16-731/15-780 Final, Spring 2002
** SOLUTION SET ***

1. Write your name and your andrew email address below.

   Name:

   Andrew ID:

2. If you need more room to work out your answer to a question, use the back of the page and clearly mark on the front of the page if we are to look at what's on the back.

3. You must answer at least 10 of the following 11 questions. Each question is worth a maximum of 10 points.

4. If you answer 11 questions then we will only count the top 10 scores (i.e. we will discard your worst score).

5. The maximum possible score is 10 * 10 = 100

6. You may use any and all notes, as well as the class textbook.

7. You have 3 hours.

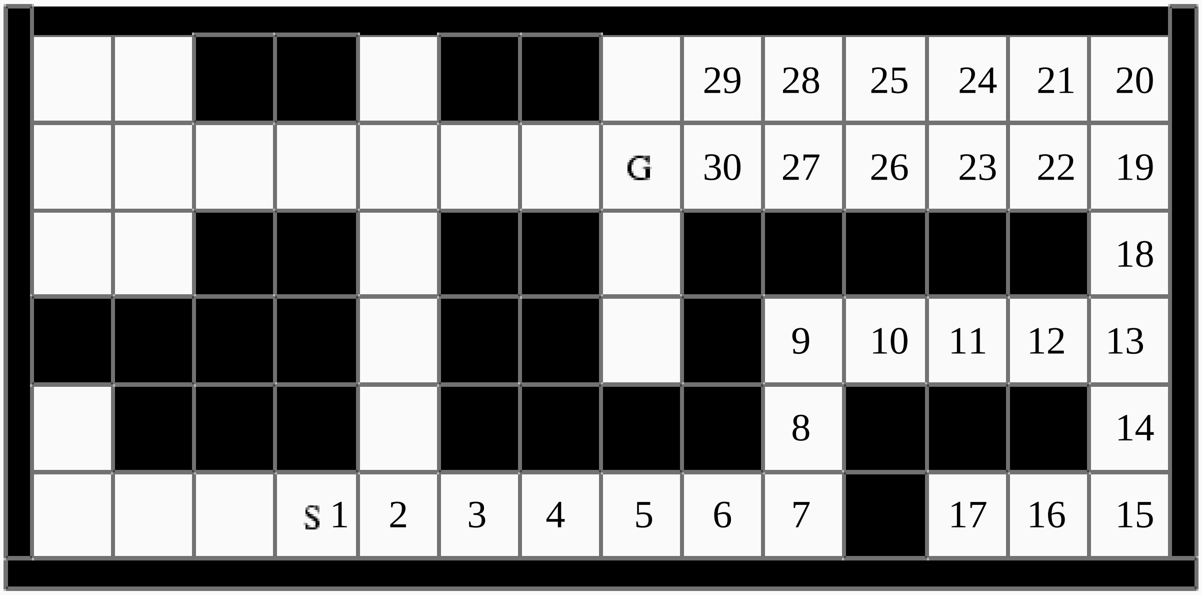
8. Good luck!

1 Depth First Search

Imagine a scenario with a robot trying to navigate in the following maze from the start position marked S to the end position marked G. At each step the robot can move in one of the four compass directions. The robot contemplates alternatives in the following order:

1. Move South
2. Move East (i.e. Right on this picture)
3. Move North
4. Move West (i.e. Left on this picture)

Mark the set of states that are expanded during the search, in the order they are expanded, by putting a 1 in the first state, a 2 in the second, and so forth. (Hint: put "1" in the cell marked "S"). Assume the search is Depth First Search (DFS). Use the version of DFS that avoids loops by never re-expanding a state that is on the current path.



		■	■		■	■		29	28	25	24	21	20
							G	30	27	26	23	22	19
		■	■		■	■		■	■	■	■	■	18
■	■	■	■		■	■		■	9	10	11	12	13
	■	■	■		■	■	■	■	8	■	■	■	14
			S 1	2	3	4	5	6	7	■	17	16	15

2 Robotics and Constraints

(a) Here is a circular robot that may move around the plane with two degrees of freedom. There are two circular obstacles. We want to plan robot motions in configuration space, where the configuration is the (x, y) position of the center of the robot.

a1. What is the diameter of the C-space representation of each obstacle?

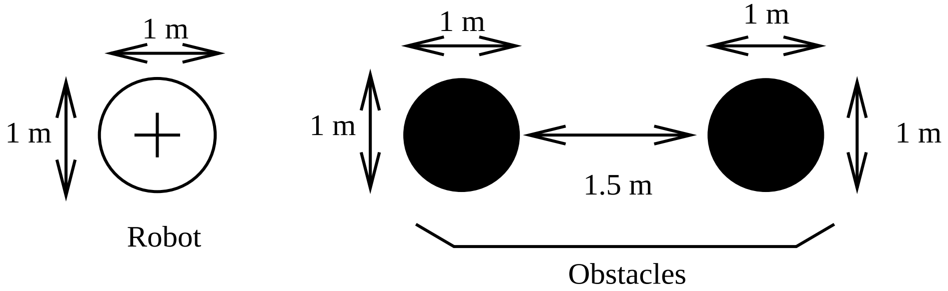
ANSWER: diameter = 2 m

a2. What is the distance between the obstacles in C-space?

