Introduction to Artificial Intelligence
CSE 473, Autumn 2017

Dieter Fox

Final Exam

December 12, 2017

Directions

This exam has 7 problems with 111 points shown in the table below, and you have 110 minutes to complete it.

- The exam is closed book. No calculators.
- If you have trouble with a question, by all means move on to the next problem—or the next part of the same problem.
- In answers incurring numbers, feel free to not resolve fractions and sums.
- If a question is unclear, feel free to ask me for clarification!
- **Please do not turn the page until I indicate that it is time to begin.**

Answer Key

Name: ______________________________

Number: ______________________________

1) True/False	/24
4) Probabilities	/10
3) Hidden Markov Models	/20
2) Reinforcement Learning	/25
5) Bayes Nets and Variable Elimination	/20
6) Bayes Nets and CSPs	/8
7) Bayes Nets Sampling	/4
Total	/111

1. (24 points: 2 pts each, -2 pts if wrong) **True or False** - Circle the correct answer (feel to not circle if you're not certain).

   (a) Assume we are running A* graph search with a consistent heuristic h. Assume the optimal cost path to reach a goal has a cost c^*. Then we have that all nodes n reachable from the start state satisfying $f(n) = g(n) + h(n) < c^*$ will be expanded during the search. . . . . T F

   True

   (b) In A* search, if $f_1(n)$ and $f_2(n)$ are both admissible heuristics, then $min(f_1(n), f_2(n))$ is also an admissible heuristic. . . . . T F

   True

   (c) If the only difference between two MDPs is the value of the discount factor then they must have the same optimal policy. . . . . T F

   False

   (d) Approximate Q-Learning with feature vectors will always converge to the optimal policy. . . . . T F

   False

   (e) For an infinite horizon MDP with a finite number of states and actions and with a discount factor γ, where $0 < \gamma < 1$, value iteration is guaranteed to converge. T F

   True

   (f) In an MDP, a discount factor $\gamma < 1$ can always be expressed as a negative living reward. . . . . T F

   False

   (g) In an MDP, a negative living reward can always be expressed using a discount factor $\gamma < 1$. . . . . T F

   False

   (h) A and B are random variables. Given no independence assumptions, it is guaranteed that $\sum_a \sum_b P(A = a|B = b) = 1$. . . . . T F

   False

   (i) The transition probabilities in a Markov Model can change over time. . . . . T F

   False

   (j) In a Hidden Markov Model, the current observation is independent of all else given the current state. . . . . T F

   True

(k) By using the most-constrained variable heuristic and the least-constraining value heuristic we can solve every CSP in time linear in the number of variables. .... T F

False

(l) When enforcing arc consistency in a CSP, the set of values which remain when the algorithm terminates does not depend on the order in which arcs are processed from the queue. .... T F

True

2. (10 points) **Probabilities**

   (a) (4 points) For each expression, state whether or not it is equal to $\mathbf{P(A, B, C)}$ given **no independence assumptions**.

      i. $P(C \mid A, B) \cdot P(A, B)$ True

      ii. $P(C \mid A, B) \cdot P(A) \cdot P(B)$ False

      iii. $P(A \mid B) \cdot P(B \mid C) \cdot P(C)$ False

      iv. $P(A \mid B, C) \cdot P(B \mid C) \cdot P(C)$ True

   (b) (4 points) Suppose A and B are independent boolean variables. Determine the missing entries x and y in the joint distribution of $P(A, B)$ shown below. **You do not have to simplify numerical results.**

$$P(A = T, B = T) = 0.15$$
$$P(A = T, B = F) = 0.45$$
$$P(A = F, B = T) = x$$
$$P(A = F, B = F) = y$$

$x =$ ______________0.1 $y =$ ______________0.3

One possible way to solve:

$$\begin{aligned}
P(A=T) &= 0.6 \\
P(B=T) &= 0.15 + x \\
P(A=T)P(B=T) &= P(A=T, B=T) && \text{(due to independence)} \\
0.6(0.15 + x) &= 0.15 \\
x &= \frac{0.15}{0.6} - 0.15 \\
y &= 0.4 - x && \text{(since it all sums to 1)} \\
y &= 0.4 - (\frac{0.15}{0.6} - 0.15)
\end{aligned}$$

(c) (2 points) Suppose we have a robot on a grid world with a noisy sensor and we know the distribution for $P(Observation|Location)$ of sensor readings given the robot's grid square. We also have knowledge about the locations in the form of a prior, $P(Location)$. However, we are really interested in the probability of being in a grid square given the sensor reading, $P(Location|Observation)$. How could you compute this probability? Short answer.

