CS 430: Introduction to Artificial Intelligence
Fall 2006
Final Exam

Name: ____

This final exam consists of 16 pages and a total of 110 marks. The exam is open book and open notes. You have 110 minutes to complete this exam. If you are taking too much time on one question, please move on to other questions.

Section	**Marks**
Pre-midterm questions	**/ 20**
Propositional Logic	**/ 20**
Bayesian Networks (Parameter Learning)	**/ 10**
Bayesian Networks (Structure Creation)	**/ 10**
Bayesian Networks (Inference)	**/ 20**
MDPs and Reinforcement Learning	**/ 30**
Total	**/ 110**

Section I: Pre-midterm questions [20 points]

1. Answer the following questions using true or false. [1 point each]

a) Recall the different characteristics of environments (ie. fully vs partially observable, deterministic vs stochastic, episodic vs sequential, static vs dynamic, discrete vs continuous, and single vs multi-agent). The most difficult type of environment for an agent is a fully observable, deterministic, episodic, static, discrete and single-agent environment.
False

b) In a deterministic environment, if you are in state *S* and you perform action *A*, the probability of reaching the next state *S'* is 1.
True

c) Rationality depends on the following 4 things: a performance measure, the agent's prior knowledge of the environment, the actions the agent can perform, and the agent's current percept.
False

d) A simple-reflex agent will work well in a partially observable environment.
False

e) Local search techniques such as hillclimbing, simulated annealing, and genetic algorithms not only find the optimal state but also return the path to that state.
False

f) A program implementing breadth-first search can be easily converted to depth-first search just by changing the fringe implementation from a FIFO queue to a LIFO queue.
True

g) Suppose you have a heuristic function h*(n) which tells you the actual cost of the cheapest path from the node n to the goal. Then h*(n) dominates all other admissible heuristics for the search problem.
True

h) A-star search has a space complexity that is linear in the number of states.
False

i) Depth first search is optimal.
False

j) Both simulated annealing and genetic algorithms are a pain to implement because they require a lot of parameter-tweaking in order to get them to work on a particular problem.
True

k) Evaluation functions are only useful for minimax or alpha-beta if we are unable to keep the entire game tree in memory.
True

l) The entire game tree for chess can be stored in memory for a standard desktop machine with 4 GB of RAM.
False

m) It is perfectly rational to play a dominant strategy.
True

n) If Player 1's payoff in a zero-sum, two-player game is +100, then the payoff for Player 2 is –100.
True

o) If a normal-form game has a single Nash equilibrium and you tell your opponent that you are playing your part of the Nash Equilibrium strategy, then your opponent (assuming he/she is rational) has no choice but to play his/her part of the Nash Equilibrium strategy.
True

p) In the Tragedy of the Commons, the farmers would make more money if they acted in their own self-interest and did not cooperate with each other when deciding on the number of goats to graze.
True

2. Circle the pure strategy Nash equilibrium in the normal-form game below or write down “None” if one doesn’t exist. [2 points]

	Player B: S1	Player B: S2
Player A: S1	A = 1, B = 3	A = -2, B = 2
Player A: S2	A = -2, B = 2	A = 3, B = -1

Nash Equilibrium at (A: S1, B: S1)

3. Write out the gradient descent rule for the function $f(x) = 4x^5 + 2x^3 + 2$, assuming the learning rate $\alpha = 0.1$. [2 points]

$x \leftarrow x - (0.1)(20x^4 + 6x^2)$

Section II: Propositional Logic [20 points]

1. You would like to order a pizza for an extremely picky friend. Fortunately, you’ve stored all of her preferences in a knowledge base and you can use that knowledge base to infer if she will be happy with the pizza you just ordered for her.

If your pizza has pepperoni, your friend will be happy. If the pizza has anchovies and feta cheese, your friend will get sick. If your friend gets sick, she will not be happy. If you order olives, the pizzeria will automatically put olives and feta cheese on the pizza. You order a pizza with anchovies and olives.

a) The paragraph above has been converted into propositional logic symbols below. However, the knowledge base below is not in CNF. Convert it to CNF. [10 points]

$P \Rightarrow H$
$(A \wedge FC) \Rightarrow S$
$S \Rightarrow \neg H$
$O \Rightarrow FC \wedge O$
$A \wedge O$

CNF notation:
$\neg P \vee H$

$\neg A \vee \neg FC \vee V$
$\neg V \vee \neg H$
$\neg O \vee FC$
A
O

b) Based on the knowledge base above, does it entail that your friend will **NOT** be happy with the pizza you ordered? Show all your resolution steps or you won't get full credit. [5 points]

