Student Number : ___________________________ UW CSE 473 Midterm Exam, Autumn 2016

Student Name (write it on ***just*** this first page):________________________

Staff use only!

**Instructions**

Please answer clearly and succinctly. If an explanation is requested, think carefully before writing. Points may be removed for rambling answers. **If a question is unclear or ambiguous, feel free to make the additional assumptions necessary to produce the answer.** State these assumptions clearly; you will be graded on the basis of the assumption as well as subsequent reasoning. **On multiple choice questions, leaving the answer blank will give you one free point!** Wrong answer = zero points. Only make informed guesses.

**You have 110 minutes -** 8:30am – 10:20am

Exam is closed book, but you may bring and use one 8.5 x 11' double-sided page of notes

**1. True/False** (2 points each – **leave answer blank for one free point**) Circle the correct answer.

T For a search problem, the path returned by uniform cost search may change if we add a positive constant C to every step cost.

T If h1(s) and h2(s) are two admissible A* heuristics, then their average h3(s) = ½ h1(s) + ½ h2(s) must also be admissible.

T A consistent heuristic (that is zero at every goal state) is always admissible.

F Pattern databases are a way to scale search to large problems, but since the resulting tabular function is usually inadmissible, they rarely return an optimal solution

F Backtracking search for CSPs is a kind of breadth-first search.

T One way that forward checking can speed CSP-solving is by detecting dead-end choices in backtracking search.

T Decreasing the discount factor, $\gamma$, in an MDP tends to make an agent favor short-term rewards over long-term rewards.

T Long term, an epsilon-greedy approach with epsilon=0.5 does a better job of optimizing simple regret compared to cumulative regret.

T Standard value iteration will update every state each iteration, even if the values of the neighbors have not changed.

F In both policy iteration (PI) and value iteration (VI), one must specify a convergence threshold, $\epsilon$, to check if the policy (in PI) or value function (in VI) has converged.

F As long as the number of particles is less than the number of states, particle filtering should never place two particles in the same state as time progresses.

F In a Bayes' net, if $A \perp\!\!\!\perp B$, then $A \perp\!\!\!\perp B \mid C$ for some variable C other than A or B.

F For a discrete Bayesian network with n variables, the amount of space required to store the "joint" distribution table is O(n).

T A hidden Markov model is dynamic Bayesian network.

F In a hidden Markov model, future states are independent of the present state given the past.

T When using a hidden Markov model for speech recognition, the observations are typically the acoustic signals and the hidden states are the recognized words.

F In hidden Markov models, particle filtering is a technique used to perform exact inference for the filtering problem.

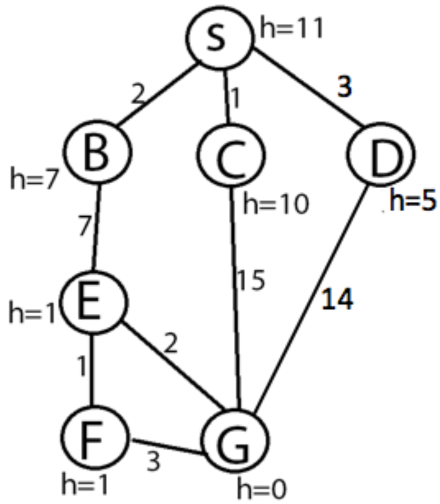
T Learning the structure of a Bayes net can be performed using local search guided by the probability that a given network might produce the observed data.

F By using expectation maximization (EM) one can learn optimal parameter values, even when a random variable is unobserved.

## 2) Search

Consider the search graph where all edges are bidirectional. Suppose we want to go from start state S to goal state G is the goal state. There are four possible solution paths.

1. S-B-E-F-G
2. S-B-E-G
3. S-C-G
4. S-D-G




Assuming ties (e.g., which child to first explore in depth-first search) should be resolved alphabetically (i.e. prefer A before Z), write the number of the solution returned to the right of each of the algorithms below.

