

Spring 2012
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, February 6, 2012

GENERAL INSTRUCTIONS:

- (a) This exam has **15** numbered pages.
- (b) Answer each question in a separate book.
- (c) 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.*
- (d) 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 Bxxx, where these two questions need *not* come from the same section, *and*
- **both** questions in the section labeled A760 or A766, again corresponding to your chosen focus area.

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

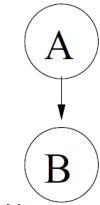
POLICY ON MISPRINTS AND AMBIGUITIES:

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.

731 Advanced Artificial Intelligence: Basic Questions

B731-1. Bayesian Networks

Consider the following Bayesian network (called N1) with two binary variables A and B:



Let its Conditional Probability Tables (CPTs) be

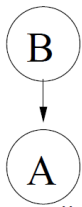
$$P(A) = a$$

$$P(B | A) = b$$

$$P(B | \sim A) = c$$

where a, b, c are real numbers between 0 and 1.

Now consider another Bayesian network (called N2) where the arrow is reversed:



Is there a set of CPTs that make N2 equivalent to N1? If yes, give the CPTs. If no, briefly explain why.

B731-2. Bayesian Network Structure Learning

Consider the problem of learning a Bayesian network structure for a task in which each variable is specified at the outset as having one of two types: *founder* or *descendent*. *Founder* variables are discrete and should not have any parents in the returned structure, whereas *descendent* variables are continuous and should have at least one parent. Design a search-based structure-learning algorithm for this task. In particular, you should discuss:

- a) how you represent conditional probability distributions in the network,
- b) how your search is initialized,
- c) the operators that are used in your search, and any conditions that a given operator application must satisfy,
- d) the scoring function and search method used by your approach.

760 Machine Learning: Basic Questions

B760-1. Overfitting Reduction

Explain how each of the following address **overfitting reduction** in machine learning. In your answers, be sure to briefly explain how one specific machine-learning algorithm would use this overfitting-reduction technique; do not use a specific learning algorithm more than once.

- a) Early stopping
- b) Minimal description length
- c) Ensembles
- d) Feature selection

B760-2. Support Vector Machines (SVMs) and other classification algorithms

- a) Show the constrained optimization problem commonly solved in SVMs. (Hint: show the objective function and the constraints for *either* the primal or dual version of your problem.)

- b) Discuss the most important way in which support vector machines:
 1. are similar to nearest-neighbor classifiers (kNN)

 2. differ from kNN

 3. are similar to artificial neural network learners based on backpropagation (ANNs)

 4. differ from ANNs

766 Computer Vision: Basic Questions

B766-1. Feature Detection, Description and Matching

- a) Describe the main steps for **detecting** local corner or interest feature points in an image based on either the SIFT detector, the Harris detector, or the Tomasi-Kanade detector.
- b) Why are corner features preferred over simply using edge point features as part of a system for matching corresponding feature points between two images?
- c) How is the characteristic scale of a feature point determined by SIFT?
- d) What properties of the SIFT descriptor make it invariant to changes in intensity?
- e) Given two images, one with n feature points and associated SIFT descriptors, and the other image with m feature points and associated SIFT descriptors, describe how a set of corresponding point **matches** between the two images can be computed.

B766-2. Stereopsis and Epipolar Geometry

- a) Formally define epipolar line, epipole, and the epipolar constraints in terms of the fundamental matrix F .
- b) If we assume two images are captured by the same camera with identical internal parameters (focal length, principal points, and no distortion), and we further assume the camera movement is only translation (with no rotation) and the translation is parallel to the x -axis in the image plane, what is the form of the fundamental matrix F ?
- c) Formally define the procedure of rectification. Give one reason why we may want to perform rectification in stereopsis.
- d) Consider stereo matching using window-based correlation versus using Dynamic Programming. Give one advantage of window-based correlation over Dynamic Programming and also give one advantage of Dynamic Programming over window-based correlation

769 Advanced Natural Language Processing: Basic Questions

B769-1. Perceptron training of Structured Linguistic Models

In structured models, predictions consist of many parts. For example, a sentence's label might be a sequence of part-of-speech tags or even a parse tree. Suppose we are given a feature function and a set of training examples consisting of input-label pairs. Our goal is to train a linear model, consisting of weights for each feature. One simple yet effective algorithm for doing so is the structured perceptron. In this algorithm, we iterate over each training instance, and update our model based on its current prediction.

- a) Write down the prediction rule for the linear structured model. This formula should involve the feature function Φ , an input X , weights W , and the set of possible labels Y .
- b) Write down the update rule for the structured perceptron algorithm, specifying when and how the weights are updated in response to a training prediction.
- c) For tasks such as part-of-speech tagging and parsing, there are an exponential number of possible labels in the length of each sentence. In two or three sentences explain how we can efficiently implement our prediction rule in these two cases.

B769-2. Part-of-Speech Tagging with HMM's and CRF's

To predict the parts-of-speech of words in sentences, we employ sequence models. This question will concern two such models: the Hidden Markov Model (HMM) and the Conditional Random Field (CRF).