$\neg O \vee FC$, O

FC

FC, $\neg A \vee \neg FC \vee V$

$\neg A \vee V$

$\neg A \vee V$, A

V

V, $\neg V \vee \neg H$

$\neg H$

$\neg H$, H

{}

3. If $\alpha \models \beta$ and $\beta \models \gamma$, does $\alpha \models \gamma$? Explain your answer. [5 points]
Yes $\alpha \models \gamma$.
By the definition of entailment, if $\alpha \models \beta$, then in every model in which α is true, β is also true. Since $\beta \models \gamma$, then every model in which β is true, γ is also true. It is easy to see that in every model that α is true, γ is also true.

Section III: Bayesian Networks (Naïve Bayes) [10 points]

In this question, you will be answering questions regarding implementing a naïve Bayes classifier that predicts a person's favorite team based on their favorite color and home town. The training data set is shown below. The FavoriteColor feature only takes two possible values of Green and Orange. The HomeTown feature only takes two possible values of Corvallis and Eugene. Finally, the FavoriteTeam class label only takes two possible values of Ducks and Beavers.

FavoriteColor	HomeTown	FavoriteTeam
Green	Eugene	Ducks
Orange	Corvallis	Beavers
Orange	Corvallis	Beavers
Green	Corvallis	Ducks
Orange	Eugene	Beavers
Green	Eugene	Beavers
Green	Corvallis	Ducks
Green	Eugene	Ducks
Orange	Corvallis	Beavers
Orange	Eugene	Beavers

1. Write down all the probabilities you need to implement a naïve Bayes classifier along with their values based on this training data set. Use the Dirichlet priors we mentioned in class. [8 points]

$$P(Beavers) = \frac{6+1}{10+2} = \frac{7}{12} \qquad P(Ducks) = \frac{4+1}{10+2} = \frac{5}{12}$$

$$P(FavoriteColor = Green \mid FavoriteTeam = Ducks) = \frac{4+1}{4+2} = \frac{5}{6}$$

$$P(FavoriteColor = Orange \mid FavoriteTeam = Ducks) = \frac{0+1}{4+2} = \frac{1}{6}$$

$$P(FavoriteColor = Green \mid FavoriteTeam = Beavers) = \frac{1+1}{6+2} = \frac{2}{8}$$

$$P(FavoriteColor = Orange \mid FavoriteTeam = Beavers) = \frac{5+1}{6+2} = \frac{6}{8}$$

$$P(HomeTown = Corvallis \mid FavoriteTeam = Beavers) = \frac{3+1}{6+2} = \frac{4}{8}$$

$$P(HomeTown = Eugene \mid FavoriteTeam = Beavers) = \frac{3+1}{6+2} = \frac{4}{8}$$

$$P(HomeTown = Corvallis \mid FavoriteTeam = Ducks) = \frac{2+1}{4+2} = \frac{3}{6}$$

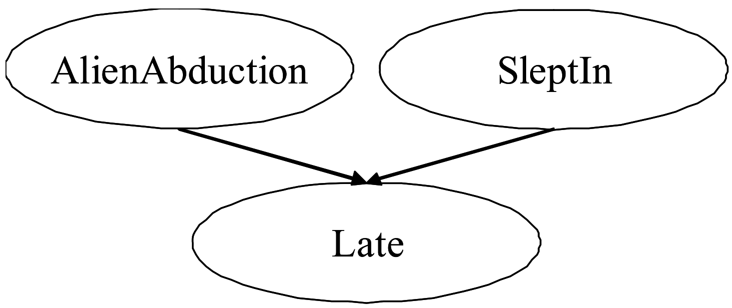
$$P(HomeTown = Eugene \mid FavoriteTeam = Ducks) = \frac{2+1}{4+2} = \frac{3}{6}$$

2. Some of the probabilities above can be derived from the other probabilities. What is the minimum number of the probabilities above that you must store in order to implement a naïve Bayes classifier? [2 points]