a) (2 points) Depth-First Search returns solution number _1________

b) (2 points) Breadth-First Search returns solution number _3________

c) (2 points) Uniform Cost Search returns solution number _2________

d) (2 points) Greedy Search returns solution number _4________

e) (2 points) A* Search returns solution number _2________

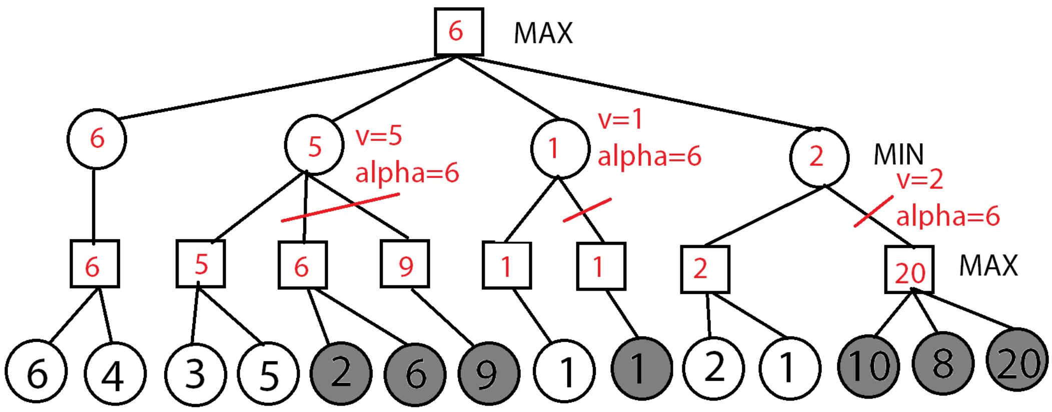
f) (2 points) Iterative Deepening (ID) Search returns solution number _3________

g) (2 points) IDA* Search returns solution number _2________

**3) Adversarial Search.** Given the search tree shown below….

a) (2 points) Fill in the squares and circles with the backed-up values resulting from a regular minimax search.

b) (4 points) Show how a depth-first search with alpha-beta cutoffs would work, indicating all $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ cut-offs by drawing a line through the unexplored branches. As usual, assume that children are explored from left to right.



**4) Constraint Satisfaction.** Consider a CSP with 5 Boolean variables, P, Q, R, S, T, and the following constraints:

i) $P \vee Q \vee S$

ii) $R \wedge \neg S$

iii) $T \vee \neg Q$

iv) $S \vee T$

| Variable | Values | |
|---|---|---|
| P | T | F |
| Q | T | F |
| R | T | ~~F~~ |
| S | ~~T~~ | F |
| T | T | ~~F~~ |

a) (3 points) Is this CSP arc consistent? **No**

If not, simulate the behavior of AC-3 and **cross off** any inconsistent values in the table to the right using a slash like this: **/**

b) (3 points) Simulate the behavior of backtracking search using the minimum remaining values (MRV) heuristic, and least constraining values (LCV) heuristic, breaking ties in lexicographic order (ie P<Q and F<T). **Circle the values** forming a solution (if one exists) or if no solution exists **put an X** in the box over on the right margin… right over here →

c) (5 points) Define the cryptarithmetic problem (shown to the right) as a CSP, using A, E, H, L, P, and T as variables with single digits as values (i.e., 0, …, 9), subject to the usual constraints: all variables must take unique values, and leading zeros are not allowed. You may add additional variables as needed, but define their ranges. Write all the necessary constraints below:

```
      E A T
+   T H A T
-----------
  A P P L E
```

Solution 1: Add two new variables, X and Y, with ranges {0, 1}

T=T = E + 10*X

A+A+X = L+10*Y

E+H+Y = P + 10

T + 1 = P + 10

A = 1

All-diff (E, A, T, H, P, L)

E,T,A != 0

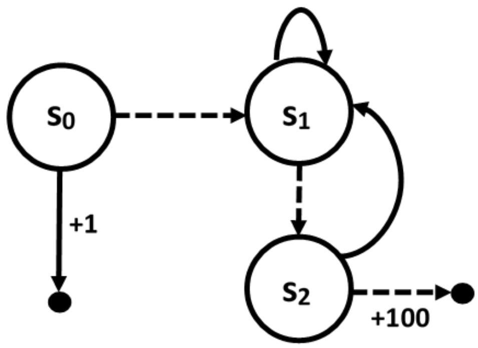
Solution 2:

(100*E + 10*A +T) + (1000*T + 100*H + 10*A + T) = (10000*A+ 1000*P + 100*P + 10*L + E)

All-diff (e,a,t,h,p,l)

E,T,A != 0

**5) MDPs.** Consider the 3-state MDP shown to the right. There are two actions, one represented by a dashed line and the other by a solid line. For simplicity, all transitions are deterministic. Transitions are labelled with the immediate reward that results from taking that transition (unlabeled transitions have zero reward). An edge terminating at a black dot means we terminate after receiving that reward. The discount factor, $\gamma$, is 0.5.