ANSWER: distance = 0.5 m

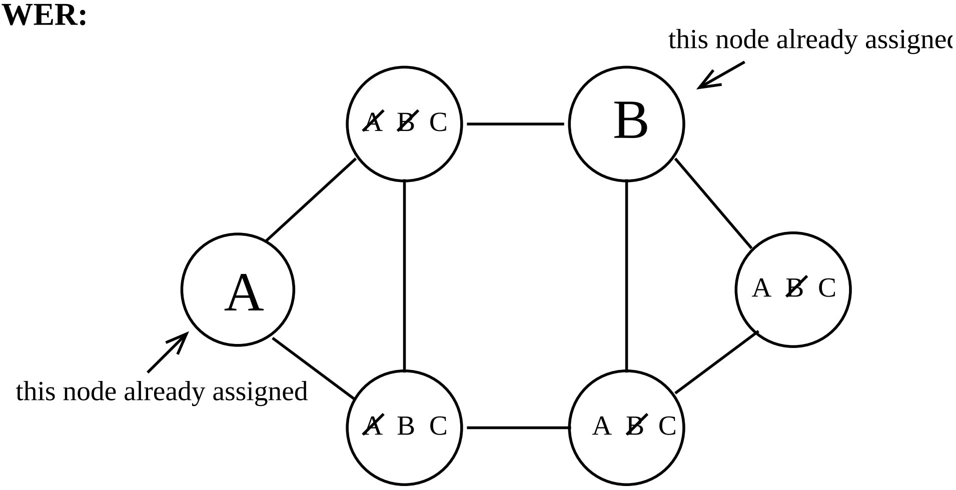
a3. Will there be a collision-free path between the obstacles in the C-space representation? (Circle one)

ANSWER: YES



(b) The following diagram concerns graph coloring problems. Each node must be labeled A or B or C and no adjacent nodes may have the same label. You run forward checking on the following problem. Remember that forward checking is the same thing as only doing constraint propagation for one step. What values remain for each variable after forward checking has finished, but before any DFS begins? (Simply cross-out any values that Forward Checking has eliminated)

ANSWER:



3 Linear Programming

(a) Re-express the following Linear Program as a new linear program in which the only inequalities are the Primary Constraints (i.e. constraints that all variables are non-negative). If you use any slack variables, call them y_1, y_2, etc.

Maximize $z = u + v + x$ subject to:

$$\begin{aligned} u + v &\geq 0 \\ v - w &\leq 0 \\ u + 3w + x &\leq 5 \end{aligned}$$

ANSWER:

$$\begin{aligned} u + v - y_1 &= 0 \\ v - w + y_2 &= 0 \\ u + 3w + x + y_3 &= 5 \end{aligned}$$

(b) Re-express the following Linear Program in Restricted Normal Form. If you use any slack variables, call them y_1, y_2, etc.

Maximize $z = 2u + v + w$ subject to:

$$\begin{aligned} u + v &= 10 \\ w - u &= 2 \end{aligned}$$

ANSWER:

$$\begin{aligned} v &= 10 - u \\ v &= 12 - w \end{aligned}$$

(c) In the following tableau, which variable gets moved to the Left Hand Side on the next iteration?

		x_4	x_5
z	2	1	2
x_1	10	-3	1
x_2	11	2	-2
x_3	12	3	3

ANSWER: We look for the variable on top that has the largest positive coefficient in the second row. This is $\mathbf{x_5}$.

(d) In the previous tableau, which variable gets moved to the Right Hand Side?

ANSWER: We look for the variable that would reach 0 first as we increase x_5. This is $\mathbf{x_2}$.

(e) What is the optimal solution (if any) to the following Tableau LP?

		x_4	x_5
z	-2	-1	-2
x_1	10	-3	-1
x_2	11	-2	-2
x_3	12	-3	-3

ANSWER: The x_4 and x_5 entries in the z row are both negative, so this tableau is already at a solution state, and we can just read off the values from the second column.

$$(\mathbf{z = -2});\ \ \mathbf{x_1 = 10};\ \ \mathbf{x_2 = 11};\ \ \mathbf{x_3 = 12};\ \ \mathbf{x_4 = 0};\ \ \mathbf{x_5 = 0}$$

(f) True or False: If an LP has no feasible solution, then it cannot be expressed in RNF.

ANSWER: When an LP is written in RNF form, setting all the decision variables to 0 is a feasible solution. So if there are no feasible solutions, there must not be a way to express it in RNF. **True.**

4 Propositional Logic

(a) The single most important idea in formal logic is that the meaning of a sentence can be represented as a set-theoretic set of possible worlds.

Given a universe with propositional symbols P, Q, R and S, how many interpretations are there in the meaning of the following KB?

$$(P \wedge Q) \Rightarrow (R \wedge S) \tag{1}$$
$$(P \wedge Q) \vee (R \wedge S) \tag{2}$$
$$\sim (P \wedge Q) \Rightarrow (\sim P \vee \sim Q) \tag{3}$$

ANSWER: Each possible interpretation assigns truth values to our four variables. One reasonable way to attack this problem is to make a big table with all $2^4 = 16$ possible interpretations and check individually if they are consistent with the KB.

But we can go faster if we are clever. First look at (3) in the KB. DeMorgan's rule tells us that $\sim (P \wedge Q) \Leftrightarrow (\sim P \vee \sim Q)$, so (3) boils down to $\sim (P \wedge Q) \Rightarrow \sim (P \wedge Q)$. That is a valid sentence, so it will be true in every interpretation and we can ignore it. Second, remembering that we can write logical OR in terms of implication, we can rewrite (2) as $\sim (P \wedge Q) \Rightarrow (R \wedge S)$. Resolve this with (1) and you realize that (1) and (2) together imply $R \wedge S$, and they don't end up saying anything about P and Q. So out of the 16 possible interpretations, the only ones consistent with the KB are the **4 interpretations** that have both R = True and S = True.

(b) The following four questions assume **we are in a world with four propositional symbols**. For each question, the answer might be "infinite".

b1. How many possible **interpretations** are there?

ANSWER: In propositional logic, an interpretation is just an arbitrary assignment of truth values to objects. There are four objects, each of which can take on one of two values (True, False). So there are **2^4 = 16 interpretations**.

b2. How many possible **meanings** are there?

ANSWER: A meaning is a set of interpretations. It can be any subset of the set of all interpretations. So there are **2^{16} meanings**.

b3. How many possible **sentences** are there?

