Your Name (very very clear, capitals):______________________________

Your Andrew userid (clear, with letters in upper case):_______________

15-381 Final Exam Fall 1997

The following exam contains 14 questions. The first nine are easier. The total number of points available is 115 (any points you get above 100 are considered extra credit).

Good luck!

Remember that if you get stuck on an early question, you may find some of the other questions easy, so keep working and come back to the difficulty-causing problem later.

The exam is OPEN BOOK. You may use Russell and Norvig and/or your notes. If you don't have a copy of the text book or are missing a classroom handout, and feel that you need it to answer a question, you can borrow one for a short period of time from the examiner

1. Search (12 points)

For each of the problems listed below, circle a search method that is likely to be the most appropriate. Assume that memory limitations are not a problem.

a. Solving a one-player puzzle, given a good evaluation function with no local minimum.

Hill Climbing Best First A* Minimax

b. Solving a one-player puzzle, given a good evaluation function but with local minima.

Hill Climbing Best First A* Minimax

c. Backpropagation can be viewed as a search through the weight space of a neural network, trying to find the lowest error. In this light, the search method it uses is

Hill Climbing Best First A* Minimax

d. Finding a good move at each turn in a game of checkers (the game that is also known as 'draughts').

Hill Climbing Best First A* Minimax

e. Finding the shortest path through a maze. (Assume that you have no information other than the moves available from your current location.)

Hill Climbing Best First A* Minimax

f. Finding the shortest path through a maze, if you have a gadget that tells you the distance from where you are to the exit. (It won't tell you the direction to the exit, and it only tells you the distance you would have to travel if there were no walls in the way).

Hill Climbing Best First A* Minimax

2. Bayesian Networks (10 points)

People sometimes leave Wean Hall. Sometimes people tamper with fire alarms. Sometimes there's a fire in Wean Hall. Sometimes there's smoke. Sometimes the alarm sounds.

a. If the alarm goes off, 88% of the time people will leave the building. However, if the alarm doesn't go off there is still a small (0.1%) chance that people will leave. Express these facts by drawing a network and filling in probability tables at each node. Use just the nodes "Leaving" (L) and "Alarm" (A). You do not need to fill in the probability table for P(A) because we haven't told you what it is.

b. If there's no fire, there is only a 1 in 100 chance of there being smoke. And, if there is a fire, there's a 90% chance of smoke. Overall, there's a 1 in 100 chance of fire. Express these facts by drawing a network and filling in probability tables at each node. Use just the nodes Smoke (S) and Fire (F).

c. The manufacturer claims that, provided no-one tampers, the alarm will work every time that there is a fire. However, if someone tampers with the alarm, there is a 50-50 chance that the alarm will sound whether or not there's a fire. According to the manufacturer, there is zero chance of the alarm ringing for no reason at all. There's a 1-in-10 chance that someone would tamper with the alarm, and a 1/100 chance of fire. Express these facts by drawing a network and filling in probability tables at each node. Use just the nodes Alarm (A), Fire (F) and Tampering (T).

d. Draw the composite network created by joining all the above facts. You may assume all the commonsense conditional independence relations and so you do not need to add any more "links" to any of the existing networks, just overlay the nodes.

e. What is the probability of smoke and fire?

f. What is the probability of smoke and no fire?

g. What is the probability of smoke?

h. You observe smoke. What's the probability that there's a fire?

i. What is the probability that there's smoke, there's an alarm, and there's tampering, and there's a fire?