We wish to run policy iteration to find the optimal policy in this MDP. Initially, our policy is to always take the solid action at each state.

(a) (6 points) Compute the first 2 iterations of policy evaluation (of the initial policy)

| | $s_0$ | $s_1$ | $s_2$ |
|---|---|---|---|
| $V_0$ | 0 | 0 | 0 |
| $V_1$ | $V_i(s) = \sum_{s'} T(s,\pi(s),s')(R(s,\pi(s),s') + \gamma V_{i-1}(s'))$ 1(1 + 0.5 * 0) (no reward after termination) =**1** | 1(0 + 0.5(0)) = 0 | 1(0 + 0.5(0)) = 0 |
| $V_2$ | 1(1 + 0.5 * 0) (no reward after termination) =**1** | 1(0 + 0.5(0)) = 0 | 1(0 + 0.5(0)) = 0 |

(b) (3 points) Imagine we now move on to the next phase, policy improvement, using $V_2$. What policy will this phase produce? (Circle one for each state)

| | | | | |
|---|---|---|---|---|
| $s_0$: | Solid Action | Dashed Action | Tie | Has Q-value 1 (vs 0) |
| $s_1$: | Solid Action | Dashed Action | Tie | Both have Q-value 0 |
| $s_2$: | Solid Action | Dashed Action | Tie | Has Q-value 100 (vs 0) |

(c) (2 points) In one or two sentences, describe what should happen next (when running policy iteration) and why.

*First we check and see if the policy has changed to decide whether or not to terminate. It has at $s_2$, so we now evaluate the policy produced in step b.*

**6) Reinforcement Learning.**

Consider doing Approximate (Linear) Q-Learning in a simple domain where states are described by integers and where there are two possible actions: 1 and 0. In this problem we use two binary features, which are defined as follows given a state action pair (s,a):

- The first feature is 1 if (s + a) is even, 0 otherwise
- The second feature is 1 if (s + a) is divisible by 3, 0 otherwise

For example, if s=1 and a=1, s+a=2, and so the feature vector is: 1,0. The learning rate ($\alpha$) is 0.5 and the discount factor ($\gamma$) is 0.5. Our agent starts off with weights of 0,0, but acts in this world to receive some experience and learn better weights.

(a) (8 points) Fill in the table below with the learned weights after each piece of experience is received. (We can only give partial credit if you show your work).

| Observed Data | Weights after seeing data | |
|---|---|---|
| | $W_1 = 0$ | $W_2 = 0$ |
| Initial State: 3<br>Action: 1<br>Reward: +20<br>Final State: 15 | $F_1 = 1$ (3+1=4, which is even)<br>$Q(s,a) = W_1 * F_1 + W_2 * F_2 = 0$ (for all s,a since weights are zero)<br>diff = $(r + \gamma \max_{a'} Q(s',a')) - Q(s,a)$<br>= (20 + 0.5*0) -0<br>= 20<br>$W_1 = W_1 + \alpha$ * diff * $F_1$<br>$W_1 = 0 + 0.5*20*1$<br>= **10** | $F_2 = 0$ (3+1 =4, which is not divisible by 3)<br><br>$W_2 = W_2 + \alpha$ * diff * $F_2$<br>$W_2 = W_2 + 0.5*20*0$<br>= 0 + 0 = **0** |
| Initial State: 15<br>Action: 0<br>Reward: +17<br>Final State: 9 | $F_1 = 0$ (15+0=15, which is not even)<br>$Q(s,a) = W_1 * F_1 + W_2 * F_2$<br>$Q(s,a) = 10 * 0 + 0 * F_2 = 0$<br>$Q(s',0) = 10*0 + 0*1 = 0$ (9+0=9 → Features 0,1)<br>$Q(s',1) = 10*1 + 0*0 = 10$ (9+1=10 → Features 1,0)<br><br>diff = $(r + \gamma \max_{a'} Q(s',a')) - Q(s,a)$<br>= (17 + 0.5*max(0,10)) -0<br>= 17+5<br>= 22<br>$W_1 = W_1 + \alpha$ * diff * $F_1$<br>$W_1 = 10 + 0.5*22*0$<br>= **10** | $F_2 = 1$ (15+0=15, which is divisible by 3)<br><br>$W_2 = W_2 + \alpha$ * diff * $F_2$<br>$W_2 = W_2 + 0.5*22*0$<br>= 0 + 11 = **11**<br><br>------------------------------------------<br>For **part (b),**<br>$Q(9,0) = 10*0 + 11*1 = 11$ (9+0=9 Features 0,1)<br>$Q(9,1) = 10*1 + 11*0 = 10$ (9+1=10 Features 1,0)<br>So the best (greedy) action is 0. |