ANSWER: We can keep adding new logical operations and variables to the end of a sentence until we are blue in the face. For instance $A \wedge A \wedge A \wedge A$ is a valid sentence, and we can keep making new ones of arbitrary length in the same form. There are an **infinite number of sentences**.

b4. How many possible **knowledge bases** are there?

ANSWER: A knowledge base is just an arbitrary collection of sentences. Since there are an infinite number of sentences, there are an **infinite number of knowledge bases**.

(c) It's rather vague to represent the meaning of a KB by a set of possible worlds. Pat proposes that it would be a better idea to simply say the meaning of a KB is going to be a single world. What is the best response to Pat's idea? (pick one of the responses below)

(i) Good idea—that will make semantics more clearly defined and automated proof more efficient, with no downside.

(ii) Bad idea—that will make logical inference more computationally expensive since we'll have to search much harder to find an interpretation that matches a knowledge base.

(iii) Bad idea—then the user who writes a knowledge base will inevitably specify the full world, and won't be able to express limited knowledge about part of the world.

(iv) Irrelevant idea—Pat's idea is mathematically equivalent to conventional propositional logical semantics anyway.

(v) My hovercraft is full of eels.

ANSWER: (iii)

(d) Is the following set of proof rules sound (yes or no)?

$$a \wedge b \vdash a \vee b$$

$$a \wedge b \vdash a \wedge b$$

ANSWER: Yes

5 First order Logic

Here are three sentences:

$$\begin{array}{lcl} S_1 : \forall x.\forall y. & P(x) \wedge Q(y) & \Rightarrow R(x) \vee S(g(x), y) \\ S_2 : \forall u.\forall v. & S(g(u), h(u, v)) & \Rightarrow R(h(v, u)) \\ S_3 : \forall w.\forall z. & P(h(w, z)) \wedge Q(h(z, w)) & \Rightarrow R(w) \end{array}$$

(a) Can you resolve sentences S_1 and S_2? If so, what sentence results when you use the most general unifier?

ANSWER: Yes. Here is the unifier and the result:

$$[x/u \;\; y/h(u, v)] \quad P(u) \wedge Q(h(u, v)) \Rightarrow R(u) \vee R(h(v, u))$$

(b) Can you resolve sentences S_1 and S_3? If so, what sentence results when you use the most general unifier?

ANSWER: No. Can't resolve.

(c) Can you resolve sentences S_2 and S_3? If so, what sentence results when you use the most general unifier?

ANSWER: No. Can't resolve.

Consider the following KB

$$\begin{array}{rc}
\exists u. & \text{Tenured}(u) \\
\forall v. & \text{Tenured}(v) \Rightarrow \text{Fat}(v) \\
\forall w. & \text{Fat}(\text{Boss}(w)) \Rightarrow \text{Fat}(w) \\
& \text{Boss}(\text{Andrew}) = \text{Bob} \\
& \sim \text{Tenured}(\text{Bob})
\end{array}$$

(d) Is “Fat” a predicate symbol, a function symbol, or a constant symbol?

ANSWER: Predicate symbol. We can tell because:

- The input is an object: Boss(w) is used as an input to Fat(), and Boss() is a function (see below), so Boss(w) is an object.
- The output is boolean: Fat(Boss(w)) is used on the left-hand side of an implication.

(e) Is “Boss” a predicate symbol, a function symbol, or a constant symbol?

ANSWER: Function symbol. We can tell because:

- The input is an object: Boss(Andrew) is used, and Andrew is an object.
- The output is an object: Boss(Andrew) is tested for equality with Bob, which is an object, so Boss(Andrew) must be an object.

(f) Assuming that Andrew and Bob are distinct objects, and assuming that there are no other objects in the universe, give the full, set-theoretic meaning of the above KB, expressed as a list of predicate-logic interpretations (Hint: The number of interpretations in the meaning is strictly greater than 1 and strictly less than 4).

ANSWER: There are three interpretations in the meaning of the KB:

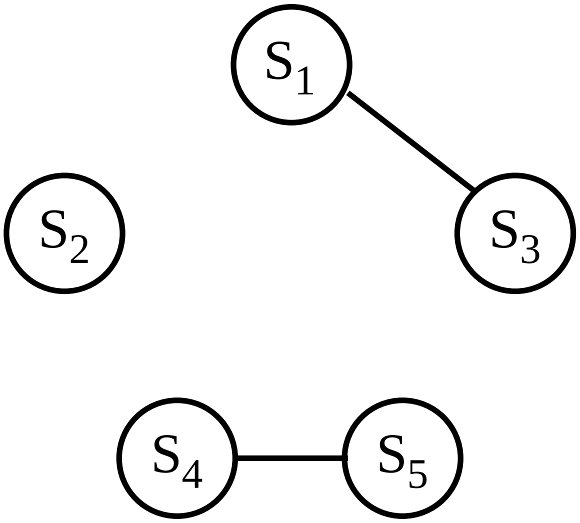
$$\begin{array}{l}
\text{Tenured} = \{\text{Andrew}\},\ \text{Fat} = \{\text{Andrew}\},\ \text{Boss} = \{(\text{Andrew}, \text{Bob}), (\text{Bob}, \text{Bob})\} \\
\text{Tenured} = \{\text{Andrew}\},\ \text{Fat} = \{\text{Andrew}, \text{Bob}\},\ \text{Boss} = \{(\text{Andrew}, \text{Bob}), (\text{Bob}, \text{Andrew})\} \\
\text{Tenured} = \{\text{Andrew}\},\ \text{Fat} = \{\text{Andrew}, \text{Bob}\},\ \text{Boss} = \{(\text{Andrew}, \text{Bob}), (\text{Bob}, \text{Bob})\}
\end{array}$$

Consider the following set of sentences

$$\begin{array}{lcl}
S_1: & \forall x.\exists y. & P(x,y) \\
S_2: & \exists y.\forall x. & P(x,y) \\
S_3: & \forall x. \sim \forall y. & \sim P(x,y) \\
S_4: & \sim \forall x.\forall y. & \sim P(x,y) \\
S_5: & \exists x.\exists y. & P(x,y)
\end{array}$$

(g) In the diagram below, draw a line between each pair of sentences that have the same meaning. For instance, draw a line between the S_1 node and the S_2 node if and only if S_1 and S_2 have the same meaning.

