SPRING 2008
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, February 4, 2008
3:00 – 7:00 p.m.

GENERAL INSTRUCTIONS:

(a) Answer each question in a separate book.

(b) Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On *one* of your books, list the numbers of *all* the questions answered. *Do not write your name on any answer book.*

(c) Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer:

- both questions in the section labeled B760, corresponding to your chosen focus area, *and*
- any two additional questions in the sections Bxxx, where these two questions need *not* come from the same section, *and*
- both questions in the section labeled A760.

Hence, you are to answer a total of six questions.

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced that a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

Answer <u>both</u> of the questions in the section labeled B760. Also answer any <u>two</u> additional questions in any of the other B sections (these two questions need NOT occur in the same section).

B760 – MACHINE LEARNING: BASIC QUESTIONS

B760-1 Evaluating Learning Methods

You are in charge of leading an international consortium to analyze an environmental dataset. You have 10,000 independent labeled examples for a two-class classification problem. You may assume this training set is representative of future examples that will be seen. Examples are feature vectors with 100 features and no missing values. Many different groups in the consortium want to try different learning algorithms. At the end of the consortium's analysis, you are to write a paper proposing one classification model and estimating what its accuracy will be on new data.

(a) How will you choose the best learning algorithm among those advocated? Be specific in describing your methodology.

(b) Having chosen the best algorithm, how will you estimate its future accuracy?

(c) Describe at least one alternative to your proposed method that also would be methodologically sound.

(d) Why do you prefer your method in (a) and (b) over the alternative in (c)?

B760-2 Supervised Learning

Two approaches for supervised learning are (i) learn a model from the training data that can predict the output for any point in feature space, e.g., decision-tree induction and (ii) use the training data itself as the representation of such a model, e.g., *k*-nearest neighbors.

(a) Using Venn diagrams, show how decision trees and *k*-nearest neighbors partition feature space. Explain your diagrams.

(b) Propose and justify <u>two</u> reasonably different ways to combine these two types (i.e., type *i* and type *ii*) of supervised learners. (Read Part c before you answer this question; in order to minimize overlap in your answers.)

(c) Discuss what you feel is the *most important* strength of each of your designs in Part b *compared* to the other design.

(d) Imagine now that at learning time you are also given the complete testset of unlabeled examples (your learned model will be discarded after it labels this testset; i.e., you need not worry about it working well on additional testset examples). Present a hybrid (of types *i* and *ii*) approach designed to exploit the properties of this particular scenario.

B766 – COMPUTER VISION: BASIC QUESTIONS

B766-1 Image Alignment and Panorama Construction

(a) There are two situations where we can exactly align two images using a homography transformation. What are these two situations? What is a homography transformation? How many free parameters does it have? To compute a homography, how many feature point correspondences do we need to establish?

(b) Outline the basic procedure for establishing feature point correspondences using the RANSAC method.

(c) Can we use a homography transformation to create a 360-degree panorama? Explain briefly why or why not. If not, how can a solution be computed?

B766-2 Visual Correspondence

(a) Describe a basic procedure for window matching used for stereo vision. What are the pros and cons of using bigger vs. smaller windows?

(b) Describe a global optimization approach to stereo matching. List two optimization algorithms that are widely used to solve the stereo matching problem.

(c) In both optical flow and stereo, a necessary step is to establish pixel correspondences between two images. What is a key difference in how correspondence search is constrained in these two problems?

(d) When we solve optical flow and stereo matching problems, a common assumption made is "brightness constancy." Explain what this means and describe two situations when it is violated (besides the imaging noise issue). Identify at least one similarity measure that is robust in situations when brightness constancy is violated.

B776 -- ADVANCED BIOINFORMATICS: BASIC QUESTIONS

B776-1 Biological Motif Finding

A biologist approaches you with a short human genomic sequence that is believed to be bound the transcription factor SPI-B. The biologist gives you a position weight matrix (PWM) of length seven for the binding site motif of this protein.

(a) Sketch an HMM for modeling a DNA sequence with zero or more occurrences of this motif.

(b) The biologist asks you to determine *how many* occurrences of this motif are present in the sequence of interest. Describe one approach for answering this question with your HMM. Give the strengths and weaknesses of this approach.

(c) Describe a second approach for answering the biologist's question in (b). Give the strengths and weaknesses of this approach.

