input and one output layer) for representing the XOR function. Explain why this network cannot represent the XOR function.

(e) (5 points) Draw an example of a neural network that can compute the XOR function. **Make sure to include all connection weights, and provide the activation functions used in your network.**

(f) (4) For an MLP *hidden* node h with output o_h connected to output nodes o_1 and o_2 with errors δ_{o1} and δ_{o2}, the error δ_h is computed as:

$$\delta_h = (\delta_{o1} \times w_{ho_1} + \delta_{o2} \times w_{ho_2}) \times o_h \times (1 - o_h)$$

i. What does $(\delta_{o1} \times w_{ho_1} + \delta_{o2} \times w_{ho_2})$ represent?

ii. What does $o_h \times (1 - o_h)$ represent?

(g) (2) Which error function is the backpropagation algorithm designed to minimize?

(Bonus: +2) For a given problem, describe how to define admissible heuristics for use with the A* algorithm.

Additional Space

Additional Space

Additional Space

Additional Space