ANSWER:



6 Bayesian Networks

Assume that Mark Stehlik wears a kilt about once a year, and Jared Cohon wears a kilt about once every five years. Also, assume that Jared walks down a particular sidewalk every day, but Mark only walks down the sidewalk every other day[1].

(a) If you see a person walking down the sidewalk in the distance, and you are sure that he is either Mark or Jared, but you can't tell which, and you also can't tell what he is wearing, what are the following probabilities?

ANSWER: We know the person is either Mark or Jared (so the probabilities must sum to 1), and it is twice as likely to be Jared.

P(Person is Mark) = 1/3

P(Person is Jared) = 2/3

(b) Now suppose the person is wearing a kilt. What is the probability that it is Mark?

ANSWER: Suppose there are D days in a year. Abbreviate

- M = Person is Mark
- J = Person is Jared
- K = Wearing a kilt

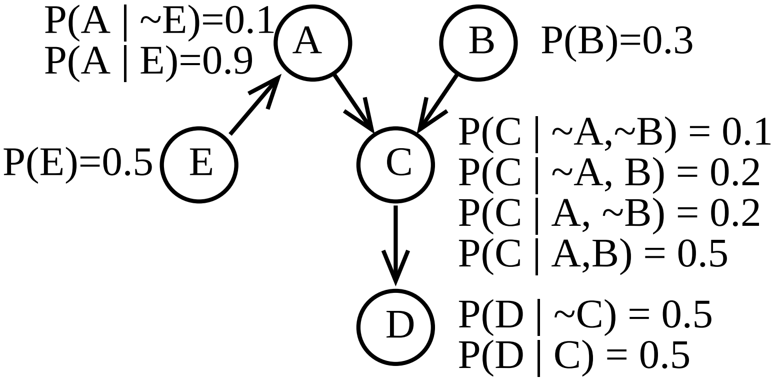
Then

$$\begin{aligned} P(M|K) &= \frac{P(K|M)P(M)}{P(K)} \\ &= \frac{P(K|M)P(M)}{P(K|M)P(M)+P(K|J)P(J)} \\ &= \frac{\frac{1}{D}\frac{1}{3}}{\frac{1}{D}\frac{1}{3}+\frac{1}{5D}\frac{2}{3}} \\ &= \frac{1}{1+\frac{2}{5}} \\ &= \frac{5}{7} \end{aligned}$$

P(Person is Mark | Wearing a kilt) = $\mathbf{\frac{5}{7}}$

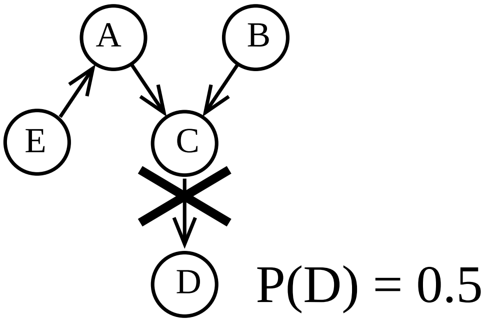
[1] every other day means once every two days

The questions on this page relate to the following Bayes net:



(c) One of the edges in the Bayes net above can be eliminated, and the conditional probabilities of the target of the edge can be changed so that we get an equivalent Bayes net. In the small diagram below, cross out the unnecessary edge, and write the new conditional probability table for the target of the edge.

ANSWER: Notice that $P(D \mid C) = P(D \mid \sim C)$, so we can remove the dependence of D on C to get the following:



For the following four questions, we ask you to calculate a quantity from the Bayes net above. Your answer should be a number. None of these calculations should take you longer than a minute or two.

(d)

$$\frac{\mathbf{P(E \mid A, C)}}{\mathbf{P(E \mid A)}} = ?$$

ANSWER: Recall from the lecture on Bayes net inference that the nodes E and C are d-separated by A. Therefore $P(E \mid A, C) = P(E \mid A)$ and the answer is

$$\frac{\mathbf{P(E \mid A, C)}}{\mathbf{P(E \mid A)}} = \mathbf{1}$$

(e)

$$\mathbf{P(C \mid A)} = ?$$

ANSWER: This is just old-fashioned cranking:

$$\begin{aligned} P(C \mid A) &= P(C \mid A, B)P(B) + P(C \mid A, \sim B)P(\sim B) \\ &= 0.5(0.3) + 0.2(0.7) \\ &= \mathbf{0.29} \end{aligned}$$

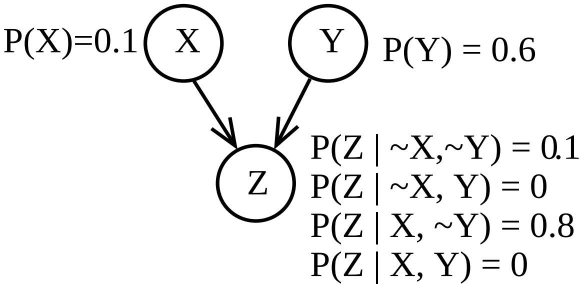
(f)

$$\mathbf{P(B \mid D)} = ?$$

ANSWER: We have already noticed in part (c) that D is effectively independent of the rest of the graph. Therefore $P(B \mid D) = P(B) = \mathbf{0.3}$.

