

Foundations of Intelligent Systems

CSCI-630-01 (Spring 2014)

Final Examination, Wed. May 21, 2014
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 (5 points)

- (a) (T / F) The minimax algorithm is guaranteed to be optimal (i.e. achieves the highest payoff) *only* against optimal opponents in two-player strategic games.
- (b) (T / F) Random Forests and Multi-Layer Perceptrons are able to represent complex regression functions using combinations of simple models.
- (c) (T / F) Neural network research was slowed substantially in the late 1960's with the publication of Minsky and Papert's book *Perceptrons*, which demonstrated that a standard perceptron was incapable of learning when only one of its two inputs was on.
- (d) (T / F) Nearly all optimization algorithms considered in our course are incremental search algorithms.
- (e) (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, \dots, v_n entropy is defined as:

$$H(P(v_1), \dots, P(v_n)) = 1 - \sum_i P(v_i) \log_2(P(i)).$$

- (f) (T / F) Predicate logic is *decidable*.
- (g) (T / F) $P(A, B | C) = P(A | C)P(B | C)$ is an example of absolute independence.
- (h) (T / F) It is always possible to convert a predicate logic knowledge base to a finite propositional logic knowledge base.
- (i) (T / F) In practice, for problems with small search spaces and high certainty in observed data, a brute-force solution may be preferable to an 'intelligent' solution.
- (j) (T / F) Random Forests as discussed in class are *generative* models, while Multi-Layer Perceptrons are *discriminative*.

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

- (a) (2) Name the four key components of a formal search problem definition.
- (b) (4) Name the four (increasingly complex) agent types discussed in lecture and the textbook. For each agent type after the simplest one, in a single sentence identify which capability the agent type adds relative to simpler models. *All four of these agent types may become 'learning' agents, so do not include the 'general learning agent.'*
- (c) (4) Give a concrete example of a problem whose solution requires a combination of logic, search, and machine learning. Briefly identify how each is needed to address the problem.

3. Logic (30 points)

(a) (4) Briefly describe how ‘facts’ are represented in propositional versus first-order logic.

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

- i. A *sound* inference rule.
- ii. A *complete* inference algorithm.
- iii. A *satisfiable* statement.

4. **Decision Trees, AdaBoost and Random Forests (20 points)**

- (a) (6) Provide the formulas for *entropy* and *information gain*, and explain how they are used to select which attribute to split on at a node in a decision tree.
- (b) (2) Why is the decision tree learning algorithm prone to over-fitting the training data?
- (c) (3) *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?

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

(a) *EZRide* prices its cars based on interior size (\$50/cubic foot) and top speed in miles per hour of the car (\$100/mile per hour). The base price of a car before considering the size of the interior and top speed is \$500.

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

ii. (4) Now suppose that over a few years, *EZRide* changes their base price, interior size and speed costs. Given a sufficient set of (cubic feet, top speed, car price) triples, we can use linear regression to estimate the new parameters of the cost function.

Provide a diagram showing the inputs and outputs of a linear regressor that can be used to learn the price model parameter weights using gradient descent.

iii. (4) Provide pseudo code for the gradient descent algorithm that will be used to learn the new weights. Make sure to identify how weights are initialized and updated.

- iv. We would like to buy a car from *EZRide*, and have exactly \$7,000 available (and no more). The current pricing has a base cost of \$500, \$100 per cubic foot and \$100 per mile/hour in the top speed of the car.
- A. (2) Provide a formula to determine whether we can afford to buy an *EZRide* car or not, given the size of the interior and top speed of the car.
- B. (6) Sketch the **weight vector** of the linear model and the decision boundary between the *affordable* and *unaffordable* classes in 2D, using labeled axes.
- v. (2) Which error function is commonly used for linear regression?

6. Machine Learning (15 points)

(a) (2) Define *over-fitting*.

(b) (4) Explain how over-fitting can be prevented when training a Multi-Layer Perceptron.

(c) (4) Explain how over-fitting is avoided in Random Forests.

(d) (5) Define *regression* and *classification* functions, and discuss their relationship.

(e) **Bonus** (2) Why is it important not to evaluate the performance of a machine learning algorithm using its training data?

Additional Space

Additional Space

Additional Space

Additional Space