Using Bayes' theorem, we know that $P(Location|Observation) = P(Observation|Location) * P(Location)/P(Observation)$.

3. (20 points) **Hidden Markov Models**

Suppose there is a decision problem called the Exam Problem, which is deciding if you will pass an exam given a set of studying intervals. Professor Kenny is attempting to classify the problem as either NP-Complete or just NP-Hard. We will model Kenny's classifications over time as a Hidden Markov Model.

Let c be the state in which Kenny classifies the Exam Problem as NP-Complete, and h be the state in which Kenny classifies the Exam Problem as NP-Hard.

Rather than using legitimate techniques, Kenny classifies the problem based on how he classified it the day before (the previous time step). Between each day, there is a 70% chance that Kenny changes his classification of the Exam Problem, otherwise his classification stays the same as the previous day.

(a) (2 points) Fill in the table with the transition probabilities as defined above, where the row corresponds to X_{t+1} and the column to X_t.

	c	h
c	$P(X_{t+1} = c \mid X_t = c) = 0.3$	$P(X_{t+1} = c \mid X_t = h) = 0.7$
h	$P(X_{t+1} = h \mid X_t = c) = 0.7$	$P(X_{t+1} = h \mid X_t = h) = 0.3$

(b) (2 points) Now suppose there is evidence in the form of Kenny listening to music. If Kenny classifies the Exam Problem as NP-Complete then there is a 60% chance he is listening to music. If Kenny classifies the Exam Problem as NP-Hard then there is an 80% chance he is listening to music. Denote listening to music as m and not listening to music as $\neg m$.

Fill in the table with the conditional probabilities of the evidence based on the description above.

	c	h
m	$P(m \mid c) = 0.6$	$P(m \mid h) = 0.8$
$\neg m$	$P(\neg m \mid c) = 0.4$	$P(\neg m \mid h) = 0.2$

(c) (2 points) There's a 50% chance Kenny's initial classification is NP-Complete, $P(X_0 = c) = 0.5$. Compute the values for the passage of time for c and h, at $t = 1$.

passage of time for c: $0.3 \cdot 0.5 + 0.7 \cdot 0.5 = 0.5$

passage of time for h: $0.7 \cdot 0.5 + 0.3 \cdot 0.5 = 0.5$

(d) (4 points) Now factor in the evidence that Kenny is listening to music at $t = 1$. Compute the beliefs $B(X_1 = c)$ and $B(X_1 = h)$. You do not have to evaluate/simplify terms when computing the beliefs.

evidence for c: $0.6 \cdot 0.5 = 0.3$

evidence for h: $0.8 \cdot 0.5 = 0.4$

$B(X_1 = c) = \frac{0.3}{0.3+0.4} = 0.429$

$B(X_1 = h) = \frac{0.4}{0.3+0.4} = 0.571$

(e) (2 points) What is the probability of producing the exact sequence C, H, H, C, H (from $t = 0$ to $t = 4$)? Ignore evidence. You do not need to evaluate or simplify your expression. Recall that there's a 50% chance Kenny's initial classification is NP-Complete, $P(X_0 = c) = 0.5$.

$0.5 \cdot 0.7 \cdot 0.3 \cdot 0.7 \cdot 0.7$

(f) (2 points) How would you change the transition probabilities so that Kenny never changes his classification in between days? Provide the corresponding values.

$P(X_{t+1} = c|X_t = c) = 1.0$
$P(X_{t+1} = h|X_t = c) = 0$
$P(X_{t+1} = c|X_t = h) = 0$
$P(X_{t+1} = h|X_t = h) = 1.0$

(g) (2 points) What is the difference between a Markov Model and a Hidden Markov Model? Provide a short answer.

Observe output/effects/evidence/emissions/observations at each time step for Hidden Markov Models but not for Markov Models (other answers acceptable if accurate).