(b) (1 point) Given the latest set of weights, at state 9, which action should a purely greedy (i.e. all exploitation, no exploration) agent take? (Circle one)

Action 0 Action 1 Tie Insufficient Information to Answer

**7) Hidden Markov Models**

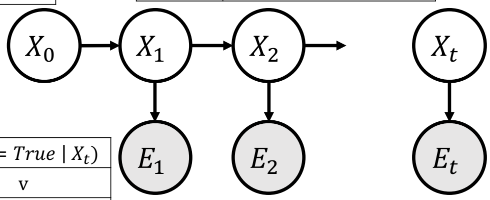
Consider a hidden Markov model with the following parameters. Let $X_{0:t}$ denote the set of $t+1$ hidden random variables and $E_{1:t}$ denote the set of $t$ observed random variables. Let $\pi$ denote the prior over the initial hidden state, $T$ denote the transition probabilities, and $O$ denote the emission probabilities. Let $e_{1:t}$ represent the values observed for the random variable $E_{1:t}$ for any $t$. Let $x_{0:t}$ represent an assignment to random variables $X_{0:t}$. Note, all random variables are Boolean and can take either value True or False.

$\pi =$

| $P(X_0 = True)$ |
|---|
| $u$ |

$T =$

| $X_{t-1}$ | $P(X_t = True \mid X_{t-1})$ |
|---|---|
| $True$ | r |
| $False$ | s |




O =

| $X_t$ | $P(E_t = True \mid X_t)$ |
|---|---|
| $True$ | v |
| $False$ | w |

a) (2 points) Given the observations $e_{1:t}$, complete the equation for the filtering problem $B(X_t)$ by filling in the blanks. Hint, the filtering problem estimates the belief of the current state $X_t$ conditioned on all the observations so far $e_{1:t}$.

$$B(X_t) = P(\, X_t \mid e_{1:t} \,)$$

b) (2 points) Given the observations $e_{1:t}$, complete the equation for the most probable explanation (MPE) problem for the remaining random variables $X_{0:t}$ by filling in the blanks in terms of the hidden and observed random variables. (Recall, the most probable explanation problem finds the optimal assignment for the for all the unobserved variables, $X_{0:t}$, conditioned on all the observed variables, $e_{1:t}$).

$$MPE(X_{0:t} \mid e_{1:t}) = \underset{x_{0:t}}{\operatorname{argmax}}\, P(X_{0:t} \mid e_{1:t})$$

c) (5 points) Compute the probability of the following event in terms of parameters of the given hidden Markov model **(show your work)**.

$$\begin{aligned} P(X_1 = True \mid E_1 = True) &\propto P(E1 = True \mid X_1 = True)\, P(X1 = True) \\ &= v \sum_{x_0} P(X_1 = True, x_0) \\ &= v \sum_{x_0} P(x_0) P(X_1 = True \mid x_0) \\ &= v\,(u\,.\,r + (1-u)\,.\,s) \end{aligned}$$

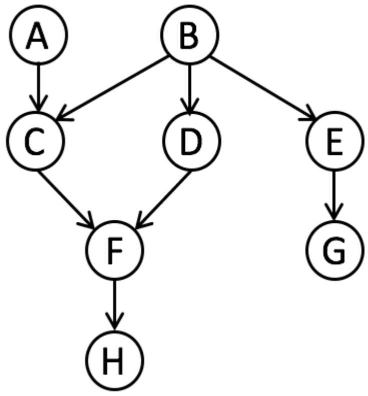
Similarly we can show,

$$P(X_1 = False \mid E_1 = True) \propto w\,(u.(1-r) + (1-u).(1-s))$$

After normalizing,

$$P(X_1 = True \mid E_1 = True) = v \frac{(u\,.\,r + (1-u)\,.\,s)}{v\,(u\,.\,r + (1-u)\,.\,s) + w\,(u.(1-r) + (1-u).(1-s))}$$

**8) Bayesian Network**. Consider the Bayes net shown to the right; determine whether the following conditional independence relations hold by circling "is" or "is not" guaranteed.