- a) Consider a training sentence: *Sam clapped after the lecture*, with the corresponding tag sequence NNP VBD PREP DET NN (where NNP indicates a proper noun, and VBD indicates a past-tense verb)
 1. Draw a HMM model structure for this sentence, including labels for the correct parts-of-speech.
 2. Draw a CRF model structure for this sentence.
- b) Consider the test sentence: *Ziad grokked the lecture*. Assume that the name 'Ziad' and the verb 'grokked' (meaning 'understood at a deep intuitive level') have never appeared in the training data.
 1. Briefly describe how an HMM model would deal with unseen words like these.
 2. Briefly describe how a CRF can deal with these words more easily and effectively.

776 Advanced Bioinformatics: Basic Questions

B776-1. Learning RNA Structural Motifs

A molecular biologist approaches you with a set of RNA sequences believed to contain a common *structural* motif, i.e., each sequence is largely unstructured (unfolded) except for a relatively short subsequence that folds into a common structure. Whereas the motif segments of each RNA share a common structure, they are *not* believed to share common sequence. Your task is to learn the structure of the motif and to infer its position in each RNA using a stochastic context-free grammar (SCFG).

- a) Specify a SCFG (i.e., a set of productions) for this task.

- b) Which parameters of your model will represent the structural motif?

- c) Name two different algorithms that you could use to estimate the parameters of your model. Briefly describe the advantages and disadvantages of the two algorithms.

B776-2. Profile HMM surgery

You are given a profile HMM (both its structure and parameter values) for a particular protein family and a set of sequences believed to belong to the family.

- a) Which algorithm would you use to obtain an alignment of the sequences using the profile HMM?
- b) You examine the alignment you obtained from the algorithm in a). What *features* of this alignment might suggest that the structure of the profile HMM is not entirely correct (i.e., that the number of match states is too small or too large)?
- c) Using your intuition from b), describe an iterative algorithm for updating the structure of the profile HMM. Give both
 1. an update step, and
 2. a stopping condition.

760 Machine Learning: Advanced Questions

A760-1. Learning Markov Logic Networks

Suppose you wish to model social network data from a source such as Facebook. For each person, you have their friends and keywords representing their interests (for this question, let's ignore text postings and other information). You wish to learn rules and parameters for an MLN so that you can predict additional (as yet un-annotated) friends and interests of individuals.

- a) What predicates and constants will you use in your representation?
- b) Discuss how you will evaluate how well MLNs perform at this task. Include both metrics and methodology in your discussion.
- c) What will be the most difficult challenge in addressing this problem with MLNs, and why?
- d) Discuss how will you address this challenge?

A760-2. Unsupervised Learning

Consider unsupervised learning in the absence of ground truth. In scenarios such as this, it is often difficult to interpret the meaning or quality of the results of an unsupervised learning algorithm. However, suppose that you are asked to cluster data and someone is able to provide a few labeled instances for each expected cluster.

- a) How would the presence of these labeled instances change the way you cluster the data? Briefly explain the approach you would use and assumptions you would be making.
- b) Would you now be able to use cross-validation to gauge the accuracy of your results? If you could not, explain why; if you could, explain how you would do so.
- c) Suppose the person who supplied the labels tells you that some of the given labels might be wrong, without telling you which ones. How would that affect your answers to parts a) and b) above? Explicitly state any assumptions you are making.

766 Computer Vision: Advanced Questions

A766-1. Structure From Motion

- a) Formally define the problem of Affine Structure from Motion. Describe the Tomasi-Kanade algorithm. What are the minimum number of images and the minimum number of feature points required by this algorithm?
- b) The Tomasi-Kanade algorithm has a few limitations. Name one of them and discuss how you would address it in practice. The limitation you describe should be different from the ones listed in c) and d) below.
- c) The Tomasi-Kanade method does not handle the projective camera model. Can we still use the rank constraint in the Tomasi-Kanade algorithm for Projective Structure from Motion? If yes, how? If no, why not?
- d) The Tomasi-Kanade method assumes the underlying 3D structure is rigid. For an object that is deforming during image acquisition, suggest an approach for how we may infer its 3D structure from its 2D motion. You can pose some constraints on the deformation; for example the deforming shapes lie within a linear subspace.

A766-2. Image Alignment

- a) Consider two monochrome images I and J which we seek to align by calculating a global displacement vector $\mathbf{u} = (u, v)$. Briefly describe how you will make use of Fourier analysis based alignment for identifying such translational shifts. Then, identify at least one advantage of this method.

- b) If one is not specifically interested in pixel to pixel alignment, feature-points extracted from a pair of images can be used. But comparing all feature-points in one image, one at a time, against all feature-points in the other to find its best (or near best) match is quadratic. This may be impractical for some applications. Identify any one idea or approach that is used to speed up this feature comparison process.

- c) Global alignment of a large number of images for simultaneous pose estimation is called "Bundle Adjustment". Bundle Adjustment is typically formulated as an optimization problem. Assume that you are provided an initial pose estimate. Describe in English what type of error the objective function minimizes, and what optimization methods are generally used.