3. Decision Trees (10 points)

You are an allergist. One of your patients complains of a food allergy, but cannot determine its cause. You ask the patient to keep a journal of all the foods they eat for a week, and whether or not they had an allergic reaction on each day. Here is the journal:

Apple	**Beef**	**Cake**	**Reaction?**
No	Yes	No	Yes
No	Yes	Yes	Yes
No	Yes	No	Yes
Yes	Yes	Yes	Yes
Yes	Yes	No	No
No	No	Yes	No
Yes	No	No	No
No	No	Yes	No

a. Which of the three food attributes has the highest information gain?

b. Construct the decision tree for this data using the standard decision tree learning algorithm for building decision trees greedily with Information Gain (assume no pruning).

c. What would you predict for these next two days:

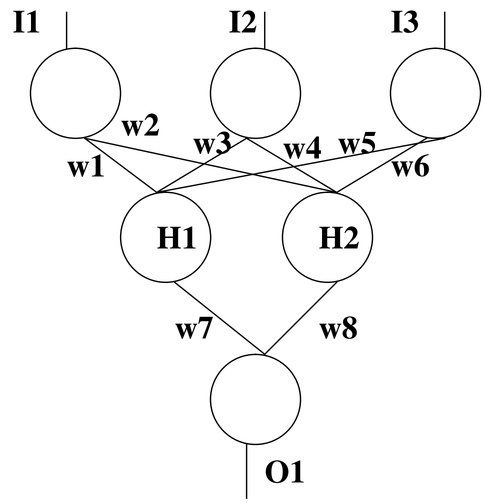
Apple	**Beef**	**Cake**	**Reaction?**
Yes	No	Yes	
No	No	No	

d. Suppose that your patient misunderstood your directions and only kept a record of what she ate on days when she had an allergic reaction. Could a useful decision tree be constructed? Explain in not more than three sentences why or why not.

4. Neural Nets (10 points)

a. At a meeting of the Machine Learning Contentious Club, someone stands up and proclaims that people should universally use neural nets and backpropagation for all learning tasks. Argue against this claim by giving two examples of other learning algorithms. For each algorithm give one reason why it might perform better than a neural net trained with backpropagation.

b. Consider the following neural net where each node is a sigmoid threshold function. Could the predictions made by this network change if we multiply each of the weights and thresholds in the network by the positive constant K? (Yes or No)



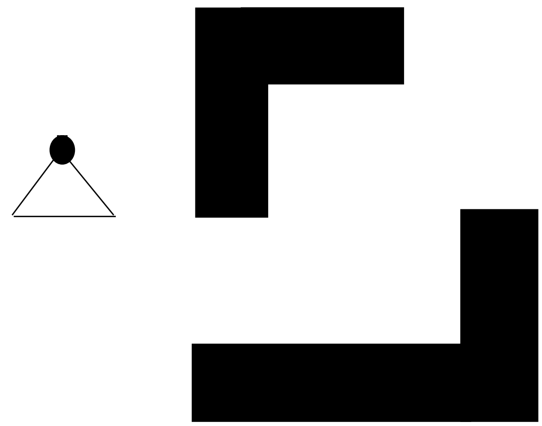
c. Every day at 9am, you must make a prediction of whether the stock market will go up or down using a neural net. You train your neural net from daily data from the last three years including data up to last night's. Inputs are yesterday's closing prices of various stocks, bonds and other financial indicators. Four data mining companies each try to sell you their neural network package. When you try them out, you get the following results:

Network Company	Time to make 1 prediction (secs)	Mean Test Set Cross Validation Error (% wrong predictions)	Time to Train (hours)	Training Set Error (% wrong predictions)
NeuroBrain	15	17	6	1.3
Nets'r'Us	17	12	5	3.2
Neural Hut	10	28	1	1.2
Neuronics	8	26	3	25

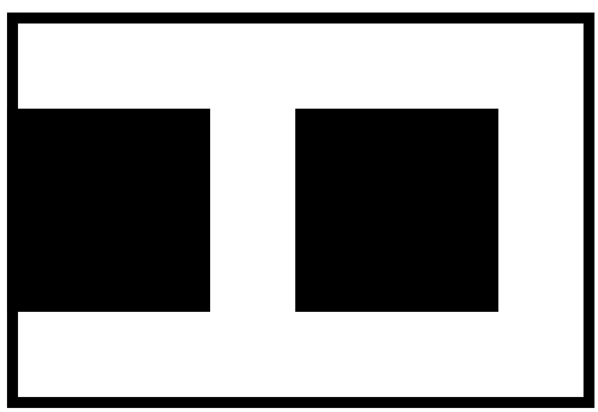
Based on this, which neural net would you choose for the job? Justify your answer.

