SPRING 2007
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, February 5, 2007
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 questions 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 both of the questions in the section labeled B7xx that corresponds to your chosen focus area. Also answer any two 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 Learning Bayesian Network Structure

(a) Give pseudocode for the *sparse candidate algorithm* for learning Bayesian network structure.

(b) Give two different widely used scoring functions for Bayesian network structures. Name one advantage of each scoring function over the other.

(c) Describe or show at least one Bayesian network that the sparse candidate algorithm will have trouble learning using either scoring function.

B731-2 Conditional Independence and *d*-Separation

In graphical models, *d-separation* rules determine whether two sets of variables A, B are d-separated by a third set C. Consider the simple case when each set has a single node: $A=\{a\}$, $B=\{b\}$, $C=\{c\}$. There are 3 basic cases in d-separation.

(a) Draw the Bayes network for each of the <u>three</u> basic cases. Each network should have the three nodes a,b,c. Make sure you mark the direction of edges, and mark which nodes are observed.

(b) Given the appropriate Bayes network structure and the state of c (either observed or not observed), one can make nodes a and b independent (or conditionally independent). For each of the 3 cases above, prove the independence relationship, that a and b are independent or conditionally independent, given the appropriate state of c. For example, if your think that a and b are conditionally independent given c, you need to show that p(a,b|c)=p(a|c)p(b|c) can be derived from the particular Bayes network.

B760 – MACHINE LEARNING: BASIC QUESTIONS

B760-1 Reinforcement Learning

Consider a robot that performs *reinforcement learning* (RL) in some environment where only one "agent" learns. Actions performed by the robot are much more expensive than the cost of using a sizable number of CPU cycles to learn from the data gathered in the environment.

Assume you want to use a function approximator (call this supervised-learning algorithm FA for short) that only works in "batch" mode, which means it gets a set of training examples and produces a model. (Conversely, an incremental learner refines its current model after each training example.)

(a) Present in high-level pseudo-code the methodology you would use to perform RL with the batch learner. Be sure to explain (or comment) your code.

(b) Assume you are doing *Q* learning and describe how your system would create training examples for FA.

(c) Assume that algorithm FA has some tunable parameters. Present an experimental methodology using *tuning* sets for selecting good values for these parameters.

In other words, algorithm FA is given a set of labeled examples and needs to set its own parameters by using only these examples. Do not simply say "perform RL with the robot multiple times, each run using different parameter settings, then choose the best."

(d) What is one significant difference between tuning sets in RL and tuning sets in traditional supervised learning? Discuss any impact of this difference.

B760-2 Linear Classifiers

Consider a binary classification problem. Each training example is represented by a d-dimensional feature vector $\boldsymbol{x}=(x_1,\ldots,x_d)$ and has label y.

(a) Describe the Naïve Bayes classifier by:

i. Showing the Naïve Bayes assumption.

ii. Showing how the assumption is used to compute $p(y|\boldsymbol{x})$.

iii. Explaining why the Naïve Bayes classifier is a linear classifier in a certain feature space.

(b) Explain why an SVM with a linear kernel is a linear classifier too.

(c) Discuss one major difference between the two linear classifiers.

(d) List one major advantage of Naïve Bayes (relative to SVMs) and one major advantage of SVMs (relative to Naïve Bayes).

B766 – COMPUTER VISION: BASIC QUESTIONS

B766-1 Image Segmentation

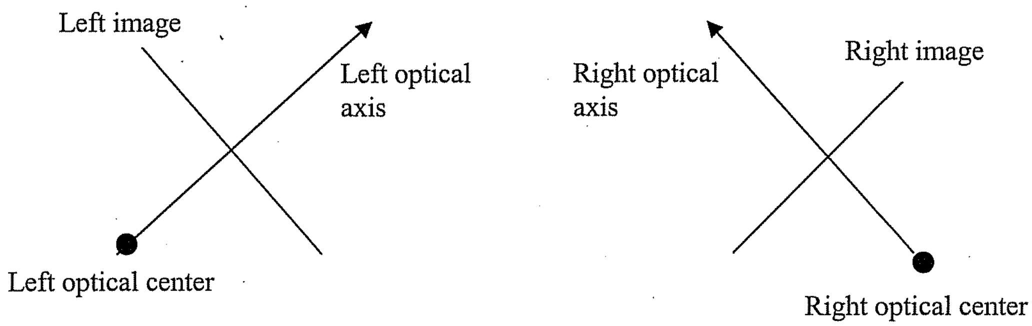
(a) Define the main steps of the Normalized-Cut segmentation method for segmenting an image into exactly two regions.

