Introduction to Artificial Intelligence
4005-750-01 (Spring 2013)

Final Examination, May 15, 2013
Instructor: Richard Zanibbi, Duration: 120 Minutes

Name: _____

Instructions

- **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.**
- **The exam questions are worth a total of 100 points.**
- The exam is closed book and notes.
- Place any coats or bags at the front of the exam room.
- If you require clarification of a question, please raise your hand
- 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 - please clearly indicate where answers to each question may be found.**
- **Please close the door behind you quietly if you leave before the end of the examination.**

Questions

1. **True/False (5 points)**

   (a) (T / F) Back-propagation, logistic regression, and the perceptron algorithm all minimize the sum of squared error of node/network output.

   (b) (T / F) Entropy is measured in *bits*, the number of binary decisions needed to predict the answer to a question with n uncertain (probabilistic) outcomes. For outcomes $v_1, v_2, \ldots, v_n$ entropy is defined as:

   $$H(P(v_1), \ldots, P(v_n)) = -\sum_i P(v_i) \log_2(P(i)).$$

   Shannon's model suggests that the information content of an answer *increases* with the randomness of the distribution of outcomes from which an observation is taken.

   (c) (T / F) Samuel 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.

   (d) (T / F) Newell won a Nobel Prize for his work in economics related to the concept of 'satisficing,' the idea that agents (e.g. people) will seek goals until an acceptable rather than optimal outcome is obtained.

   (e) (T / F) Nearly all optimization algorithms considered in our course are local algorithms. Ironically, by using local search they may return sub-optimal solutions at local minima of an error function, but this is tolerated due to the often complex structure of the error function.

   (f) (T / F) Resolution is more powerful than forward or backward chaining for showing entailment in propositional logic.

   (g) (T / F) All functions may in theory be represented in a table, mapping inputs to outputs, such as for selecting actions based on the history of percepts collected by an agent.

   (h) (T / F) The minimax algorithm is optimal (i.e. achieves the highest payoff) against any opponent in two-player strategic games.

   (i) (T / F) Conditional and absolute independence allow us to determine which entries in a joint probability distribution table should be increased or decreased to produce a more accurate distribution.

   (j) (T / F) Propositional logic represents the world as a series of propositions without internal structure, whereas predicate logic represents the world using a set of objects, for which logical predicates assert the truth of relationships between objects in the domain.

2. **Miscellaneous Topics (10 points)**

(a) (4) Name the four key components of a formal search problem definition.

(b) (2) When do we need to use the *expecti-minimax* algorithm rather than the minimax algorithm for two-player strategic games?

(c) (4) Why are random numbers used so frequently in this course?

3. **Logic (30 points)**

(a) (4) Convert the following statements to two sentences each: one in propositional logic, and one in predicate logic.

i. Tom has a nice laptop.

ii. Tom will go to the restaurant only when Mary goes to the restaurant.

(b) (6) Define and provide an example for each of the following.

i. Entailment.

ii. A *complete* inference algorithm.

(c) (10) The following is a propositional knowledge base representing relationships between available flavors at an ice cream store.

1. $Vanilla \vee Chocolate$
2. $Vanilla \rightarrow Strawberry$
3. $Chocolate \rightarrow CookieDough$
4. $Strawberry \rightarrow Pistachio$
5. $CookieDough \rightarrow Pistachio$

i. (4) Convert the knowledge base to conjunctive normal form (CNF).

ii. (6) Prove that Pistachio ice cream is available using resolution. (Hint: resolution proofs are a form of *proof by contradiction*). You may use a proof tree or a list of statements.

(d) (10) The following Prolog program represents a legal matter involving Canada.

```
ally(spain,china).
ally(china,belgium).
ally(X,Z) :- not(X=Z), ally(X,Y), ally(Y,Z).

has(spain,beer).
canadian(colonel_molson).

criminal(X) :- sold(X,beer,Y), canadian(X), ally(Y,belgium).
sold(colonel_molson,beer,Y) :-  has(Y,beer), ally(Y,belgium).
```

i. (3) Given this knowledge base, will Prolog say that the query `ally(canada,belgium)` is true, false or unknown, and why?

ii. (7) Show how Prolog would process the query `criminal(A)` for the program. You may use a tree such as the ones seen in class to illustrate the execution and unifications (variable bindings).

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

(a) (7) **Briefly** explain how the next attribute to select for splitting at a node is chosen during decision tree construction. Also, identify the three stopping conditions, when the algorithm will *not* split the training samples associated with a node in the tree.

(b) (5) *Chi-squared pruning* may be used to prevent over-fitting by pruning a decision tree after its construction. At a node whose children are being considered for pruning, what difference does the Chi-square Δ measure?

(c) (3) What does it mean when we say that a Chi-square Δ value is statistically significant, with a p value of 0.05 (5%)?

(d) (5) The AdaBoost algorithm is used to crate an ensemble (a set) of classifiers that work together to make classification decisions, where classifiers are trained one-at-a-time.

i. (1) What common problem for machine learning algorithms does AdaBoost often avoid? (Hint: this was probably the most surprising result when the algorithm appeared)

ii. (2) What is different about how AdaBoost handles training samples, vs. a regular decision tree or the backpropagation algorithm?

iii. (2) How are the decisions of the individual classifiers (e.g. decision trees) combined to make a final classification decision?

5. **Linear Regression and Classification (20 points)**

(a) *SimpleFloat* prices its motor boats based on the boat length ($500/foot) and top speed in miles per hour of the boat ($100/mile per hour). The base price of a boat before considering length and top speed is $2,000.

i. (2) Provide a function *cost* that defines boat pricing using a linear model.

ii. (8) Draw a picture illustrating the inputs and outputs for a linear regressor ('node') that could be used to learn the weights of your pricing model using *gradient descent*. Briefly state how weights for the regressor are initialized. **Finally, show how error is computed, and how weights are updated after a single training sample has been passed through the regressor.**

iii. (8) Colonel Molson would like to buy a boat from *SimpleFloat*, for less than $8,000.

A. (3) Provide a formula defining a linear classifier that determines whether Colonel Molson will consider a boat to be *expensive* or *affordable*. **Note: the formula should use a threshold ('step') activation/output function.**

B. (5) Sketch the **weight vector** of the linear model and the decision boundary between the *expensive* and *affordable* classes in 2D, using labeled axes.

iv. (2) Which algorithm studied in class can be used to automatically learn the weights of our model, and what well-known limitation does the algorithm have?

6. **Neural Networks (15 points)**

(a) (4) Draw a multi-layer perceptron with 3 input nodes, 2 hidden nodes, and 3 output nodes.

(b) (4) Identify two situations in which it is better to use a multi-layer perceptron (MLP) rather than a logistic regressor to learn a classification function from data.

(c) (2) How do the functions that may be represented by an MLP with one hidden layer consisting of 100 hidden nodes differ from one with 10 hidden nodes?

(d) (5) Are MLPs regressors or classifiers? Explain your answer.

Additional Space

Additional Space

Additional Space

Additional Space