5. Robot Motion Planning (10 points)

a. Dra w the configuration space of the following scene for the triangular robot using the indicated distinguished point:



b. Dra w a rough sketc h of the V oronoi skeletonization of the following space:



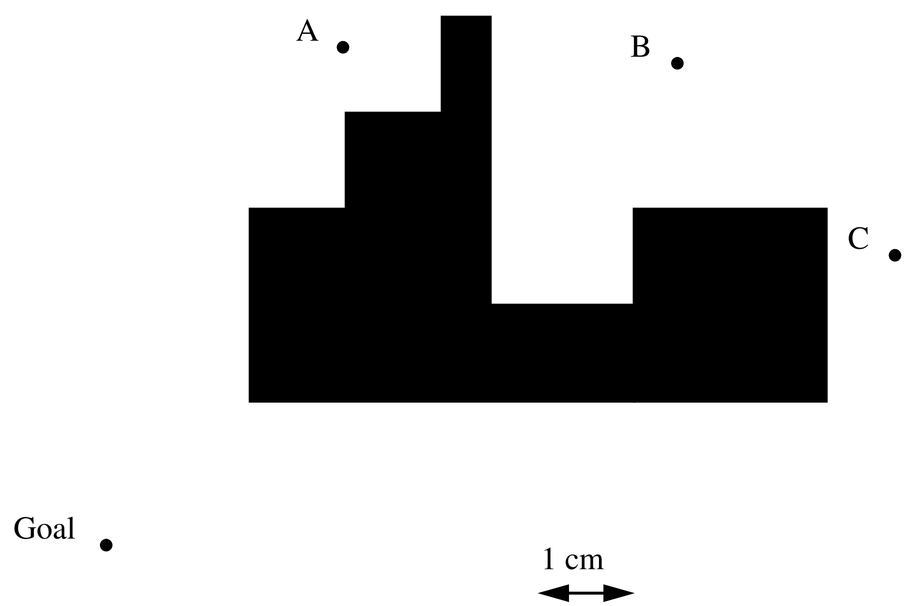
c. In the following maze , let

- d_0 = distance from robot to closest point on the obstacle in centimeters .
- d_g = distance from the robot to the goal in centimeters .

Suppose the robot uses the potential field method for path planning , with the field value defined as

$$d_g + 1/d_0$$

Draw (roughly) the path that the robot would take starting from point A on the following diagram.



d. Draw (very roughly) the path that it would take starting from point B on the above diagram.

e. Draw (very roughly) the path that it would take starting from point C on the above diagram.

6. Logic (6 points)

For each of the following sentences, circle the best translation into first-order predicate calculus:

a. Not all students take both History and Biology:

$\forall x. \sim (TakeHistory(x) \wedge TakeBiology(x))$

$\exists x. \sim (TakeHistory(x) \wedge TakeBiology(x))$

$\sim (\exists x. TakeHistory(x) \vee TakeBiology(x))$

$\exists x. TakeHistory(x) \vee TakeBiology(x)$

$\forall x. TakeHistory(x) \vee TakeBiology(x)$

b. There is a barber who shaves all and only those people who do not shave themselves:

$\exists x. \forall y. Shaves(y,y) \Rightarrow (Barber(x) \wedge \sim Shaves(x,y))$

$\exists x. Barber(x) \wedge \forall y. (Shaves(x,y) \Rightarrow \sim Shaves(y,y))$

$\forall y. \exists x. (Barber(x) \wedge Shaves(x,y) \Leftrightarrow \sim Shaves(y,y))$

$\exists x. \forall y. Barber(x) \wedge (Shaves(x,y) \Leftrightarrow \sim Shaves(y,y))$

$\forall y. \sim Shaves(y,y) \Leftrightarrow (\exists x. Barber(x) \wedge Shaves(x,y))$

c. FOR THIS QUESTION, choose all first order predicate calculus sentences that accurately represent the English sentence. There may be more than one.

