SPRING 2006
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, January 30, 2006
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 or B766, corresponding to your chosen focus area, *and*
- any two additional question in the sections B731, B760, B766, and B776, where these two questions need *not* come from the same section, *and*
- both questions in the section labeled A7xx that corresponds to your focus area.

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

POLICY ON MISPRINTS AND AMBIGUITIES:

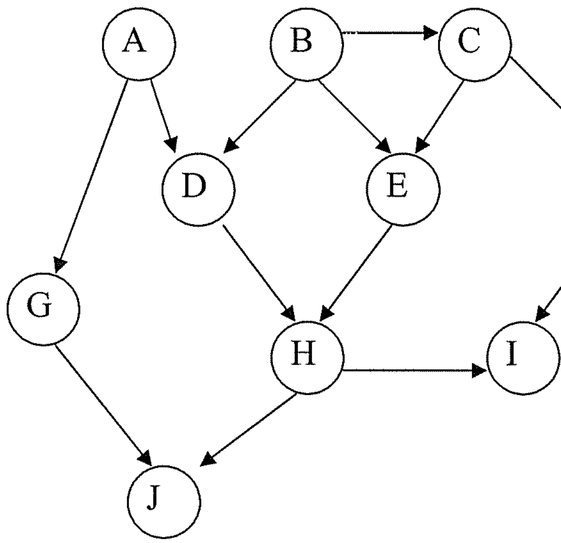
The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced that a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area 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 B7xx that corresponds to your chosen focus area. Also answer any <u>two</u> 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. Bayesian Network Inference

Both parts of this question refer to the Bayesian network below.



(a) Construct a <u>junction tree</u> for this Bayesian network. Be sure to show moralization and triangulation steps based on identifying simplicial nodes as soon as possible.

(b) Which nodes are **not** <u>d-separated</u> from node *G* given evidence at nodes *D*, *E* and *H* ? For each such node, give one path through which evidence can flow.

B731-2. Bayesian Network Learning

Suppose you wish to learn the parameters (CPT probabilities) for a Bayesian network from data. Assume the data has some missing values, and the Bayesian network includes some hidden variables. Suppose further that you have Dirichlet priors for all parameters.

(a) Discuss one advantage of Expectation-Maximization (EM) for this task over Gibbs sampling.

(b) Discuss one advantage of Gibbs sampling over EM for this task.

(c) Describe an alternative to both EM and Gibbs sampling for this task.

B760 – MACHINE LEARNING BASIC QUESTIONS

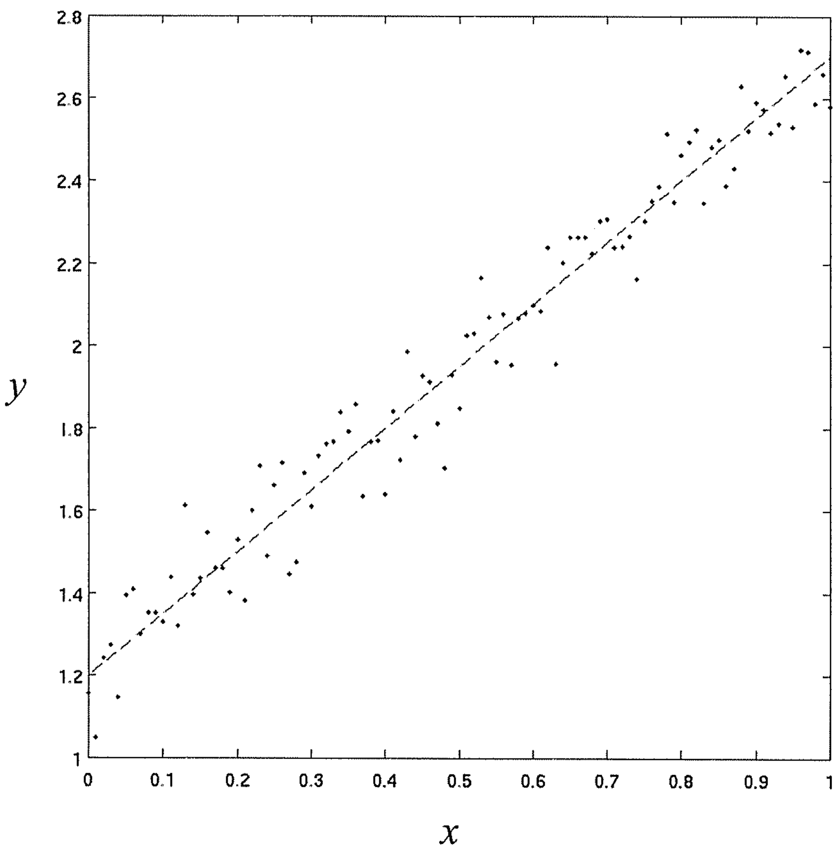
B760-1. *k*-Nearest Neighbor Learning

Consider the use of case-based learning, in particular *k*-nearest-neighbor (*k*NN) methods, in machine learning.

(a) *k*NN can be used for regression instead of classification. Explain how this is done; be sure to include pseudocode or a formula.

(b) One can extend standard *k*NN by weighting the contribution of each neighbor. Give one example of a method for assigning the weights. Describe one significant advantage and one significant disadvantage of this weighted variant.