B776-1 Anytime Clustering

An *anytime algorithm* is one that (i) can be interrupted at any time and will return a solution, but (ii) will continually try to find better solutions as it is allowed to run longer. Suppose you have a large amount of computing power available to exploit in finding good clusterings, and you are willing to let your method run for hours or even days.

(a) Describe how you would adapt standard *k*-means clustering to be an anytime algorithm.

(b) Describe how you would adapt standard hierarchical clustering to be an anytime algorithm.

B838 -- ADVANCED NATURAL LANGUAGE PROCESSING: BASIC QUESTIONS

B838-1 Latent Topic Models

Latent Dirichlet Allocation (LDA) is a model for a collection of text documents.

(a) In English, describe the generative process of a single document under the LDA model. Clearly state each step. You do not have to give the formula.

(b) Briefly discuss two major differences between:

1. LDA with k hidden topics
2. Text classification with k classes

(c) Sometimes we might have prior knowledge about the hidden topics. Consider the case where we prefer the first hidden topic to have words $w_1 \ldots w_m$ in it. Briefly discuss how you might do this.

B838-2 Information Theory

The Kullback-Leibler (KL) divergence measures the difference between distributions.

(a) Given two unigram probability distributions p, q over the same vocabulary, define the KL-divergence from p to q.

(b) A distance metric d(x,y) satisfies four conditions:

1. $d(x,y) \geq 0$
2. $d(x,y)=0$ iff $x=y$
3. $d(x,y)=d(y,x)$
4. $d(x,y) \leq d(x,z)+d(z,y)$

Is KL-divergence a distance metric over distributions? Be sure to justify your answer.

(c) Let us assume there is a corpus which is generated from the underlying unigram distribution p. Given the corpus, we want to find the Maximum Likelihood Estimate (MLE) unigram distribution p^{MLE}. State the connection between finding the MLE and the KL-divergence, when the corpus size goes to infinity.

Answer <u>both</u> of the questions in the section A760.

A760 – MACHINE LEARNING: ADVANCED QUESTIONS

A760-1 Prediction with Time Series Data

Consider a problem domain in which we are interested in learning a model to predict the recurrence of some type of event. For example, perhaps we are interested in predicting which patients with a particular medical condition (e.g. cancer) are likely to have a recurrence of the condition, or when they are likely to have a recurrence. Assume you are given a data set that consists of records of the form $(p_i, t_i, x_{i,1}, \ldots x_{i,n})$ where p_i represents a patient, t_i represents the time at which the condition occurred, and $x_{i,1}, \ldots x_{i,n}$ represent features characterizing the patient at the given time. Note that there will be some patients who will have recurrences that aren't represented in our data set, simply because they haven't happened yet.

(a) Describe how you would frame this problem as a *classification* task.

(b) Describe how you would frame this problem as a *regression* task.

(c) How would you evaluate the accuracy of your approach? Be sure to discuss aspects of your evaluation methodology that pertain to the data coming from a time series.

(d) Now suppose we want to train a model using all of the data, or as much of it as possible (even recent occurrences of the condition). Describe how you would adapt <u>either</u> the approach described in (a) or the approach described in (b) so that you could use nearly all of the records for training.

A760-2 Reinforcement Learning, Genetic Algorithms and Relational Learning

Consider applying a genetic algorithm (GA) to a reinforcement-learning (RL) task, involving two robots playing a game, and where:

- the states are represented by N real-valued features, relating the robots to one another and other portions of the game, denoted $f_i(game\#, step\#)$, $1 \leq i \leq N$
- rewards are produced by the function $r(game\#, step\#)$
- at each step, the robots simultaneously choose either action 1 or action 2
- your task is to train the policy for robot A; robot B's policy is fixed

(a) Design a GA approach for this task, where the model(s) learned are in *first-order predicate calculus*. (Your algorithm need not be an incremental learner. It is fine if you assume that after every M games you collect a set of relevant training examples and execute your GA in "batch" mode, then use your learned model for the next chunk of games.) Be sure to justify your design decisions.

(d) Does your answer to Part (a) use *policy iteration* or *value iteration*? Explain.

(e) If your answer to Part (a) used policy iteration, explain what would need to change for it to employ value iteration; vice versa if your Part (a) design used value iteration.

This page intentionally left blank. You may use it for scratch paper. Please note that this page will NOT be considered during grading.