"Only one student failed the exam"

$\exists x. FailedExam(x)$

$\forall x. FailedExam(x) \Rightarrow (\forall y. FailedExam(y) \Rightarrow (x=y))$

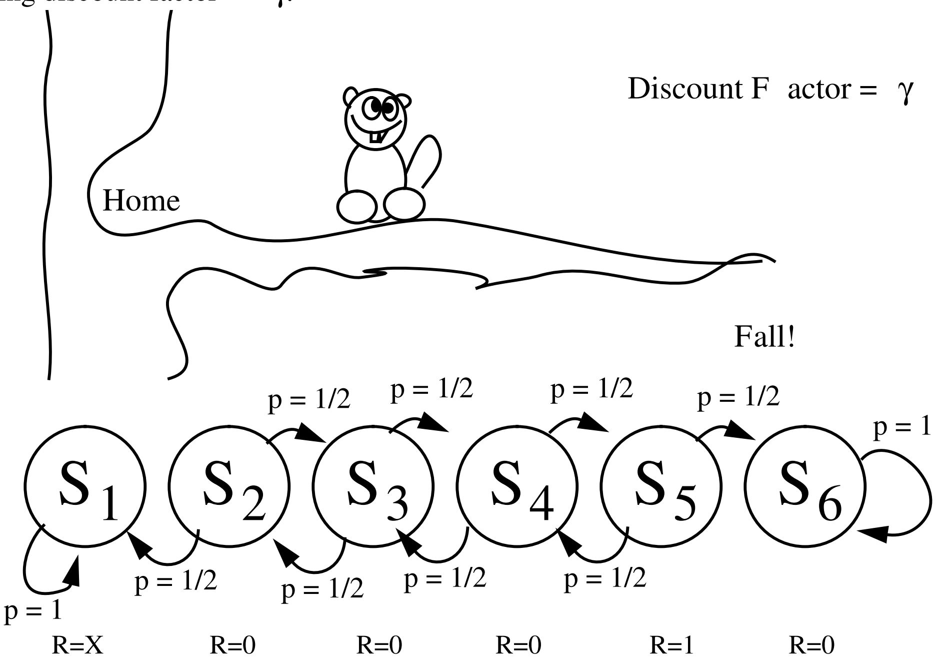
$\forall x. \forall y. FailedExam(x) \wedge FailedExam(y) \Rightarrow (x = y)$

$\forall x. FailedExam(x) \wedge (\forall y. FailedExam(y) \Rightarrow (x=y))$

$\exists x. (FailedExam(x) \wedge (\forall y. FailedExam(y) \Rightarrow (x=y)))$

7. Markov Systems (12 points)

You are a drunk squirrel wandering randomly left and right on a tree branch. Whenever you do a transition from your home state (S1) you get a reward of X where X is some positive number. No other states give you a reward, except for S5 which gives a reward of 1 unit because of its nice view. You move left or right at random at each time step, unless you're at home (in which case you stay at home permanently) or you've fallen off the tree (in which case you stay fallen off permanently). We are using discount factor γ.