(c) Consider using *k*NN for regression shown in the figure below. The input variable $x \in [0,1]$. The dashed line is the unobserved truth, while the dots are noisy observed data points. There is one major problem when we perform *k*NN regression for x close to 0 or 1. Explain what the problem is.



B760-2. Reinforcement Learning

A common approach to addressing the reinforcement learning (RL) task is to learn value functions.

(a) Explain what *Q(state,action)* estimates in traditional *Q* learning.

(b) Repeat part (a), but this time use SARSA learning. Be sure to explain the motivation of SARSA.

(c) Assume you choose to use a function approximator to learn the *Q* function for SARSA. Explain two different ways "supervised" training examples can be created, and briefly discuss one relative advantage of each.

(d) What is needed if one wants to learn the function *Value(state)* instead of *Q(state, action)*? How would actions be chosen once a good function for scoring states was learned?

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. Feature Detection

(a) Describe the main steps of Lowe's SIFT operator for detecting feature "interest" points in an image.

(b) What types of local features does this operator detect? That is, describe the extent to which image features such as points, lines, and corners are detected by this operator, and what distributions of the underlying image intensity values the operator is most sensitive to.

(c) What is the main advantage of implementing this operator using difference-of-Gaussians instead of using Laplacian-of-Gaussians?

(d) Briefly explain whether or not this operator is invariant to (i) translation, (ii) rotation, (iii) scale, and (iv) affine transformations.

B766-2. Object Recognition

Consider Turk and Pentland's method for appearance-based 2D object recognition called Eigenfaces.

(a) The Eigenfaces algorithm projects an input image to a point in a k-dimensional subspace called "face space." What does each dimension in face space correspond to? Give a method for selecting a good value for k.

(b) How could the algorithm be used for face *detection* instead of face recognition? That is, indicating if the input image is of *a* face or not, without regard to who it might be.

(c) Say you want to modify the algorithm so that it can be used for face *synthesis* instead of face recognition. Given three people whose faces correspond to three points in face space, $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$, specify how to synthesize a new virtual person's face that combines the three given faces equally. Your computation should *not* use the original face images, but can use the $\mathbf{W}_i$'s eigenvectors, $\mathbf{E}_i$'s, and the "average face" image, $\mathbf{A}$. Explain what one of the challenges is of using this method to synthesize realistic-looking faces.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. Active Learning with HMMs

In an active learning setting, a learner may selectively ask queries about unlabeled instances. In response to a query, an oracle (e.g. a human expert) will provide a label for an instance or some part of it.

(a) Suppose that you are training an HMM for a sequence classification task (i.e. each sequence is associated with a single label) and you have a large set of unlabeled sequences in addition to a small set of labeled sequences. Describe a method for selecting which unlabeled sequence you would most like to have labeled for training?

(b) Now suppose that you are training an HMM for a sequence segmentation or tagging task (i.e. each element in each sequence has a label) and you have a set of partially labeled sequences. Describe a method for selecting which *individual element*, from *any* of the sequences, you would most like to have labeled for training.

B776-2. Hierarchical Clustering

(a) Describe how one can obtain a <u>partitional</u> (i.e. flat) clustering from a distance-based hierarchical clustering. You can assume that a user specifies the number of clusters to return.

(b) Suppose that we want a <u>flat</u> clustering that is <u>soft</u> (i.e. the membership of an instance in a cluster is associated with a measure of confidence) and <u>disjunctive</u> (i.e. some instances truly belong to more than one cluster). Describe how you might adapt or employ a standard hierarchical clustering method to return this type of clustering.

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

A760 – MACHINE LEARNING ADVANCED QUESTIONS

A760-1. Co-training

Co-training is a learning algorithm that uses both labeled data and unlabeled data to build classifiers. It makes several assumptions about the features. In particular, it assumes the features can be divided into two disjoint subsets, so that each subset is sufficient to train a good classifier, and the subsets are conditionally independent given the class.

(a) Define the co-training algorithm using pseudocode.

(b) Explain why the subsets need to be conditionally independent. Give an example of what can go wrong if the subsets are conditionally dependent.

(c) If the features do not satisfy the assumptions (i.e. there is no natural division), but you have more than two different learning algorithms (e.g. *k*NN, neural net, SVM, decision tree, etc.), propose an algorithm that can use the unlabeled data to help learning.

A760-2. Transfer Learning

Assume you had previously used supervised machine learning to (separately) learn Boolean concepts *A*, *B*, *C*, *D*, and *E*.

(a) Now assume that someone has asked you to learn Boolean concept *Z* from a set of 1000 labeled training examples. This person also tells you that concept Z is most similar to concept *E*. Your task is to figure out how to leverage what you learned about concept *E* to better learn concept *Z*.

For any two of the three representations below, describe and motivate how you would use a model learned for concept *E* in the chosen representation when learning a model for concept *Z*.

i. A single decision tree
ii. A linear, weighted sum of features
iii. An ensemble of non-linear SVMs

The training examples for concept *E* and concept Z share some but not all of the same features.

(b) Describe an experimental methodology that would allow you to demonstrate in a statistically significant manner that your use of concept *E* improved the accuracy of the model you learned for concept *Z*.

(c) Now assume you are given training examples for concept *Y*, but are not told which of concepts *A*-*E* are most similar to it. Describe how you might choose which previously learned concept is likely to be the most helpful for learning concept *Y*.

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