(h) (2 points) Consider a sequence of evidence $e_1, \dots, e_n$ in a Hidden Markov Model. We want to perform inference on X_t. What is the difference between filtering and smoothing in the context of computing a distribution over, or performing inference on, X_t with the given evidence? Provide a short answer.

Filtering is inference based on previous evidence,$P(X_t|e_{1:t})$, and smoothing is inference based on all evidence, $P(X_t|e_{1:n})$.

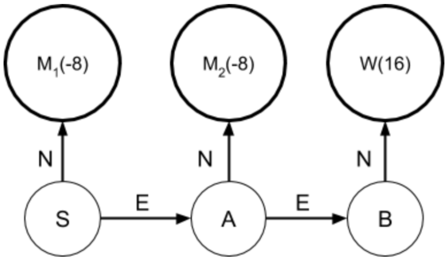
(i) (2 points) How would you find the most probable explanation in a Hidden Markov Model? Provide a short answer.

Use Viterbi's algorithm to compute the values, extract path from values.

4. (25 points) **MDP and Reinforcement Learning**

You are living in the year 2050 and are given the task to find water on Mars. You remember the material from CSE473 and design a robot that uses the MDP shown below to represent the state space on Mars. Martians live in M_1 and M_2 and water can be found in state W. The states S, A, and B are useless and provide 0 reward. Once your robot reaches one of the M and W states its only option is to send a boolean value back to Earth representing whether or not it found water, and it's not getting any additional reward for doing that.

The robot should stay away from the Martians, so the robot receives a reward of -8 when entering an M state, and a reward of 16 when entering the W state (as indicated in the figure). The only actions it can take is to either go East (E) or go North (N) (it can only go North from state B). You are the best at designing robots and hence, there is no noise when transitioning to different states and actions always succeed.



(a) (3 points) What are the optimal values, V^* of each state if $\gamma = 0.5$?

$V^*(S) = 4$ or $V_1(S) = 0$
$V^*(A) = 8$ or $V_1(A) = 0$
$V^*(B) = 16$ or $V_1(B) = 16$

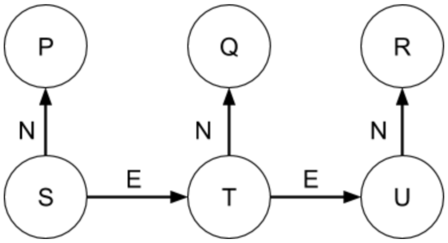
(b) (2 points) What are the Q-values for the state S if $\gamma = 0.5$?

$Q^*(S, N) = -8$, $Q_1(S, N) = -8$
$Q^*(S, E) = 4$, $Q_1(S, E) = 0$

(c) (3 points) What is the optimal policy, π^*, for this MDP?

$\pi^*(S) =$ E
$\pi^*(A) =$ E
$\pi^*(B) =$ N

After achieving great success you are supposed to present your exploration to your boss. However, while walking to the presentation room a Martian attacks the robot and puts it in a different location on Mars. You know this new location has a similar state space as the one before, though the rewards at “Done” states P, Q, and R are different and unknown. Your agent now has a stochastic transition function due to the damage. You remember that you had some code for online learning of values as backup. Help yourself learn these values before the final presentation.



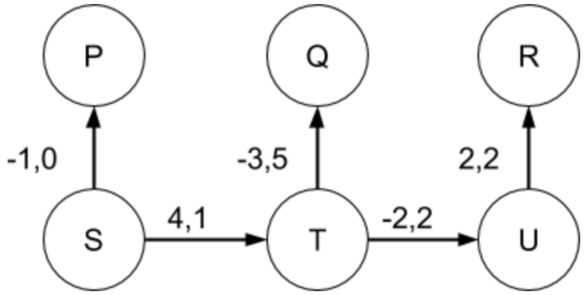
(d) (6 points) Consider the following episodes performed in this state space. The experience tuples are of the form [s, a, s’, r], where the robot starts in state s, performs action a, ends up in state s’, and receives immediate reward r, which is determined by the state entered. Let $\gamma = 1.0$ for this MDP. Fill in the values computed by the Q-learning algorithm with a learning rate of $= 0.5$. All Q values are initially 0, and you should fill out each row using values you have computed in previous rows.

