

**University of Wisconsin-Madison
Computer Sciences Department**

CS 760 — Machine Learning

Spring 1992

Midterm Exam

(one page of notes allowed)

100 points, 90 minutes

April 28, 1992

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	_____	35
2	_____	35
3	_____	20
4	_____	10
Total	_____	100

1. Empirical Learning (35 points)

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

Shape \in {triangle, diamond}
Font \in {small, large}
Color \in {red, blue, green}

Also assume that ID3 is given the following set of classified examples. Using Quinlan's max-gain formula, produce a decision tree that accounts for these examples. **Show all your work.** You may use the abbreviations used to specify the examples.

S = t	F = l	C = g	+
S = d	F = s	C = r	+
S = t	F = s	C = g	+
S = t	F = l	C = r	-
S = d	F = s	C = g	-

ii) Briefly discuss the relations, if any, between the G and S sets in Mitchell's Version Spaces algorithm and Michalski's concepts of *discriminant* and *characteristic* descriptions.

iii) Define, in terms of concept space, the following types of *inductive biases*.

(a) preference biases

(b) restricted hypothesis-space biases

iv) Which type of biases (preference, restricted hypothesis space, both, or neither) do ID3, Version Space, and AQ have? Briefly justify your answers.

(a) ID3

(b) Version Space

(c) AQ

v) Briefly define the experimental question addressed by *t-tests*. (Note: you need not describe the *calculations* of a *t-test*.)

2. Artificial Neural Networks (35 points)

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

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

Draw the network after processing (once) each of the above four (4) 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 $\eta=0.25$.

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

ii) Discuss two (2) useful ways of using *tuning sets* in neural network training.

(a)

(b)

iii) Briefly describe one (1) major problem that can arise from using a neural network with:

(a) too *few* hidden units

(b) too *many* hidden units

iv) Assume you are given a linearly-separable data set and have trained a *perceptron* so that it properly classifies all the examples in your training set. Consider the two network alterations listed below. For each, either (1) argue that it would have no impact on the perceptron's accuracy on the training set or (2) show (mathematically or with a specific case) that the accuracy can change.

(a) you multiply all the weights and the output unit's threshold by α , a positive constant

(b) you add α to all the weights and to the output unit's threshold

v) Repeat part (iv-a), but assume now that your data is *not* linearly separable and that you have to use a neural network with one layer of sigmoidal hidden units to completely separate (i.e., learn) the training data.

(a) you multiply all the weights and each unit's bias by α

3. Explanation-Based Learning (20 points)

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

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

Assume the following problem-specific facts are asserted:

$$E(1,2) \quad E(3,2) \quad F(2,3,3) \quad F(3,1,2) \quad G(1,2) \quad D(3,3)$$

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 predicate B is operational, what rule would Mooney's EGGs algorithm learn?

- ii) Define the *utility problem* in EBL and briefly explain Minton's solution to it. Be sure to discuss why this is a central problem to EBL.

4. Learning without a Teacher (10 points)

Fisher's COBWEB system is designed to work incrementally - it assumes examples arrive periodically and it adjusts its current hierarchy following the receipt of each new example. Sketch the design of a *batch* version of COBWEB that also does hill-climbing search. That is, assume all the training examples are available at the start of learning and describe how your approach would heuristically search for a good concept hierarchy. (Do not merely convert this problem to that addressed by the standard, incremental COBWEB algorithm.)