SPRING 2004
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, February 2, 2004
3:00 – 7:00 p.m.
Room 3345 Engineering Hall

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 (2) questions in the section labeled B760 or B766, corresponding to your chosen focus area, *and*

- any two (2) additional question in the sections B731, B760, B766, and B776, where these two questions need *not* come from the same section, *and*

- both (2) questions in the section labeled A7xx that corresponds to your focus area.

Hence, you are to answer a total of *six* (6) 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 both (2) of the questions in the section labeled B7xx that corresponds to your chosen focus area. Also answer any two (2) additional questions in any of the other sections (these two questions need NOT occur in the same section).

B731 – ADVANCED AI BASIC QUESTIONS

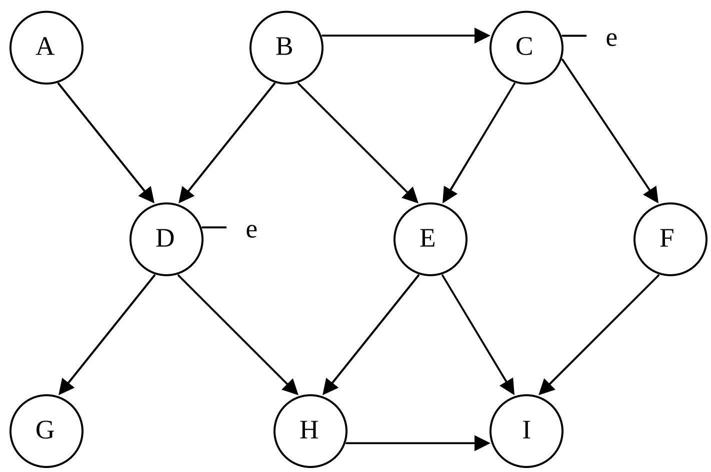
B731-1. Consider the *Bayesian network* below, with evidence given (i.e., values known) for variables *C* and *D*. Be sure to justify your answers to the questions below.

a) Define *d-separation*. Be sure to describe all of the possible connection types.

b) Is *E* d-separated from *F*?

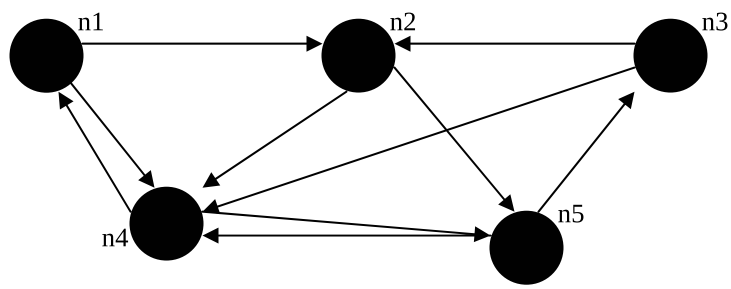
c) Is *A* d-separated from *H*?

d) Is *G* d-separated from *I*?



B731-2. Suppose you apply Quinlan's *FOIL* algorithm (as published in *Machine Learning*, 1990) to the following learning task. Each example is a directed graph. An example is labeled "positive" if and only if it contains as a subgraph a cycle of 3 nodes. An illustration of the chosen representation is given below.

If it affects your answers below, you may assume a probability distribution over examples where the number of nodes is chosen uniformly between 5 and 10, inclusive, and each possible edge has probability 1/3 of appearing.



Examples	**Background Knowledge**	
positive(g1).	*node(g1,n1).*	*arc(g1,n1,n2).*
	node(g1,n2).	*arc(g1,n1,n4).*
	node(g1,n3).	*arc(g1,n2,n4).*
	node(g1,n4).	*arc(g1,n2,n5).*
	node(g1,n5).	*arc(g1,n3,n2).*
		arc(g1,n3,n4).
		arc(g1,n4,n1).
		arc(g1,n4,n5).
		arc(g1,n5,n3).
		arc(g1,n5,n4).

a) Why might you expect Quinlan's FOIL algorithm to have difficulty with this task?

b) What *change* to the FOIL algorithm might permit it to perform better on this task (short of replacing FOIL's greedy search with a complete or exhaustive search)?

B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Pat Hand is assigned the task of improving a spam filter that is created using *supervised machine learning*. Pat adds 20% more features to the set used to describe individual email messages, but when Pat runs a 10-fold cross-validation experiment, the resulting test-set accuracy is 5 percentage points lower than before, even though the new features Pat added are meaningful ones (rather than irrelevant features). You may assume that all features are Boolean valued.

a) *Informally* explain why this might have happened.

b) Explain, using the ideas from *PAC learning*, why accuracy can become lower even though additional informative features are used. Be explicit and precise in your answer.

c) Sketch the algorithm for using *backward selection* to choose a good set of features. What are one (1) major strength and one (1) major weakness of this approach?

d) Explain how one could employ *information theory* to choose a good set of features. What are one (1) major strength and one (1) major weakness of this approach?