	Records	Q-Values
1	[S, E, T, 4]	Q(S, E) = 2
2	[T, E, U, 4]	Q(T, E) = 2
3	[U, N, R, 16]	Q(U, N) = 8
4	[S, E, P, -4]	Q(S, E) = -1
5	[T, E, Q, -4]	Q(T, E) = -1
6	[S, E, T, 4]	Q(S, E) = $3/2$

(e) (3 points) You also want to know the transition probabilities for a few state - action pairs. Use the information from the records above to calculate T(s, a, s’) for the given state-action pairs where we start in state s, perform action a and end up in state s’.

T(S, E, T) = $2/3$
T(T, E, U) = $1/2$
T(U, N, R) = 1

In order to impress your boss even more, you also perform approximate q-learning using features, as shown on the MDP below. You find out that 2 features would allow you to represent this MDP accurately enough. However, you are stumped by the calculations needed to compute all the values.



(f) (7 points) Help your future self get started by computing the weights w_1 and w_2 using approximate Q-Learning using the following 2 records. Calculate the weights after each record. Initially $w_1 = 1$ and $w_2 = 0$. Use a learning rate of α=0.5 and a discount factor of γ=1.0. **Please show your work**.

i. [S, E, T], reward: 8

$w_1 = 1 + 0.5 * [8 + 1 * (-2 * 1 + 2 * 0) - (4 * 1 + 1 * 0)] * 4 = 5$
$w_2 = 0 + 0.5 * [8 + 1 * (-2 * 1 + 2 * 0) - (4 * 1 + 1 * 0)] * 1 = 1$

ii. [T, E, U], reward: 12

$w_1 = 5 + 0.5 * [12 + 1 * (2 * 5 + 2 * 1) - (-2 * 5 + 2 * 1)] * -2 =$ -27
$w_2 = 1 + 0.5 * [12 + 1 * (2 * 5 + 2 * 1) - (-2 * 5 + 2 * 1)] * 2 = 33$

(g) (1 points) Give one possible issue when using approximate Q-Learning to compute Q-values for a MDP (1 sentence).

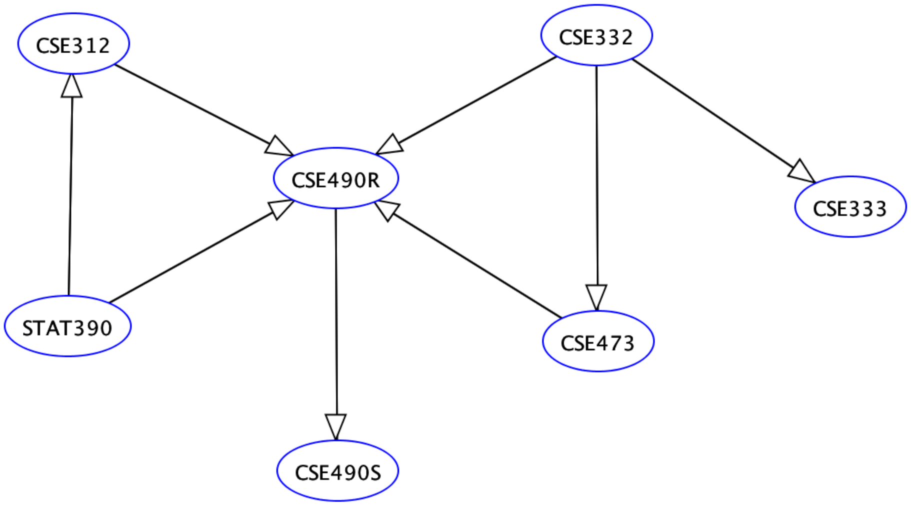
possible problem when states may share features but might be really different in values.

5. (20 points) **Bayesian Networks and Variable Elimination**

Consider the Bayesian Network of UW Courses which students might take at a given quarter. Each variable can be either "Registered" (True) or "Not Registered" (False).