Example: $B \perp\!\!\!\perp G$ is / is not guaranteed

Two points for each correct answer to a question; one for leaving a question completely blank.

a) $A \perp\!\!\!\perp H$ is / is not guaranteed

b) $A \perp\!\!\!\perp H \mid F$ is / is not guaranteed

c) $A \perp\!\!\!\perp B$ is / is not guaranteed

d) $A \perp\!\!\!\perp B \mid D$ is / is not guaranteed

e) $A \perp\!\!\!\perp B \mid F$ is / is not guaranteed

f) $A \perp\!\!\!\perp B \mid D,H$ is / is not guaranteed

g) $A \perp\!\!\!\perp B \mid G$ is / is not guaranteed

h) $C \perp\!\!\!\perp D$ is / is not guaranteed

i) $C \perp\!\!\!\perp D \mid A,B$ is / is not guaranteed

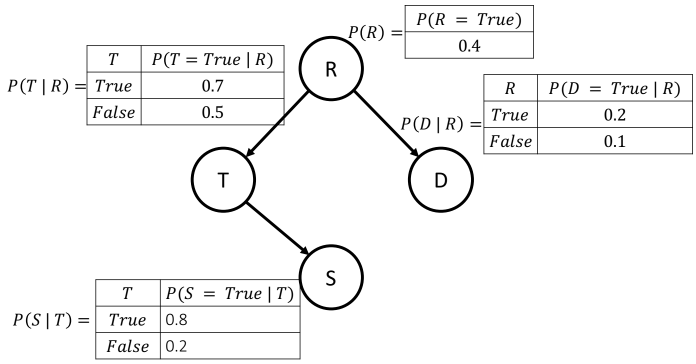
j) $F \perp\!\!\!\perp G \mid D$ is / is not guaranteed

k) $C \perp\!\!\!\perp G \mid B$ is / is not guaranteed

l) $A \perp\!\!\!\perp G$ is / is not guaranteed

**9) Bayesian networks (Redux)**

For the given Bayesian network, answer the following questions:




a) (2 points) Complete the equation for the joint distribution in terms of the CPTs given with the given Bayesian network. (Your answer should be a symbolic formula – don't substitute numbers)

$$P(R,T,D,S) = P(R)P(T\,|R)\;P(D|R)P(S|T)$$

b) (6 points) Use variable elimination compute the probability of the following event **(show your work).** Note, all random variables are binary and can take either value True or False.

P(R=True, S=True) =

$$P(R,S) = \sum_{T}\sum_{D} P(R,T,D,S)$$

$$= \sum_{T}\sum_{D} P(R)P(T\,|R)\;P(D|R)P(S|T)$$

$$= P(R)\sum_{T} P(T|R)P(S|T)$$

When $R = True$ and $S = True$, we get

$$P(R = True, S = True) = \;0.4 \times (0.7 \times 0.8 + 0.3 \times 0.2) = \;0.248$$

c) (2 points) Compute the probability of the following event (show your work).

*P(S=True | R=True) =*

Using Bayes rule we get,

$$= \frac{P(S = True, R = True)}{P(R = True)}$$

$$= \frac{0.248}{0.4}$$

**10 Learning (4 points)**

Suppose that your boss tells you that she thinks the Bayes net from the previous problem had errors in its CPTs. Specifically, she asks you to use machine learning to compute a new CPT for P(D | R) from the data in the table to the right. Because there isn't much data, she suggests you use Laplace smoothing. After learning, what is the new CPT for P(D | R)? Show your work.

| R | D |
|---|---|
| T | T |
| F | F |
| T | T |
| T | F |
| T | T |

| *R* | *P(D = True \| R)* |
|---|---|
| *True* | (3+1) / (4+2) = 4/6 = 0.67 |
| *False* | (0+1) / (1+2) = 1/3 = 0.33 |