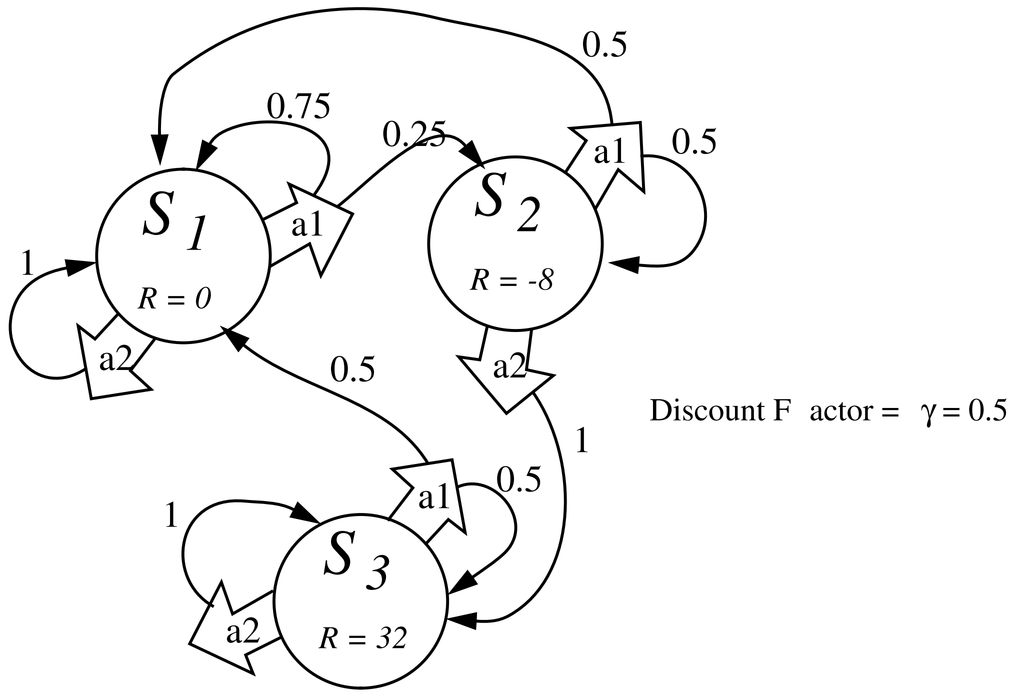
Define J_i = Discounted sum of future rewards starting from state S_i.

a. Which one of the following statements is true?

- $J_1 = X$
- $J_1 = X + \gamma J_1$
- $J_1 = X + \gamma X$
- $J_1 = X + J_1$
- $J_1 = \gamma + X J_1$

b. In addition to the equation you selected above, there are five other linear equations that together determine the values J1 through J6. Write down these five other equations. You don't need to solve them.

c. Define p_i = Probability that you'll eventually end up fallen off the tree starting from state s_i. Write down six linear equations relating the values p_1, p_2, p_3, p_4, p_5, and p_6



d. How many possible policies are there in the above Markov Decision Process?

e. What is the optimal action to apply in state S_1 (assuming you are trying to maximize your expected discounted sum of future rewards)?

f. What is the optimal action to apply in state S_2 (assuming you are trying to maximize your expected discounted sum of future rewards)?

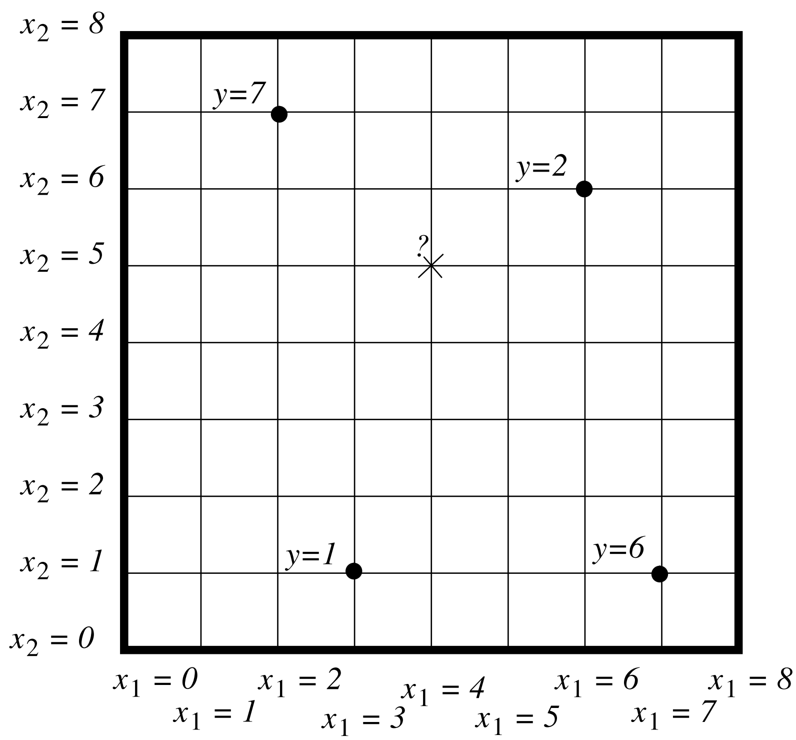
g. What is the optimal action to apply in state S_3 (assuming you are trying to maximize your expected discounted sum of future rewards)?