(a) (8 points) **Bayesian Network Independence**

Determine if the following conditional independence statements are enforced by the structure of the Bayes' Net. (2 points each, -1 point if wrong)



i. Is 332 conditionally independent of 312 given 490S? .... T F

ii. Is 312 conditionally independent of 333 given 490R and 332? .... T F

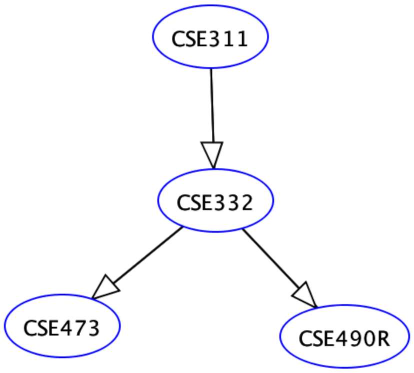
iii. Is 333 conditionally independent of 473 given 490R and 332? .... T F

iv. Is 332 conditionally independent of STAT390 given 312? .... T F

A. False
B. True
C. True
D. True

(b) (2 points) **Bayesian Network Inference**

Consider the following subset of the above network.



Complete the Conditional Probability Tables below with the **names of the distributions** that fully describe the given Bayesian Network. Write one value in each box marked with a ⇒.

⇒ CSE311	⇒ P(CSE311)
$True$	0.5
$False$	0.5

$CSE311$	$CSE332$	⇒ $P(CSE332 \mid CSE311)$
$True$	$True$	0.5
$True$	$False$	0.5
$False$	$True$	0.0
$False$	$False$	1.0

$CSE332$	$CSE473$	⇒ $P(CSE473 \mid CSE332)$
$True$	$True$	0.5
$True$	$False$	0.5
$False$	$True$	0.0
$False$	$False$	1.0

$CSE332$	$CSE490R$	⇒ $P(CSE490R \mid CSE332)$
$True$	$True$	0.5
$True$	$False$	0.5
$False$	$True$	0.0
$False$	$False$	1.0

(c) (10 points) **Variable Elimination**

Using Variable Elimination, list the steps to calculate $P(CSE332|\neg cse473)$. Clearly mark any new factors and how they are introduced. Eliminate variables in any choice of order. Please use the following notation to sum over the values of a variable:

For example, $X : \sum_x P(X = x)$.

Define new factors as $f_2(X, y) = \sum_z f_1(X, y, Z = z)P(X|Z = z)$.

Variable Elimination Strategy:

i. List factors: $P(CSE311)$, $P(CSE332|CSE311)$, $P(CSE473|CSE332)$, $P(CSE490R|CSE332)$

ii. Remove irrelevant factors: $P(CSE490R|CSE332)$

iii. Observe $\neg cse473$

iv. Eliminate $P(CSE311)$: $f_1(CSE332)$, $P(\neg cse473|CSE332)$

$$f_1(CSE332) = \textstyle\sum_{cse311} P(cse311) * P(CSE332|cse311)$$

v. Join $f_1(CSE332)$ and $P(\neg cse473|CSE332)$

$$f_2(CSE332, \neg cse473) = f_1(CSE332) * P(\neg cse473|CSE332)$$

vi. Normalize for $CSE332$

$$P(CSE332|\neg cse473) = \frac{f_2(CSE332, \neg cse473)}{f_2(CSE332, cse473) + f_2(CSE332, \neg cse473)}$$

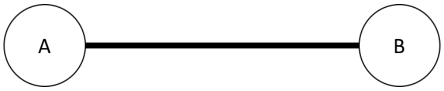
6. (8 points) **Bayes Nets and CSPs**

Let us consider how to solve CSPs by converting them to Bayesian networks.

We can solve for values of variables that satisfy CSPs by running inference for the query on the constraints, where the setting(s) of the variables with nonzero probability are those that satisfy the constraints.

College Level Mathematics

Let us analyze a CSP of only two integer variables A and B each with domain $\{+1, -1\}$. **The two variables are subject to the constraint that** $A + B \neq 0$. This is shown below with the edge between the two nodes representing the constraint.

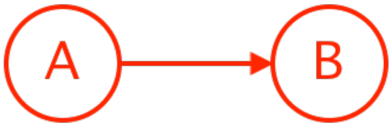


(a) (4 points) Enforce the above constraint by representing the variables in a Bayesian Network. This formation must represent each variable as a node with a **directed** edge showing influence between the two variables. **Draw the Bayes' net** and **define all necessary conditional probability tables** with values such that consistent settings of the variables correspond to nonzero probabilities.

