Second Midterm, Answers and FFQ CSC 242

Name:_____

*Write your **NAME** legibly on your bluebook(s) **AND EXAM, WHICH YOU WILL ALSO TURN IN**. There are 75 minutes worth of problems and 15 minutes of extra or alternative credit. Work all the problems you can. You may use two double-sided pages of notes. Please hand your notes in with your bluebook.* **Please start each new problem on a separate page. Any order is fine.**

1. NewPage! Conditional Probabilities (10 min)

Prove these statements or give a counterexample:

A) (5 min) If $P(a \mid b, c) = P(b \mid a, c)$, then $P(a \mid c) = P(b \mid c)$.

B) (5 min) If $P(a \mid b) = P(a)$, then $P(b \mid c) = P(b)$.

(Hint: full credit for math (or probability assignments), half credit for semantic arguments in English.)

Ans: A) $P(a \mid b, c) = P(b \mid a, c)$ means, by definition, that $P(a, b, c)/P(b, c) = P(b, a, c)/P(a, c)$, from which "by a simple arithmetical process you'll easily discover" that $P(b, c) = P(a, c)$, so divide both sides of that by $P(c)$ and you get $P(a \mid c) = P(b \mid c)$.

B) If a and b are independent, then b and c have to be independent too? Puh-leeeeze. If you don't pay attention to what I say, does that prove I don't pay attention to what State law says? So by the def. of cond. prob. the claim is equivalent to $P(a, b) = P(a)P(b) \Rightarrow P(b, c) = P(b)P(c)$. Since this is just silly we should be able to construct a contradictory joint pretty easily.

Here's a 3-D boolean proto-joint that only needs normalization (divide all numbers by their sum) to be a true PDF.

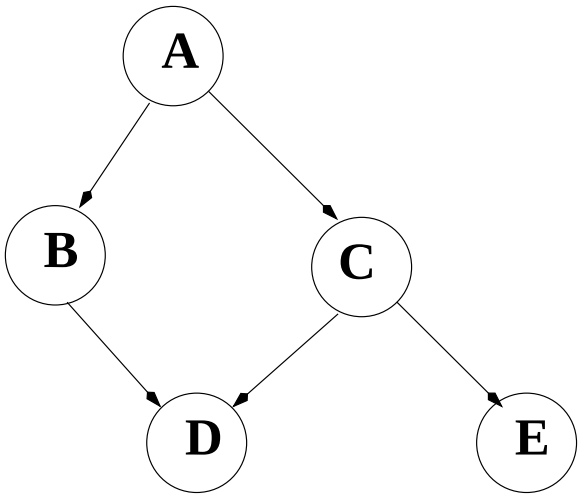
```
        B                       B
    T       F               T       F
  T 1       1               0       0
A
  F 2       2               1       1

       C is T                  C is F
```

Here, P(A) =1/4, PB =1/2, P(C) =3/4, P(AB) = 1/8 and P(A)P(B) = 1/8 (so IF condition satisfied), but P(AC) = 1/4 and P(A)P(C) = 3/16, (so the THEN isn't).

2. New Page! Bayes Nets (25 Mins)

Given this Bayes net structure over Boolean random variables and all its associated prior and conditional probabilities:



A) (5 min) Write out $P(D \mid B, C)$ using as few of the following symbols as you can: $(,), *, +, P(A), (P(B), P(C), P(E), P(B \mid A), P(C \mid A), P(D \mid B), P(D \mid C), \alpha$.

Ans. FFQ number 1. This ME making the 'capital vs lowercase' mixup that lots of you did here and there. Capital letters are for PDFs, little ones for values. But once I mistyped that, I got totally off track, so it was a typo followed by a thinko.

This question should have been: "How do we figure out the value of $P(d \mid \neg b, c)$?", or somesuch, since $P(D \mid B, C)$ is one of the "associated prior and conditional probabilities", and not obviously derivable from anything else we know in a useful way. So I gave everybody 3 points and myself 0.

B) (5 min) Using the properties of the Bayes Net, what is the formula for the probability of the particular atomic event $P(a, \neg b, c, d, \neg e)$?

Ans. $P(a)P(\neg b \mid a)P(c \mid a)P(d \mid \neg b, c)P(\neg e \mid c)$

C) (5 min) Use inference by enumeration to write down a formula for $P(d \mid \neg b, e)$.

