Foundations of Intelligent Systems
CSCI-630-01 (Fall 2015)

Final Examination, Fri. Dec 18, 2015
Instructor: Richard Zanibbi, Duration: 120 Minutes

Name: _____

Instructions

- **The exam questions are worth a total of 100 points.**
- **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.**
- **Remain in the exam room if you finish during the final five minutes of the exam.**
- **Close the door behind you quietly if you leave before the end of the examination.**
- 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 - clearly indicate where answers to each question may be found.**

Questions

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

   (a) (T / F) Shannon's entropy metric suggests that the information content of a prediction *increases* as the probability distribution for possible outcomes becomes more random.

   (b) (T / F) Training algorithms for multi-layer perceptrons, logistic regression, and the (classical) perceptron algorithm minimize the sum of squared error for output values.

   (c) (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.

   (d) (T / F) Nearly all optimization algorithms that we studied in the course, including genetic algorithms, min-conflicts (used to solve the 8-queen's problem), simulated annealing, decision tree learning, and backpropagation are variations of hill climbing.

   (e) (T / F) ...(continued from previous question)...and a common variation is to use randomness in some way to increase the amount of exploration during the search for optimal parameters, in the hope of avoiding local maxima or minima.

   (f) (T / F) Predicate logic is *semi-decidable*.

   (g) (T / F) Bayes' rule is defined by: $P(A, B) = P(A|B)P(B)$.

   (h) (T / F) Turing created his imitation game as a way to determine when human intelligence had been acquired by a machine, i.e. a true Artificial Intelligence.

   (i) (T / F) In practice, for problems with small search spaces and little or no noise in variable values, a brute-force solution may be preferable to an 'intelligent' solution.

   (j) (T / F) A logical statement α is *satisfiable* when every combination of truth values that makes a knowledge base true also makes α true.

2. **Miscellaneous Topics (18 points)**

   (a) (4) Define what it means for an agent to act *rationally*.

   (b) (2) Name the four parts of an *incremental* search problem definition.

(c) (4) Identify how the parts you identified in part (b) change for each of the following.
- Game search
- Local search

(d) (2) Which type of attributes is a decision tree better suited for than a neural network?

(e) (3) Who coined the term 'Artificial Intelligence?'

(f) (3) When is it necessary to use the *expectiminimax* algorithm in a game playing program?

3. **Logic (26 points)**

(a) (6) Convert each of the following statements into 1) a sentence in propositional logic, and 2) a sentence in predicate (first-order) logic.

   i.   Gary has a nice cell phone.
   ii. All people that attend RIT are fun.
   iii. If Gary goes, Mary goes, and vice-versa.

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

    i.   A *complete* inference algorithm.

    ii.   A *constant* in predicate logic.

    iii. A contradiction.

(c) (14) The knowledge base below represents a Canadian legal matter.

1. $\forall x, y, z \; ally(x, y) \wedge ally(y, z) \wedge \neg(x = z) \rightarrow ally(x, z)$
2. $\forall x, y \; sold(x, Beer, y) \wedge canadian(x) \wedge ally(y, Belgium) \rightarrow criminal(x)$
3. $\forall x \; has(x, Beer) \wedge ally(x, Belgium) \rightarrow sold(ColonelLabatt, Beer, x)$
4. $ally(Spain, India)$
5. $ally(India, Belgium)$
6. $has(Spain, Beer)$
7. $canadian(ColonelLabatt)$
8. $\neg(Spain = Belgium)$

i. (6) Use *forward chaining* to prove that Colonel Labatt is a criminal ($criminal(ColonelLabatt)$). Make sure to show the unification of variables at each step of your proof.

5

ii.   (8) Convert the *rules* from the knowlege base to Conjunctive Normal Form (CNF), and then prove $criminal(ColonelLabatt)$ using *resolution*. Again, make sure to show the unification of variables at each step of your proof.

4. **Decision Trees and AdaBoost (18 points)**

   (a) (3) Identify the three stopping conditions (base cases) for the decision tree learning algorithm. These cases are where the algorithm decides *not* to split the training samples at a node.

   (b) (4) *Chi-squared pruning* can be used to prevent over-fitting, by reducing the size of a decision tree after its construction. The *Chi-square* metric computes a difference between two probability distributions. At a node whose children are being considered for pruning, specifically which two distributions are being compared? After this comparison is made, when will pruning be applied?

   (c) (4) Using the *entropy* formula, show how to compute the *entropy* for a distribution of 10 people: 8 have a sandwhich (+), and 2 do not have a sandwhich (-). **You do not need to compute the final value.**

   (d) (2) Why does the decision tree learning algorithm split samples using the attribute that reduces entropy the most?

(e) AdaBoost creates an ensemble of classifiers that work together to make classification decisions, with member classifiers being trained one-at-a-time.

    i. (3) What is different about how AdaBoost handles training samples versus other machine learning algorithms such as regular decision trees or the backpropagation algorithm?

    ii. (2) What determines the weight of each classifier in the vote to select the final classification of an AdaBoost classifier?

5. **Linear Regression and Neural Networks (28 points)**

(a) *EZRide* prices its cars based on interior size ($100/cubic foot) and top speed in miles per hour of the car ($50/mile per hour). The base price of a car before considering the size of the interior and top speed is $1500.

    i. (2) Provide a linear model for the cost of an *EZRide* car.

    ii. (4) Over a ten year period, *EZRide* changes their pricing. We have obtained a long list of (cubic feet, top speed, actual car price) samples in a spreadsheet. Assuming that we want to minimize the *sum of squared differences* between our model's predictions and actual car prices, how can we accurately estimate the new prices *without* machine learning?

(b) (6) Draw a *linear regressor* 'neuron' for our EZRide model. Identify the inputs, the outputs, and where the prices in the model are represented in the neuron.

(c) (4) What would you need to change in order for the linear regressor in the last question to become a *logistic regressor?*
   - Also, what would the range of output values be after making this change?
   - And what is the derivative of the output?

(d) (4) Define the problem of fitting the weights in a neural network as a modified local search problem. For a multi-layer perceptron (MLP), what part of the search problem definition does backpropagation address?

(e) (8) When training an MLP to classify three different types of cat image, how are 'correct' outputs represented in the training data? Also,

- How likely is the MLP to produce this target outputs in the training data exactly?
- Why is it necessary to use backpropagation to compute nodes in layers other than the output layer?

(Bonus: +2) In what order should you read the sections of a research paper?

Additional Space

Additional Space

Additional Space

Additional Space