

Fall 2014  
COMPUTER SCIENCES DEPARTMENT  
UNIVERSITY OF WISCONSIN–MADISON  
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 15, 2014

## GENERAL INSTRUCTIONS

1. This exam has 8 numbered pages.
2. Answer each question in a separate book.
3. 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.
4. Return all answer books in the folder provided. Additional answer books are available if needed.

## SPECIFIC INSTRUCTIONS

You should answer:

1. both questions in the section labeled 760 – MACHINE LEARNING
2. two additional questions in another selected section, 7xx, where both questions *must* come from the same section.

Hence, you are to answer a total of four 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.

## 760 – MACHINE LEARNING: REQUIRED QUESTIONS

### 760-1 Parallelized decision tree learning

Assume you have a task with  $n$  features and  $m$  examples, plus you have  $p$  processors. You wish to learn  $k$  decision trees.

1. Assume  $k = 1$  in this part of the question. Consider choosing the root node.

You decide to give each of  $p-1$  processors an equal and disjoint fraction of the examples. The remaining processor (the ‘master’) sends out the candidate features and asks the  $p-1$  ‘helper’ processors to score them.

*Concretely* describe what information each helper processor needs to send back to the master.

2. Keep  $k = 1$  but now consider the recursive calls after the root has been chosen by the master processor. What additional steps are needed beyond what you described in your answer to Part 1? Be concrete.
3. Now consider arbitrary values for  $k$ . Present **two** ways of learning  $k$  decision trees in parallel using *bagging*. Describe the merits of each approach depending on the relative values of  $n$ ,  $m$ ,  $p$ , and  $k$ .

## 760-2 SVMs, ensembles, and active learning

Suppose you want to build an ensemble of support vector machine classifiers.

1. Describe two different ways you can create diversity in your collection of trained SVM classifiers.
2. The parameter commonly labeled  $C$  is used to trade off the two goals of large margin and consistency with training example labels. Show how  $C$  does this by presenting either the primal **or** the dual formulations of the SVM optimization problem with  $C$  as the coefficient on the slack variables.
3. How does the value of the  $C$  parameter affect diversity within the ensemble? How might the optimal value for  $C$  in the ensemble setting differ from the non-ensemble setting, and why?
4. Having constructed an ensemble of SVM classifiers, suppose you are given the opportunity to improve it further by active learning. Describe one active learning approach that will use your ensemble to select the example whose label you will request next.

## 769 – ADVANCED NATURAL LANGUAGE PROCESSING QUESTIONS

### 769-1 Part-of-speech tagging noisy input

You are employed by NLP Corp. and you have been assigned the task of POS tagging the output of a speech recognizer. You have already trained a first-order (bigram) HMM and you must figure out how to obtain the most likely label sequence  $y$  for a given sentence  $x$ .

1. Describe a dynamic program that will allow you to obtain  $y$  without enumerating all possible sequences. Pseudo-code is not required, but you should use precise notation to describe your data structures, the base cases, and the recurrence equations. What is the time complexity of your algorithm?
2. Now assume that the speech team gives you instead a *set* of 10,000 unscored sentence interpretations for each input voice utterance. Your job is to simultaneously select one of the sentences along with its label sequence which together have the highest probability under your HMM model. You notice that most of the 10,000 sentences tend to differ by only a word or two. Describe how you could exploit the similarity of the sentences to avoid running 10,000 separate dynamic programs.

## 769-2 Partial parsing

In your next assignment at NLP Corp. you are given the task of syntactically parsing sentences. You have already trained a simple PCFG model for English.

1. Describe a dynamic program that will allow you to obtain the most likely parse  $y$  for sentence  $x$ . Pseudo-code is not required, but you should use precise notation to describe your data structures, the base cases, and the recurrence equations. What is the time complexity of your algorithm?
2. For some sentences, the CEO of NLP Corp. will supply you with a span  $(i, j)$  which he demands form a constituent in the parse tree you predict. Describe a modification to the dynamic program of Part 1 that will ensure that you produce the most likely parse  $y$  subject to the constraint that words  $i$  through  $j$  in sentence  $x$  are dominated by a node in the tree.
3. The CEO of NLP Corp. is becoming increasingly erratic. He has now instituted a policy whereby your pay will depend *only* on the correctness of the parse of the span  $(i, j)$  within the sentence. If you simply run the constrained dynamic program of Part 2 would your parse be optimal for maximizing your pay? If not, describe a further modification to produce optimal partial parses of word span  $(i, j)$ . (Hint: you need not even provide parses for the rest of the sentence, but your predictions should still take the rest of the sentence into account.)

## 776 – ADVANCED BIOINFORMATICS QUESTIONS

### 776-1 Constructing a phylogenetic tree using gene networks

Recall the problem of constructing phylogenetic trees using DNA sequence data. We will apply concepts of constructing a phylogenetic tree from sequence data to construct a phylogenetic tree of gene networks. A phylogenetic tree of gene networks is defined as a phylogenetic tree with a gene network at each leaf instead of a DNA sequence.

1. There are three approaches to constructing phylogenetic trees: (a) the distance-based approach, (b) the parsimony-based approach, and (c) the probabilistic approach. Briefly describe each approach and give one advantage of each.
2. Assume you are given as input 10 gene networks, each with at most 20 genes. Devise an algorithm to construct a phylogenetic tree of these gene networks based on one of the approaches stated in Part 1. Describe any assumptions you are making about the gene networks. Justify your choice of approach over the other two approaches.
3. How will you assess the confidence in your constructed phylogenetic tree of gene networks?

## 776-2 Open reading frame motif finding

Suppose you are given 100 DNA sequences, each of length 1000 bases. You are told that within each sequence, there is a single weakly-conserved open reading frame (ORF) of unknown but fixed length  $L$  (coding for a peptide of  $\frac{L}{3} - 1$  amino acids in length). The starting position and reading frame (including strand) of the ORF in each sequence are unknown, and may vary between the sequences. Note that there may be many ORFs in each sequence, but only one is conserved across all of the sequences. Your task is to design an efficient method for discovering the weakly-conserved ORF in each sequence. There is no detectable conservation between these sequences at the DNA level and thus classical DNA motif finding methods fail to address this task. In your approaches to this task, you may *not* use any sequence data outside of the 100 DNA sequences provided as input.

1. Describe a probabilistic model-based approach for this task.
2. Describe a BLAST-based approach for this task.
3. Briefly describe the strengths and weaknesses of the two approaches as compared to each other.

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