7 Inference in Bayesian Networks

Consider the following Bayes network:



where the nodes represent the following events:

X. There is an enormous dust storm on Mars tonight. “There is a dust storm”.

Y. It is raining at my house tonight. “It’s raining”.

Z. I see what appears to be an enormous dust storm on Mars tonight from my backyard observatory. “I see a dust storm”.

(a) Common sense dictates that X and Y are independent; but according to our Bayes net inference techniques, if we know the value of Z, then knowing X might be able to help us predict Y, or vice versa. You tell your friend Trey this, but he is not very smart and doesn’t believe you. Give an example of a case when

(i) You start by knowing the value of Z (either true or false).

(ii) Somebody tells you the value of either X or Y.

(iii) The information given to you in step (ii) helps you to get a better prediction of the remaining unknown variable.

You must explain your example in the space below *using the English descriptions of the events* without saying X, Y, Z, or mentioning the exact probabilities. Your example should make sense given the probabilities in the net above, and it should convince even Trey that his common-sense notion is incorrect. Feel free to use the short versions of the English descriptions to save time.

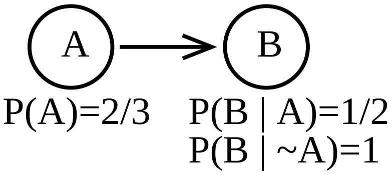
ANSWER: It will help to clarify this problem if we change it to the second person. Suppose I start by knowing that you didn’t see a dust storm, and then I learn that there was a dust storm. Now I can infer that it was probably raining at your house, and that’s why you didn’t see the dust storm.

Here is the tricky part of this question that caught some people: if I start by knowing that you *did* see the dust storm, then there is no way that information about X can help me to deduce Y or vice versa. Examine the probabilities: if Z is true, I can immediately infer that Y is false. Then if somebody tells me X, it doesn’t help me to infer Y (because I already knew Y). If somebody tells me Y, it doesn’t help me infer X (telling me Y is not new information).

Now we consider representing a knowledge base K using a Bayes net, so that Bayes net inference is equivalent to logical deduction, in the following sense: let p and q be propositions. Then

- Whenever $K \vdash p \Rightarrow q$, then Bayes net inference returns $P(q \mid p) = 1$.
- Whenever $K \vdash p \Rightarrow \sim q$, then Bayes net inference returns $P(q \mid p) = 0$.
- Otherwise, Bayes net inference returns $P(q \mid p) = c$, where $0 < c < 1$.

Here is an example: suppose our knowledge base K contains only the sentence $A \vee B$. The following Bayes net is equivalent to K:



Here are some tests that K and the net above are equivalent:

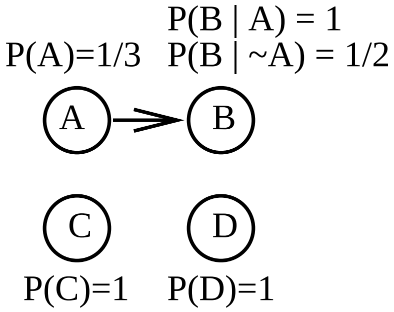
- According to K, $\sim A \Rightarrow B$. According to the Bayes net, $P(B \mid \sim A) = 1$. Check.
- According to K, $\sim B \Rightarrow A$. According to the Bayes net, $P(A \mid \sim B) = 1$ (in order to verify that, you need to apply Bayes rule and some algebra). Check.
- According to K, knowing that A is true does not allow us to infer the value of B. According to the Bayes net, $P(B \mid A) = 1/2$. Check.

(b) Suppose that the knowledge base L contains only the following sentences:

$$A \Rightarrow B$$
$$C \wedge D$$

In the space below, draw a Bayes net that represents L, labeling all the conditional probabilities appropriately. Use the smallest possible number of edges.

ANSWER: The derivation of this Bayes net is explained below.



The $C \wedge D$ sentence in the KB tells us immediately that C and D are true, so they must have probability 1, and it would be redundant to make any connections with them. It is similarly easy to guess that we will have a link from A to B to represent the implication.

The question then is: how do we assign probabilities such that $P(B \mid A) = 1$ and $P(\sim A \mid \sim B) = 1$ as the implication requires? In fact, there are many assignments of probabilities that do the trick, but you are unlikely to hit on a correct assignment by chance. Here is a way to generate working probabilities. We will enumerate all the interpretations for A and B that are consistent with the KB, and we will assign them each a non-zero probability so that the probabilities sum to 1. Then we can calculate any probability in the Bayes net by looking at our table. Here is the table that generated our solution:

A = True, B = True :	$1/3$
A = False, B = True :	$1/3$
A = False, B = False :	$1/3$

Notice we left out A = True, B = False, because it is not consistent with $A \Rightarrow B$. From this table, it is easy to calculate, for instance, $P(A) = 1/3$ and

$$P(B \mid \sim A) = \frac{P(\sim A \wedge B)}{P(\sim A)} = \frac{1/3}{2/3} = \frac{1}{2}$$

If you did not figure this technique out and got $P(A)$ wrong, you only lost one point.

8 Decision Trees

(a) This part of the question is going to create a counterexample to the hypothesis that we will always find the smallest decision tree consistent with the data. Assume we are using the decision tree algorithm described in the notes. Assume that all attributes (inputs and output) are binary, and assume that we do no pruning.

Construct an example dataset in which

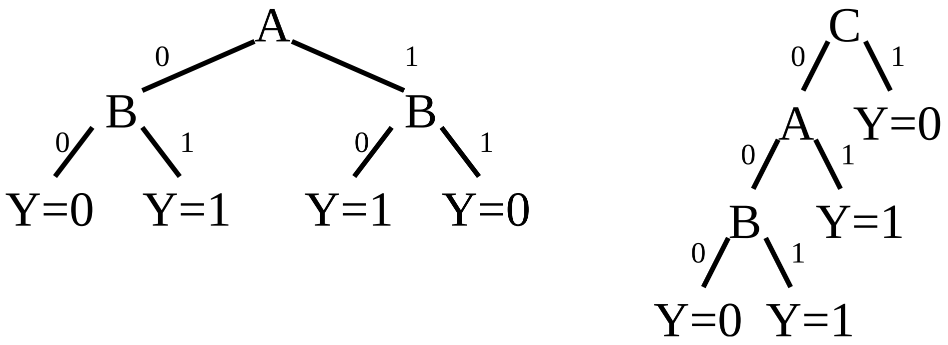
- ...it is possible to find a tree that gets zero training set error
- ...and our decision tree learning algorithm succeeds in finding a tree with zero training set error
- ...but there is a smaller, simpler, tree that also has zero training set error but which our decision tree learning algorithm fails to find.