(b) Describe an affinity measure that would be appropriate to use with the Normalized-Cut algorithm, explain what it measures, and give one example of an image for which it would be well-suited, and one example for which it would not be well-suited.

(c) Describe how you could generalize your algorithm in (a) so that it could segment an image that contained more than two regions and the number of regions is unknown *a priori*.

B766-2 Stereo

(a) A stereo rig consists of a pair of cameras oriented so that their optical axes form an angle of 90 degrees as shown below; the y axes of the two cameras are parallel to each other and point out of the page. Draw the two 2D images and sketch on them the epipolar geometry, indicating clearly the positions of the epipoles and the epipolar lines. Also, define these terms in general.



(b) Given the camera configuration in (a), describe the position of a point P in the scene such that it projects to two conjugate points in the two images that have disparity 0.

(c) Define the *ordering constraint* (also known as the monotonicity constraint) that is used in stereo correspondence algorithms to disambiguate point matches on corresponding epipolar lines. Sketch a configuration of scene surfaces and cameras where the ordering constraint is valid, and a configuration where it is not valid.

B776 -- ADVANCED BIOINFORMATICS: BASIC QUESTIONS

B776-1 Clustering with High Confidence

A limitation of most clustering algorithms is that they will return a clustering whether there is any real "structure'' in the data or not. In other words, they will form a clustering even for a data set that consists purely of noise. Assume that we are interested in clustering genes where each gene is represented by a vector of expression (i.e. "activity") levels measured over a wide range of conditions. Further assume that every gene in the data set has been measured for all of the same conditions, that is, all instances are represented by the same features.

(a) Design and describe a clustering approach that has the following properties:

(i) the user does not have to specify the number of clusters to be returned,

(ii) not every instance has to be placed in a cluster,

(iii) the clusters returned are those that, with high confidence, represent real structure in the data.

Your approach may use a standard clustering algorithm as a subroutine.

(b) Discuss one significant limitation of your approach.

B776-2. Learning HMM Parameters with Partially Labeled Data

Consider a task in which we would like to distinguish two classes of sequences, and within each sequence class we would like to identify several specific types of subsequences. For example, suppose our sequences consist of the words in newspaper articles and our classes of interest are *international* stories and *domestic* stories. For international stories, we want to identify phrases that refer to *people* and *geographic locations*. For domestic stories, we want to identify phrases that refer to *people* and *organizations*.

Suppose that we are training an HMM for this task, and we have available as training data articles that have been labeled to various levels of detail and completeness.

(a) Sketch an HMM for this task.

(b) Describe how you would use standard HMM algorithms to do parameter learning from a subset of the articles for which both the class (*international* or *domestic*) and the phrases of interest (*people*, *locations*, *organizations*) are labeled.

(c) Describe how you would do parameter learning from a subset of the articles for which only the class (*international* or *domestic*) is labeled. Be sure to note how you would handle this case differently than (b).

(d) Describe how you would do parameter learning from a subset of the articles that belong to the *domestic* class and for which only the phrases about *organizations* are labeled. Be sure to note how you would handle this case differently than (c).

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

A766 – COMPUTER VISION: ADVANCED QUESTIONS

A766-1 Image Mosaics

(a) Define one circumstance (in terms of cameras poses or scene assumptions) in which two partially overlapping images can be aligned and combined to create a seamless mosaic image when scene depth information is not known.

(b) Describe a robust, feature-based approach to registering two images that includes detecting feature points in both images, determining a sufficient number of corresponding points, and then computing the transformation that aligns one image with respect to the other one.

(c) Using your answer to (b) as one component, describe the main steps associated with the problem of creating a mosaic image from two input images.

(d) Assuming there are multiple images from a video sequence to mosaic together and there may be moving objects in the foreground of a static scene, how could the moving foreground objects be automatically removed after the alignment between the images has been determined?

A766-2 Word-to-Picture

Consider the problem of finding a good, representative image for a given word from a set of candidate images that are found using an image search engine such as Google Image Search. That is, given a word, a set of *n* images are returned by an image search engine and the goal is to automatically select one of them as the "best" depiction of the word among these *n* candidates. Describe two important computer vision challenges that this problem contains, and then outline an approach to solving this problem using image analysis techniques. Clearly state any assumptions, additional knowledge, or bias that you employ.

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