P(A) has been provided to get you started.

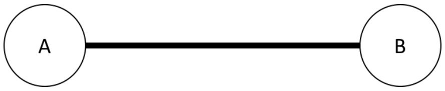
A	$P(A)$
$+1$	$1/2$
-1	$1/2$

Solution Bayesian Network:



A	B	$P(B\|A)$
$+1$	$+1$	1.0
-1	$+1$	0.0
$+1$	-1	0.0
-1	-1	1.0

(b) (4 points) The same CSP is printed for reference with variables $A, B \in \{+1, -1\}$ and the constraint that $A + B \neq 0$:

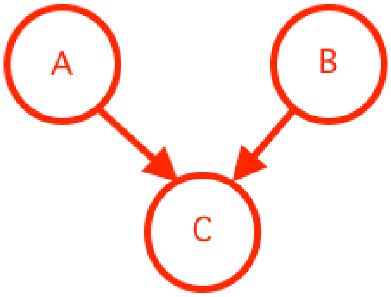


Once again, we will convert this CSP into a Bayes' net. However, this time we want to represent the constraint by adding a third Boolean variable C whose conditional distribution enforces the constraint. **Draw the Bayes' Net with the nodes A, B, C** to show the influence between the three nodes, and **provide all conditional probability tables** with values such that consistent settings of the variables correspond to nonzero probabilities.

As before, P(A) is provided.

A	$P(A)$
$+1$	$1/2$
-1	$1/2$

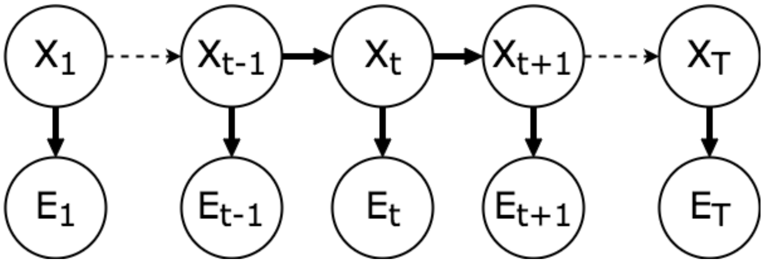
Solution Bayesian Network:



B	$P(B)$
$+1$	$1/2$
-1	$1/2$

A	B	C	$P(C\|A,B)$
$+1$	$+1$	T	1.0
$+1$	$+1$	F	0.0
$+1$	-1	T	0.0
$+1$	-1	F	1.0
-1	$+1$	T	0.0
-1	$+1$	F	1.0
-1	-1	T	1.0
-1	-1	F	0.0

7. (4 Points) **Bayesian Network Sampling** Consider the HMM shown below. We want to use Gibbs sampling to generate samples from the distribution over the hidden states X_i conditioned on all T evidence values.



In Gibbs Sampling, we pick one of the variables, say X_t, and sample a new value for it, conditioned on all the other sampled values. How can you efficiently compute this distribution using the sampled values and the conditional distributions specifying the HMM? Provide the equation and any additional operations necessary.

Solution: $p(X_t|x_{t-1})p(e_t|X_t)p(x_{t+1}|X_t)$ followed by normalization over X_t

Equation sheet:

$$
\begin{aligned}
V^*(s) &= \max_a Q^*(s,a) && \text{(Bellman equations)}\\
Q^*(s,a) &= \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] &&\\
V^*(s) &= \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] &&\\
\pi^*(s) &= \operatorname*{argmax}_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] && \text{(Optimal policy)}\\
Q(s,a) &= (1-\alpha)Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a')) && \text{(Q-Learning)}\\
Q(s,a) &= w_1 f_1(s,a) + w_2 f_2(s,a) + ... + w_n f_n(s,a) && \text{(Approximate Q-Learning)}\\
\text{diff} &= [r + \gamma \max_{a'} Q(s',a')] - Q(s,a) &&\\
w_i &= w_i + \alpha[\text{diff}] f_i(s,a) &&
\end{aligned}
$$

Additional space

Additional space