You **must** explain the idea behind your example dataset, and why our decision tree learning algorithm fails to find he smallest solution.

ANSWER: Imagine we have the following dataset, where A, B, and C are inputs and Y is the output:

A	B	C	Y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

In the diagram below, we show the simplest zero-error decision tree on the left, and the tree our algorithm would find on the right. The reason our algorithm chooses to branch on C first is that $IG(Y \mid A) = IG(Y \mid B) = 0$. With this particular dataset, knowing A and B together lets you predict Y perfectly, but knowing either one alone doesn't help at all. Our algorithm is greedy, so it immediately chooses C instead of thinking about combinations of inputs.



Comparing the two trees, the one our algorithm generates is deeper, but it has the same number of nodes... but we could make our tree arbitrarily more complicated by adding more nuisance variables like C.

(b) Let $C(k, m)$ be the computational cost for our decision tree algorithm to build a non-pruned tree from a dataset with m attributes (all binary) and 2^k records.

Assume the time talken to compute one entropy calculation $H(Y|X)$ is $\alpha 2^k$, when there are 2^k records, and where α is some constant multiplicative factor. Furthermore, assume that all other costs in building the tree are negligable.

Assume that every time we split into two subtrees, the two subtrees contain equal numbers of records.

Define $C(k, m)$ in terms of $\alpha, k, m, C(k, m-1), C(k-1, m)$ and $C(k-1, m-1)$. Not all of these terms will appear in your formula!

ANSWER:

$$C(k, m) = m\alpha 2^k + 2C(k-1, m-1)$$

Note: $C(k-1, m-1)$ assumes that we ignore the split attribute in subsequent sub-trees. It's also fine to answer

$$C(k, m) = m\alpha 2^k + 2C(k-1, m)$$

9 Neural Networks

(a) Suppose we want to learn a perceptron but we are worried about the computational cost of computing sigmoids. We decide to use the following piecewise linear model instead:

$$y = h(w_1 + w_2 x)$$

where

$$\begin{aligned} h(z) &= 0 \text{ if } z \leq -1 \\ h(z) &= z \text{ if } -1 < z < 1 \\ h(z) &= 1 \text{ if } z \geq 1 \end{aligned}$$

we then define

$$\begin{aligned} \frac{dh}{dz} &= 0 \text{ if } z \leq -1 \\ \frac{dh}{dz} &= 1 \text{ if } -1 < z < 1 \\ \frac{dh}{dz} &= 0 \text{ if } z \geq 1 \end{aligned}$$

The gradient descent update rules are as follows. You must fill in the eight blanks in the equations below. You'll notice that we have broken the sum over datapoints into two sums. Define

$$\text{Zone} = \{k \text{ such that } |w_1 + w_2 x| < 1\}$$

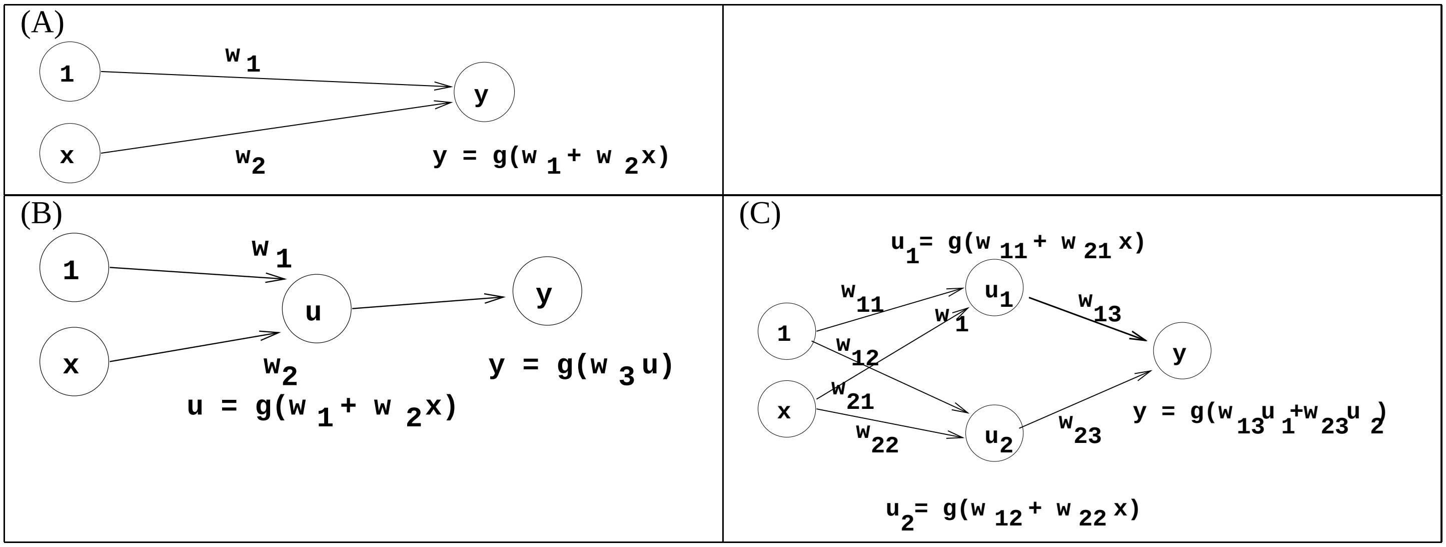
Assume we are using learning rate η. If you wish, you may use the variable δ_k in your answer, where δ_k is understood to denote $y_k - w_1 - w_2 x_k$.