Ans. $P(d \mid \neg b, e) = \alpha P(d, \neg b, e) = \alpha \sum_a \sum_c P(a)P(\neg b \mid a)P(c \mid a)P(d \mid \neg b, c)P(e \mid a)$, where $\sum_a$, for instance, means "sum over all possible values of a".

D) (5 min) Recall that in a Bayes Net a random variable (node) is conditionally independent of its non-descendents given its parents. The *Markov blanket* of a node is its parents, its children, and its children's parents. Convince me (or mathematically prove) that a variable is independent of all other variables in the network, given its Markov blanket.

Ans. We know that a node depends on its parents by construction, and that by Bayes rule we can also say the values of its children influence it as, by construction, it must influence them. BUT it's not all that influences them. Any other parents of a child must influence the child as well, by construction., and hence their parent node we're discussing. However that's the end of local influence on the node, since any more nodes involved would violate the conditional independence construction. Mathematically proving this is an exercise in Russell and Norvig.

E) (5 min) We are running a Markov Chain Monte Carlo Gibbs Sampling process to determine node (RV) probabilities. The current state is $(a, \neg b, \neg c, \neg d, e)$. We decide to sample for a new value of B. What distribution do we sample?

Ans. The Markov blanket of B is [A,D,C], so

$P(b) = \alpha P(b \mid a) P(\neg d \mid b, \neg c) P(\neg c | a)$
$P(\neg b) = 1 - P(b)$.

3. New Page! Reinforcement Learning: 25 Min.

You are a passive learning agent; you follow a fixed policy. You always start in state S1. You make three trials in your state space, each of which ends in the terminal state S3, which has a reward of 10. Your experience (actions, rewards) are as follows (your state sequence reads left to write interspersed by the actions that cause state transitions, with reward below each state). Trials 1 and 3 are identical, that's not a misprint.

Trial	State	action	State	action	State	action	State
	Reward		Reward		Reward		Reward
1	S1	A	S1	A	S2	B	S3
	-1		-1		-3		10
2	S1	A	S2	B	S2	B	S3
	-1		-3		-3		10
1	S1	A	S1	A	S2	B	S3
	-1		-1		-3		10

For the following, initially your guess at the utilities of all states is 0. When you hit a terminal state you are entitled immediately to set its utility to its reward. Your discounting constant $\gamma = .9$. Your learning constant $\alpha = .5$,

First, assume you're using Adaptive Dynamic Programming (ADP) to learn.

A) (5 min) At this point (end of trial three), what is your estimate of $T(s, a, s')$?

Ans. T(S1, A, S1) = 2/5, T(S1,A,S2) = 3/5, T(S2, B, S2) = 1/4, T(S2, B, S3) = 3/4. Rest are 0.

B) (5 min) At the end of trial three you decide to use value iteration to compute the utilities of states. What is your first estimate of U(S2)? Show the relevant numerical expression (i.e. substitute numerical values into the definition) but don't evaluate it.

Ans $U(S2) = R(S2) + \gamma \max_{A,B}[T(S2, B, S2)U(S2) + T(S2, B, S3)U(S3)]$
$= -3 + .9[(1/4) \cdot 0 + (3/4) \cdot 10] = .9(7.5) - 3.$

C) (3 min) Besides value iteration, how might you have computed this utility? (hint: you would also get U(S1)).

Ans. You could just solve the Bellman equations as in Russell and Norvig equation 17.10. Or use direct utility estimation (average your experiences).

D) (5 min) Instead of value iteration, assume instead you use Temporal Difference (TD) to learn. After trial 1 above and the TD computation, what is your estimate of U(S2)? Show

the relevant numerical expression (i.e.substitute numerical values into the definition) but don't evaluate it.

Ans. the TD equation is Russell and Norvig eq. 21.3: $U(S2) \leftarrow U(S2)+\alpha[R(S2)+\gamma U(S3)-U(S2)]$, or
$U(S2) \leftarrow 0 + .5[-3 + .9 \cdot 10 - 0] = 3$.

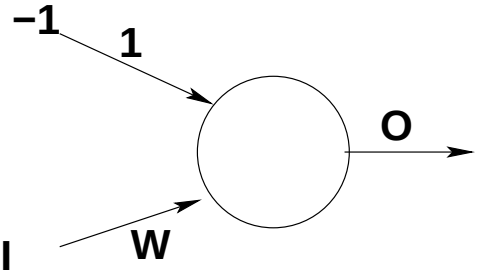
E) (4 min) What is *Q-learning*? (An update equation and explanation would be best, otherwise explain its idea and characteristics as precisely as you can).