5 probabilities

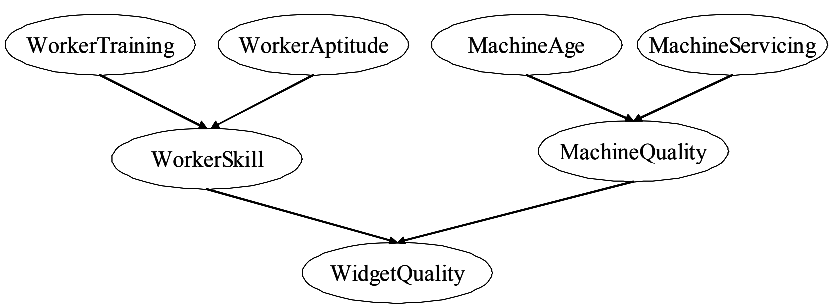
Section IV: Bayesian Networks (Structure creation) [10 points]

1. We would like to model the event that Bob is late for class. If we know that Bob is late because he slept in, then this knowledge would "explain away" the event that Bob was abducted by aliens. Draw the Bayesian network that represents this information using the three nodes of Late, AlienAbduction, and SleptIn. [3 points]



2. You have been asked to build a Bayesian network for a widget manufacturing process. The quality of the widget depends on the skill of the worker and the quality of the machine. The worker's skill is in turn influenced by the training the worker has received and the aptitude of the worker. Knowing the worker's skill makes the worker's training and aptitude independent of the widget quality. Two factors affect the quality of the machine: its age and the frequency of servicing. Knowing the quality of the machine makes the age and the frequency of servicing independent of the quality of the widget.

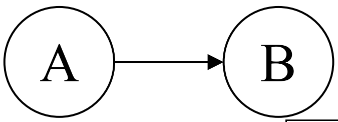
Draw a Bayesian network which captures this manufacturing process. You will use the nodes WidgetQuality, WorkerSkill, WorkerTraining, WorkerAptitude, MachineQuality, MachineAge, MachineServicing. [7 points]



Section V: Bayesian Networks (Inference) [20 points]

1. Calculate P(A = True | B = True) using the Bayesian network structure and conditional probability tables below. Hint: Use Bayes rule.[8 points]

A	B	C	P(C|A,B)
False	False	False	0.75
False	False	True	0.25
False	True	False	0.75
False	True	True	0.25
True	False	False	0.75
True	False	True	0.25
True	True	False	0.75
True	True	True	0.25



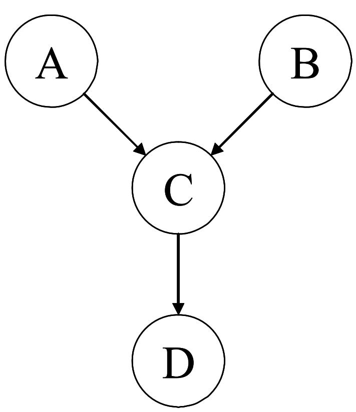
A	P(A)
False	0.2
True	0.8

A	B	P(B|A)
False	False	0.5
False	True	0.5
True	False	0.7
True	True	0.3

$$P(B = true) = \sum_a P(A = a, B = true) = \sum_a P(A = a)P(B = true \mid A = a)$$
$$= P(A = true)P(B = true \mid A = true) + P(A = false)P(B = true \mid A = false)$$
$$= (0.8)(0.3) + (0.2)(0.5)$$
$$= 0.24 + 0.1 = 0.34$$

$$P(A = true \mid B = true) = \frac{P(B = true \mid A = true)P(A = true)}{P(B = true)}$$
$$= \frac{(0.3)(0.8)}{0.34} = \frac{0.24}{0.34} = 0.71$$