ANSWER:

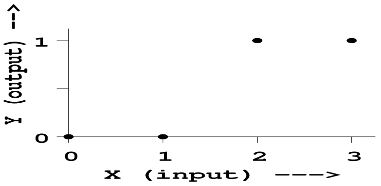
$$w_1 \leftarrow w_1 + \left(\mathbf{2}\eta \times \sum_{k \in \text{Zone}} \delta_\mathbf{k}\right) + \left(\mathbf{0} \times \sum_{k \notin \text{Zone}} \mathbf{0}\right)$$

$$w\_2 \leftarrow w\_2 + \left(\mathbf{2}\eta \times \sum_{k \in \text{Zone}} \delta_\mathbf{k}\mathbf{x_k}\right) + \left(\mathbf{0} \times \sum_{k \notin \text{Zone}} \mathbf{0}\right)$$

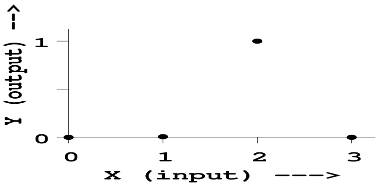
Consider the following three neural net architectures. All take one real-valued input x and predict output y which is constrained to be between 0 and 1 by a conventional sigmoid function $g(z) = 1/(1 + e^{-z})$.



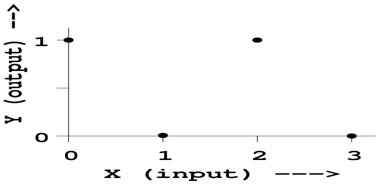
(b) Which (if any) is capable of achieving a sum-squared-error of less than 0.01 on the following four datapoints? Your answer should either be "none" or else some subset of $\{A, B, C\}$. **ANSWER:** A**,** B**,** C



(c) Which (if any) is capable of achieving a sum-squared-error of less than 0.01 on the following four datapoints? Your answer should either be "none" or else some subset of $\{A, B, C\}$. **ANSWER:** C



(d) Which (if any) is capable of achieving a sum-squared-error of less than 0.01 on the following four datapoints? Your answer should either be "none" or else some subset of $\{A, B, C\}$. **ANSWER: NONE**



10 Markov Decision Processes

(a) Suppose that Pat is going to use value iteration to find the optimal policy for a very large MDP. Pat notices that state S_{17} has five actions, and three of the actions have the following behavior.

$$\begin{aligned}
P(\text{NextState} = S_{25} | \text{ThisState} = S_{17} \wedge \text{Action} = a_1) &= 1 \\
P(\text{NextState} = S_{25} | \text{ThisState} = S_{17} \wedge \text{Action} = a_2) &= 0.3 \\
P(\text{NextState} = S_{33} | \text{ThisState} = S_{17} \wedge \text{Action} = a_2) &= 0.7 \\
P(\text{NextState} = S_{33} | \text{ThisState} = S_{17} \wedge \text{Action} = a_3) &= 1
\end{aligned}$$

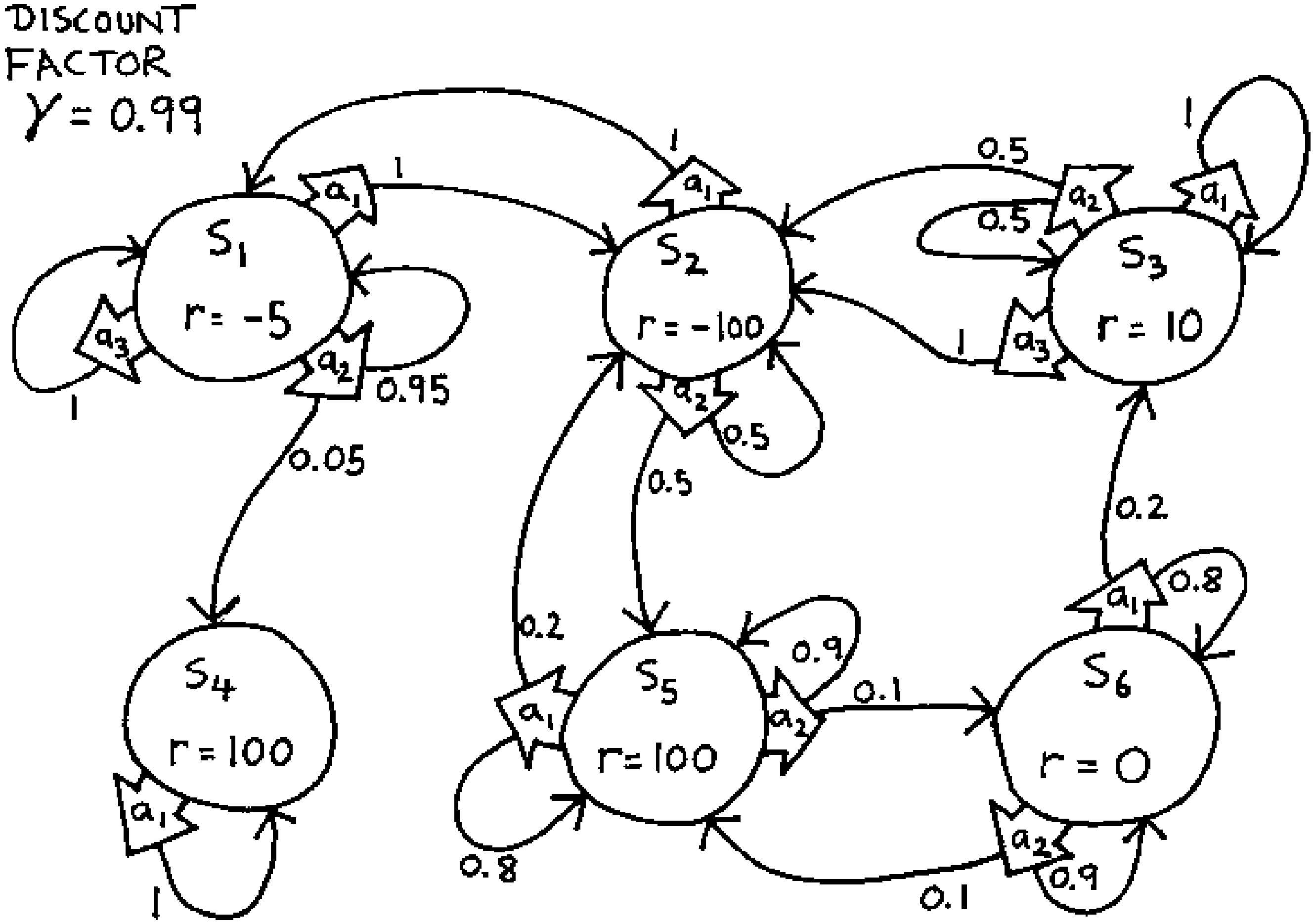
Pat decides to ignore the possibility of choosing action a_2 in state S_{17}. It will not be considered during value iteration and it will not be considered as a possible choice in the final policy. Is there any danger that Pat's decision will cause Pat to miss out on obtaining an optimal policy? *You must briefly explain your answer.*

