

University of Wisconsin-Madison
Computer Sciences Department

CS 760 — Machine Learning

Spring 1993

Midterm Exam

(calculator and one page of notes allowed)

100 points, 90 minutes

April 19, 1993

Write your answers on these pages and show your work. If you feel that a question is not fully specified, state any assumptions you need to make in order to solve the problem. You may use the backs of these sheets for scratch work. Notice that all questions do not have the same point-value. Divide your time appropriately.

Before starting, write your name on this and all other pages of this exam. Also, make sure your exam contains four (4) problems on 9 pages.

Problem	Score	Max Score
1	_____	25
2	_____	15
3	_____	30
4	_____	30
Total	_____	100

1. Learning Decision Trees (25 points)

- i) Assume you are given the following three nominal features with the possible values shown.

Owner \in {betty, jim, pat}
Size \in {small, med, large}
Wheels \in {one, two, four, eight}

Using Quinlan's ID3 and its max-gain formula, produce a decision tree that accounts for the following classified examples. **Show all your work.** You may use the abbreviations used to specify the examples.

O = b	S = s	W = e	+
O = p	S = s	W = o	-
O = b	S = l	W = e	-
O = p	S = m	W = t	+

- ii) Imagine you have a collection of training examples for some Boolean-valued function that you wish to learn (you may assume the features of the examples are also Boolean-valued). Also imagine that you are given a *magic ML program* that takes as its only input the list of features used to represent the examples.

Somehow, without requiring any training examples, every 0.1 seconds this magic program outputs - without duplication - a decision tree (one that only uses the features provided in the call to the magic program). Briefly discuss how you would use this magic program to solve your learning task. Discuss the major strength and the major weakness in your approach.

system design:

major strength:

major weakness:

2. Learning as a Search Problem (15 points)

- i) Describe how Michalski's AQ formulates learning as a traditional AI search task.

states:

operators:

heuristics:

end test:

- ii) What do you see as the most fundamental difference between the *spaces* searched by Mitchell's version-space algorithm and Rumelhart et al.'s backpropagation technique? Justify your answer.

3. Artificial Neural Networks (30 points)

- i) Consider using an artificial neural network (without hidden units) to learn, using the *delta rule*, the following examples:

Input				Output
0	1	0	1	1
1	1	0	1	0
1	0	0	1	1

Draw the network *after* processing (once) each of the above three (3) training examples. Use a linear threshold unit as the output and initialize its threshold to 0 (assume the node is "active" if it equals or exceeds its threshold). Also assume all weights are initially 0 and that $\eta=0.1$.

What output does your final network give for the input "0 0 1 1"? _____

ii) Briefly describe two (2) ways of addressing the *overfitting problem* in backpropagation training.

(a)

(b)

iii) Somehow you have chosen the "right" number of hidden units for the learning task you have at hand. Assume that you still have time to train, using your collection of examples, five networks with this topology. One approach would be to train the five networks and somehow estimate which one is best, discarding the other four. However this seems wasteful; can you imagine any way of using all five trained networks on the testset? Sketch a design and discuss its major strength.

iv) Assume someone uses the backpropagation algorithm with the error function:

$$error = \frac{1}{2} \sum (teacher - 0.5)^2 . \quad \text{What would happen and why?}$$

4. Explanation-Based Learning (30 points)

- i) Consider the following EBL domain theory. Terms beginning with ?'s are implicitly universally-quantified variables.

$$\begin{array}{lll}
 A(?x, ?y) \text{ and } B(?y, ?x, ?z) & \rightarrow & C(?x, ?y, ?z) \\
 D(?x, ?y) \text{ and } E(?x, ?y) & \rightarrow & B(?x, 1, ?y) \\
 F(?x, ?x) \text{ and } F(?y, ?y) & \rightarrow & D(?x, ?y)
 \end{array}$$

Assume the following problem-specific facts are asserted:

$$\begin{array}{llllll}
 A(1, 2) & A(3, 1) & E(2, 3) & F(3, 2) & F(2, 2) & G(3, 3) \\
 A(2, 2) & A(3, 2) & E(1, 2) & F(3, 3) & F(2, 3) & G(3, 2)
 \end{array}$$

Explain, with a proof tree, that $C(1, 2, 3)$ is true. Draw to the *right* of your proof tree the corresponding *explanation structure* (before pruning at operational nodes). Clearly indicate the necessary unifications.

Assuming that predicates A, E, F, and G are operational, what rule would Mooney's EGGS algorithm learn? Explain your answer.

Assuming that predicate B *also* is operational, what rule would Mooney's EGGS algorithm learn?

- ii) Considering using EBL to speed up a planner. Briefly describe the approach of learning *macrooperators* and the approach of learning *search control*. For each approach, discuss one of its advantages over the other.

macrooperators:

search control:

Name one system that learns macrooperators: _____

Name one system that learns search control knowledge: _____

iii) What would you claim is the closest analog in empirical learning (i.e., SBL) to the EBL *utility problem*? Justify your answer.