8. Nearest Neighbor Learning (10 points)

You are given the following simple dataset with inputs x1 and x2, and output y ..

x1	x2	y
2	7	7
3	1	1
6	6	2
7	1	6

These datapoints are depicted in the figure below , whic h positions the datapoints according to their x1 and x2 values .



a. What output value does nearest neighbor (using a conventional Euc lidean Distance metric) predict for query inputs ($x1 = 4$, $x2 = 5$). (This is shown by a question mark on the diagram).

b. What output value would 2-nearest-neighbor predict for the same query?

c. Sketc h the region of (x1,x2)-space for whic h nearest neighbor (i.e . 1-nearest neighbor) would give a prediction of 7.

d. What is the a verage absolute lea ve-one-out error of nearest neighbor (i.e . 1-nearest neighbor) on the dataset. "A verage absolute" means you should give the mean unsigned difference between predictions and true values .

9. Simulated Annealing (7 points)

You have been given a large propositional knowledge base containing hundreds of propositional symbols and thousands of sentences. You want to find an interpretation that makes the knowledge base evaluate to True. Describe how simulated annealing could be used for this purpose. Make sure you mention:

- What representation would you use for the thing you are optimizing?
- What would be the move-set?
- What would be the objective function?

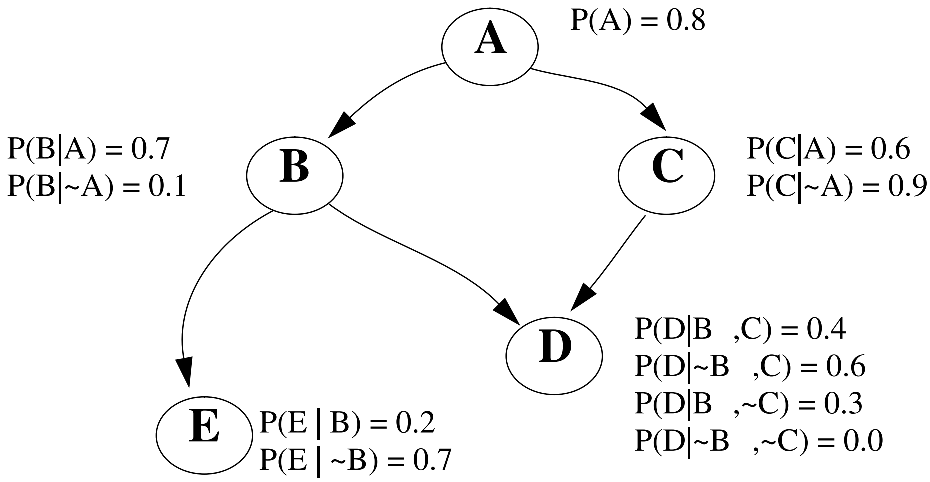
10. Harder Decision Tree Question (5 points)

Assume you have 100 input attributes, 1 output attribute, and 512 records (i.e. datapoints). Suppose each attribute value in each record is generated entirely randomly ...set to true with 50-50 probability, else set to false. Suppose the output values are also generated randomly in the same way. It is, of course, pretty futile to do learning in this case.

a. But suppose you did decide to try to learn using the standard decision tree algorithm (with no pruning). How many nodes do you expect the resulting decision tree to have? (Leaf values do not count as nodes).

b. Now, repeat your answer to the above, except suppose that there are only 5 input attributes instead of 100. Does that change your answer? If so, how?