ANSWER: Pat won't miss the optimum.

- If $J^*(s_{25}) > J^*(s_{33})$, then a_1 is better than a_2.
- If $J^*(s_{25}) < J^*(s_{33})$, then a_3 is better than a_2.
- If $J^*(s_{25}) = J^*(s_{33})$, then a_1, a_2, and a_3 have the same value.

So in no case is it essential to use action a_2.

(b) By thinking carefully, and perhaps hitting a few keys on your calculator, it should be possible to deduce the optimal policy for the following MDP without needing to run Value Iteration.



You must write down the optimal policy π below:

ANSWER:

$$\begin{aligned}
\pi(S_1) &= a_2 \\
\pi(S_2) &= a_1 \\
\pi(S_3) &= a_3 \\
\pi(S_4) &= a_1 \\
\pi(S_5) &= a_1 \\
\pi(S_6) &= a_1
\end{aligned}$$

We get this answer by noticing that γ is very close to 1, so we are most interested in the long-term reward. The best spot to be in for the long haul is S_4, because it has the highest reward and once in it we never leave. All of the actions were chosen to take us toward S_4.

11 Reinforcement Learning

Imagine an MDP with two states (S_1 and S_2) and in which S_1 has two actions (a_1 and a_2) and S_2 has only one action (a_1). Suppose the discount factor is γ and suppose you run Q-learning with a Q-table initialized with all zero values and with learning rate α.

(a) On the first transition, you start in state S_1, apply action a_1, receive an immediate reward of 1 and then land in state S_2. What is the resulting value of $Q(S_1, a_1)$? Give your answer as an algebraic expression that may include one or both of the symbols γ and α.

ANSWER: α

(b) On the second transition, you apply action a_1, receive an immediate reward of 0 and then land in state S_1. What is the resulting value of $Q(S_2, a_1)$? (Give your answer as an algebraic expression that may include one or both of the symbols γ and α.)

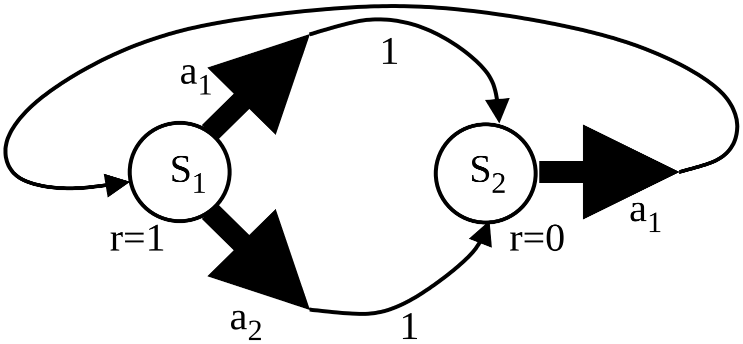
ANSWER: $Q = \alpha Q_{new} + (1-\alpha)Q_{old} = \alpha(\gamma\alpha) + 0 = \alpha^2\gamma$

(c) On the third transition, you apply action a_2, receive an immediate reward of 1 and then land in state S_2. What is the resulting value of $Q(S_1, a_2)$? (Give your answer as an algebraic expression that may include one or both of the symbols γ and α.)

ANSWER: $Q = \alpha Q_{new} + (1-\alpha)Q_{old} = \alpha(1 + \gamma(\alpha^2\gamma)) + 0 = \alpha + \alpha^3\gamma$

(d) Suppose that you were using Certainty Equivalent learning. What would be the estimated values of $J^*(S_1)$ and $J^*(S_2)$ after having observed those first three transitions? (Give your answers as two algebraic expressions that may include the symbol γ.)

ANSWER: Certainty equivalent learning would learn the following MDP structure from the observations so far:



Abbreviate $J_1 = J^*(S_1)$ and $J_2 = J^*(S_2)$. From the MDP we get the recurrence:

$$\begin{aligned} J_1 &= 1 + \gamma J_2 \\ J_2 &= \gamma J_1 \end{aligned}$$

which we can solve to find

$$\begin{aligned} J_1 &= 1 + \gamma^2 J_1 \quad \Rightarrow \quad J_1 = \frac{1}{1-\gamma^2} \\ J_2 &= \gamma J_1 = \frac{\gamma}{1-\gamma^2} \end{aligned}$$