B760-2. Briefly describe and explain the significance of the following key ideas in s*upport vector machines*:

a) *margins*

b) *slack variables*

c) *kernels*

For each of the above three concepts, what do you feel is the most similar idea from some other type of supervised learning. Be sure to justify your answers.

d) Most similar idea to *margins* in another approach to supervised learning?

e) Most similar idea to *slack variables*?

f) Most similar idea to *kernels*?

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1.

a) Show with a figure the epipolar geometry between two images taken by two different cameras. Include the epipoles, an epipolar plane, an epipolar line, the two cameras' optical centers, and a conjugate pair of points in the left and right images. Formally define epipolar line, epipole, and the epipolar constraint in terms of the fundamental matrix, **F**.

b) Describe the epipolar lines when the line through the two cameras' optical centers is parallel to (i) exactly one of the image planes, (ii) both image planes, and (iii) both image planes and both images' rows?

c) Consider the case where a single camera is translating, but not rotating and without changing any of its intrinsic parameters, with images taken at the initial position and the final position 10cm away. What are the homogeneous coordinates of the right image's epipole when (i) the translation is along the camera's optical axis, and (ii) the translation is parallel to the image's x-axis?

B766-2.

a) Describe the main steps for *segmenting* an image that contains regions distinguishable based on their texture and color, using (i) a threshold-based algorithm, and (ii) a graph-based algorithm such as normalized-cut.

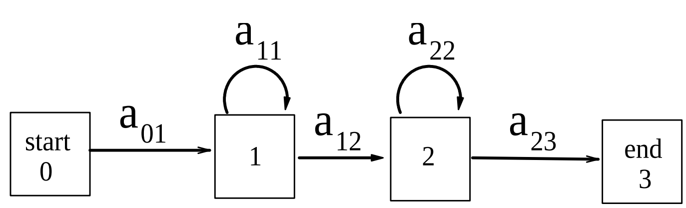
b) Describe one (1) major advantage and and one (1) major disadvantage of each of the two methods in (a) for segmenting this type of image.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. Consider the *hidden Markov Model* shown below. The *start* and *end* states are silent, but both *state 1* and *state 2* have emission parameters for the four DNA bases {**A**, **C**, **G**, **T**}.

Suppose that we are learning the emission and transition parameters using the *Forward-Backward* (Baum-Welch) algorithm. We are training using two sequences, **TCG** and **TC**.

Show how the Forward-Backward algorithm updates the transition probabilities for transitions emanating from *state 1* (i.e. a_{11} and a_{12}). You can assume that we have already run the Forward and Backward algorithms on both sequences and calculated all of the relevant *f* (Forward) and *b* (Backward) values for them. Be sure to explain any notation you use.



B776-2. Suppose we want to develop a *clustering* approach for instances that are represented by trees. For example, the instances might be parse trees of English sentences or trees describing RNA secondary structures. Describe an *Expectation-Maximation* (EM) based approach for this setting. In particular, answer the following questions:

a) How are clusters represented in your approach?

b) How are clusters initialized?

c) What calculations are done in the *E* step of your approach?

d) What calculations are done in the *M* step?

Answer both (2) of the questions in the section labeled A7xx that corresponds to your chosen focus area.

A760 – MACHINE LEARNING ADVANCED QUESTIONS

A760-1. Consider the task of learning a model for *ranking* instances. For example, suppose we are learning a model for ranking documents according to their relevance to some topic. The training data consists of a set of documents and a set of relations of the form, $d_i > d_j$, where d_i and d_j are both documents and the '>' indicates that d_i is more relevant than d_j. The $d_i > d_j$ relations specify a partial order over the documents in the training set.

a) Describe an approach for learning a ranking model from this type of training data. Describe both the representation and learning algorithm you would use.

b) Given a set of documents, describe how you would use your model to rank them.

c) Does your ranking procedure produce a partial or total ordering of the test-set documents? Does is produce a consistent ordering?

d) Describe one (1) significant strength and one (1) significant weakness of your approach.

A760-2. Consider the *bias-variance tradeoff* in machine learning.

a) Precisely define both *bias* and *variance* for the case of regression tasks.

b) Explain why there is a tradeoff between these two. If you wish, you may answer this question in the context of classification tasks.

c) Using the concept of the bias-variance tradeoff, explain why *bagging* often improves classifier performance.

d) Suppose you have a machine learning algorithm that has a parameter with which you can specify how tightly you wish to fit the training data (e.g., degree of pruning in decision trees, slack parameter in SVMs, etc.). If you are going to bag your classifiers, would you expect in general to get better results with a tight or loose fit to the training data? Explain your answer based on the bias-variance tradeoff.