11. Harder Bayesian Networks (8 points)



a. If this information had been expressed in a joint probability table instead of a bayesian network, how many entries would have been required?

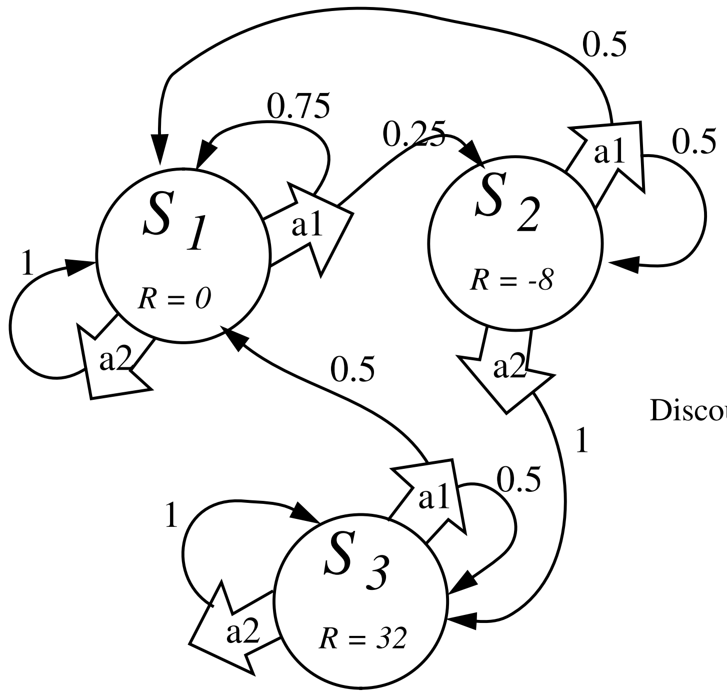
b. What is P(E)?

c. What is P(A ^ B ^ C ^ D | E)?

d. A, ~B and E have been observed. What is the probability that D is true?

12. Harder Markov Decision Process Question (6 points)

Returning to the same MDP you saw earlier ...



Discount Factor = $\gamma = 0.5$

Write $J^*(s_i)$ = The expected discounted sum of future rewards starting from state s_i assuming you act optimally.

a. What are the values of $J^*(s_1)$, $J^*(s_2)$, and $J^*(s_3)$?

b. Assume you are doing value iteration. Write $J^k(s_i)$ = the maximum possible expected discounted sum of future rewards during the next k timesteps. Fill in the remaining entries in the first four rows of the following value iteration computation (some of the entries have already been filled in for you).

k	$J^k(s_1)$	$J^k(s_2)$	$J^k(s_3)$
1	0	-8	32
2		8	48
3			
4			

13. Harder Simulated Annealing (5 points)

Returning to your answer to the question about using simulated annealing for finding an interpretation that satisfies your knowledge base. Construct a small example showing how it is possible for the search to be in a local minimum state in which any move chosen from the moveset would make the objective function worse.

Make sure your answer very clearly describes why it is a local minimum. Unclear answers will not receive any marks.

14. Harder Neural Nets (4 points)

I am trying to learn a function with one real number (x) as input and one real number (y) as output. I am convinced it is some kind of simusoid and so I use the following predictive function:

$$y = w_1 + w_2 \, sin(w_3 x + w_4)$$

Give the gradient descent rule for training this function.