CS 527A Machine Learning **Spring Semester, 2001**

Midterm Exam

March 7, 2001

- For any problem that involves any arithmetic you need not give the final answer. Just be sure to put it in a form where all that would be left to do is plug it into a calculator.
- If you have any questions about what is being asked in a problem, please come ask me to clarify the problem.

1. (10 pts) Answer true or false for each question and then give a 1 or 2 sentence explanation to back your answer.

   (a) Assuming you are not concerned with the training time, when using an artificial neural network it is best to include enough hidden units so the training error can be reduced as much as possible.

   (b) It is best to pick a learning algorithm (and hypothesis space) that minimizes bias.

   (c) Decision tree post-pruning techniques are good because they help reduce the training error.

   (d) Reinforcement learning is a special case of supervised learning where the learner gets extra reinforcement.

   (e) Backpropagation uses gradient descent to descend the error surface corresponding to the true error of the network?

2. (10 points). Consider the hypothesis class $\mathcal{H}$ of decision trees over boolean variables x_1 and x_2.

   (a) Let the training data $D = \{\langle 01, + \rangle, \langle 00, - \rangle\}$. Show the version space $VS_{\mathcal{H},D}$ using the general-to-specific partial ordering. If there are two syntactically different decision trees that are equivalent semantically, you can show whichever you would like in the version space

   (b) Suppose $D' = D \cup \{\langle 10, - \rangle\}$. What is the most specific set S for the version space $VS_{\mathcal{H},D'}$?

   (c) What is the decision tree that would be output by ID3 using D' as the training data (with no pruning)?

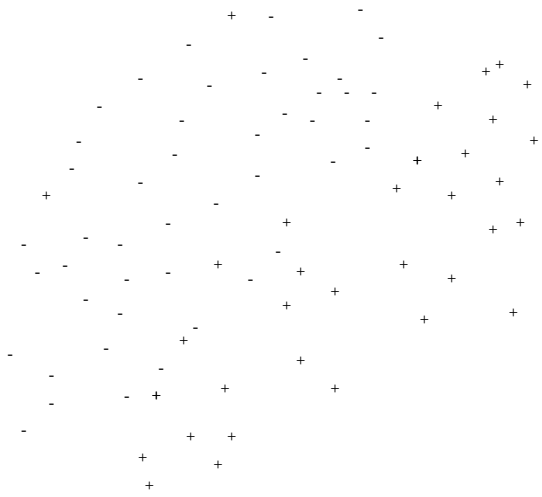
   (d) Explain your answer to part (b) and part (c) in terms of the bias of the two learning algorithms.

3. (8 points) For each of the learning algorithms shown below, give an advantage and one disadvantage that is NOT shared by the other three.

Learning Algorithm	Advantage	Disadvantage
ID3		
Backpropagation		
FindS		
Weighted Majority		

4. (4 points) Suppose you are given that hypothesis h correctly predicts 75 out of 100 examples randomly drawn from distribution $\mathcal{D}$. Give the smallest b you can such that you can state there is a 90% chance that $error_{\mathcal{D}}(h) \leq b$.

5. (3 points) In the diagram below show the hypothesis that would be output by the delta (LMS) rule on the given data if it were run long enough to converge. I don't expect you to exactly compute the right answer. To be sure we knew what you were aiming to do, briefly explain what criteria you used to select the hypothesis you marked.



6. (10 points) Recall we showed in class on Monday that VCD(monotone monomials) = n.

   (a) Give the best a mistake-bound algorithm for monotone monomials that you can. You are NOT required to make a prediction in polynomial time. Be sure to clearly describe your learning algorithm and derive a bound on the number of mistakes you make in the worst case.

   (b) Now try to prove your algorithm achieves the optimal mistake bound. If you cannot, then give the best lower bound you can on the number of mistakes made by any mistake-bound algorithm for learning monotone monomials.

7. (15 points) Let X be the real numbers. Let concept class $\mathcal{C}_I$ be the class composed of intervals. That is, each $c \in \mathcal{C}_I$ is defined by reals a and b where $a \leq b$ where for any $x \in X$, $c(x) = 1$ if and only if $a \leq x \leq b$.

   (a) What is $|\mathcal{C}_I|$?

   (b) What is VCD($\mathcal{C}_I$)? Be sure to argue your answer is correct.

   (c) Let $\mathcal{C}_{UI}$ be the union of any number of concepts from C_I. That is it is the union of any number of intervals. What is VCD($\mathcal{C}_{UI}$)? Again, be sure to argue your answer is correct.

NOTE: If you could not solve any of the above, then to help with (d) and (e) you can come to the front and ask me to give you the answers. There will be a deduction of 1 pt for part (a) and 2 points each for parts (b) and (c) for me to write the answers. You can then receive partial credit to then explain why they are correct. For half this deduction, I will check your answers.

(d) Is $\mathcal{C}_I$ efficiently (i.e. polynomial time and sample complexity) PAC-learnable? If not, clearly explain why. If so, give a PAC-algorithm for it including the number of examples that will be needed.

(e) Is $\mathcal{C}_{UI}$ PAC-learnable? If not, clearly explain why. If so, give a PAC-algorithm for it including the number of examples that will be needed.

Problem	Points Possible	Points Received
1	10	
2	10	
3	8	
4	4	
5	3	
6	10	
7	15	
total	60	