Ans. Q(s,a) measures the value of performing action a in state s. Q is related to utility but can be computed by a TD-like computation without knowing anything about T(s,a,s'). It's like learning a policy without the underlying exact probabilies or utilities. The update equation is R&N 21.8:
$Q(s,a) \leftarrow Q(s,a) + \alpha[R(s) + \gamma \max_\beta Q(t,\beta) - Q(s,a)]$, recalculated whenever doing a in s lands you in state t.

F) (3 min) What is an *exploration function*? What is it trying to achieve and can you give and justify an example exploration function?

Ans. An exploration function raises utility of less-explored states, so it's a function $f(s,u,n)$ of the state, its current utility estimate, and the number of times it's been visited, like:
$f(s,u,n) = SomeOptimisticValue \;\; if \;\; n < N$
$f(s,u,n) = u \;\; if \;\; n \geq N$
for some desired-amount-of-visits N.

4. New Page! Perceptron Learning: 15 Min.



Here is a one-input perceptron with its threshold θ implemented as a constant -1 input multiplied by a constant weight (here set to 1). The input I (that's not a "one") is weighted by W, and initially $W = 1$. Its output is labeled O (that's not a "zero").

Its activation function is a step function that generates the output 1 (accepts the input) if $WI + \theta \geq 0$ and 0 (rejects) if $WI + \theta < 0$. The learning rate $\alpha = .1$.

We want the perceptron to accept (have $O = 1$) for any input $I \geq 5$, and reject any input less than 5.

A) (5 min) It's simple to solve for the proper weight W that implements the test in the previous paragraph. What should it be?

Ans. We know the output should go from 0 to 1 at the point $W \cdot 5 - 1 = 0$, so $W = .2$

B) (10 min) Now let's learn W. You present three inputs [4,6,4], and use the corresponding correct answers [0,1,0] along with the perceptron learning rule to adjust W from its initial value of 1. What happens? That is, please fill in the blanks in the table below. It shows

current Weight W, Input, Activity (WI + threshold) , Output, true answer, signed error, and new Weight. (Hint: you only need these three trials).

Trial	Curr. W	I	WI+thresh.	O	True Ans	Err	New W
1	1	4			0		
2		6			1		
3		4			0		

Ans.

Trial	Curr. W	I	WI+thresh.	O	True Ans	Err	New W
1	1	4	3	1	0	-1	.6
2	.6	6	2.6	1	1	0	.6
3	.6	4	1.4	1	0	-1	.2

5. New Page! Extra or Alternative Credit: Wanna Bet? (15 min)

Mr. A believes that a particular coin has probability .8 of coming up heads and .1 of coming up tails. He doesn't think there are any other possible outcomes. Mr. B was listening in 242 class and wants to pull a DiFinetti on A to take A's money. The trick is to take both sides of the bet and exploit A's violation of some axiom of probability. (Aside: what axiom is it?) B thus proposes a wager of two separate bets with A, each on terms that A thinks fair, given his beliefs. B's two bets are: "H will come up", and "T will come up".

A) (10 min) What happens? That is, please fill in the blanks in the table below. Also summarize the results: i.e. for the two wagers taken together, does A always win, always lose, always break even, or only win sometimes?

B's Bet	A's Belief	B:A stakes	A's gain if H	A's gain if T
H	.8	8:2		
T	.1			

(Hint: Reading the first line, B bets that H will come up. A believes H comes up .8 of the time, so B must put up 8 (dollars, say) and A must only put up 2 for A to perceive the bet as fair. If H actually comes up, A loses so his (negative) 'gain' is .... etc.)

Ans:

B's Bet	A's Belief	B:A stakes	A's gain if H	A's gain if T
H	.8	8:2	-2	8
T	.1	1:9	1	-9

So A loses in every such 2-bet wager: no matter what happens he pays 1 to B.

B) (5 min) Mr. A believes that Mr. B's coin has probability .5 of coming up heads and .5 of coming up tails. In fact it has probability .75 of coming up heads and .25 of coming up tails. Can B create a wager, possibly of multiple bets, to win off A *on every toss*? What is it, or why not?

Ans. No. for any finite set of bets, B could just get unlucky. Possible Moral?: it's better to be totally ignorant of outcome probabilities than consistently to violate an axiom of rational behavior.