2. Use the Bayesian network structure and conditional probability tables below to calculate P(C = false). [8 points]



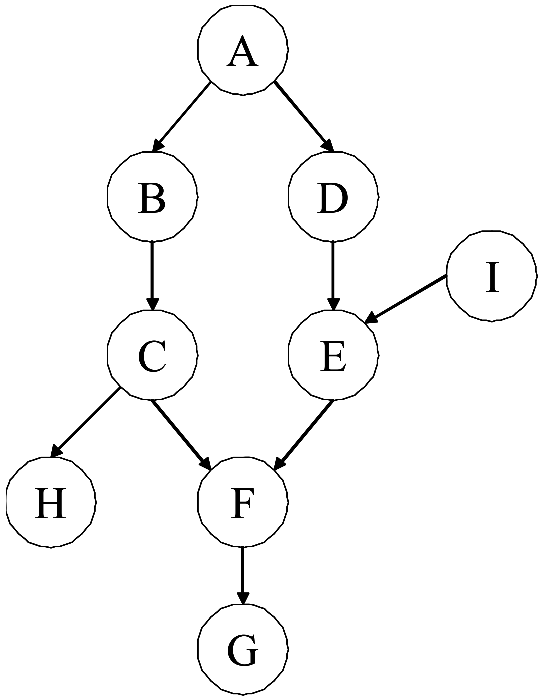
A	P(A)
False	0.6
True	0.4

B	P(B)
False	0.2
True	0.8

C	D	P(D|C)
False	False	0.75
False	True	0.25
True	False	0.9
True	True	0.1

$$P(C = false) = \sum_{a,b,d} P(A = a, B = b, C = false, D = d)$$
$$= \sum_{a}\sum_{b}\sum_{d} P(A = a)P(B = b)P(C = false \mid A = a, B = b)P(D = d \mid C = false)$$
$$= \sum_{a} P(A = a)\sum_{b} P(B = b)P(C = false \mid A = a, B = b)\sum_{d} P(D = d \mid C = false)$$
$$= \sum_{a} P(A = a)\sum_{b} P(B = b)P(C = false \mid A = a, B = b)(1)$$
$$= P(A = false)[P(B = false)P(C = false \mid A = false, B = false) +$$
$$P(B = true)P(C = false \mid A = false, B = true)] +$$
$$P(A = true)[P(B = false)P(C = false \mid A = true, B = false) +$$
$$P(B = true)P(C = false \mid A = true, B = true)]$$
$$= (0.6)[(0.2)(0.75) + (0.8)(0.75)] + (0.4)[(0.2)(0.75) + (0.8)(0.75)]$$
$$= (0.6)[0.15 + 0.6] + (0.4)[0.15 + 0.6]$$
$$= (0.6)(0.75) + (0.4)(0.75)$$
$$= 0.75$$

3. Answer true or false to the conditional independence statements using the graph below. Write out the blocked or non-blocked paths for partial credit.

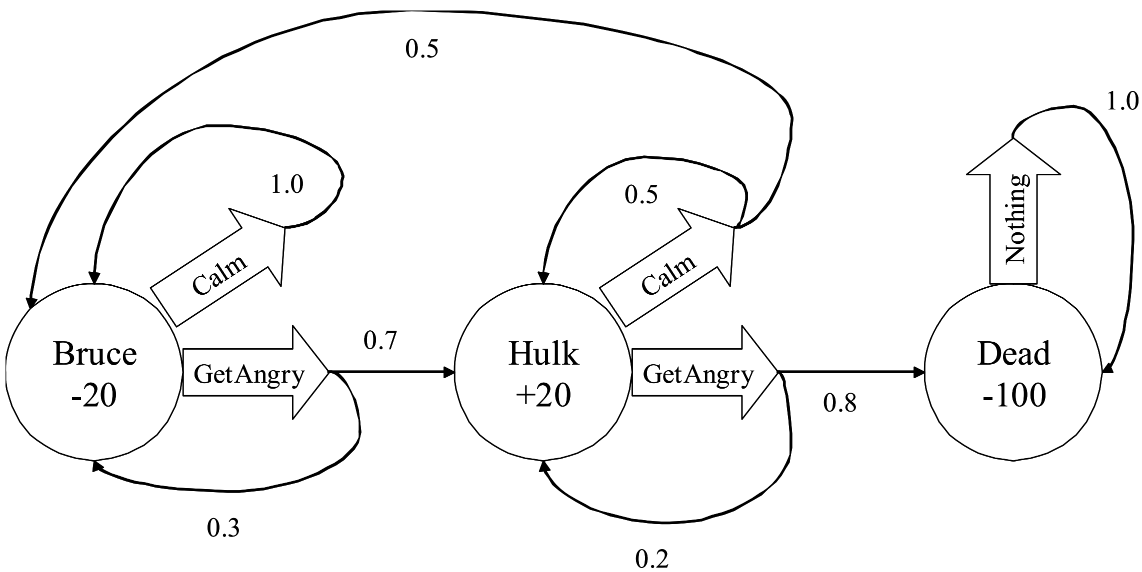


a) I(B, D | A) [2 points]
True
B<-A->D is blocked
B->C->F<-E<-D is blocked

b) I(H, I | E) [2 points]
False
H<-C->F<-E<-I is blocked
H<-C<-B<-A->D->E<-I is not blocked

Section VI: MDPs and Reinforcement Learning [30 points]

1. The Markov Decision Process below represents a model of how Bruce Banner becomes the Incredible Hulk.



a) If we perform value iteration on this MDP, what are the initial values of the utilities of each state ie. what are U_1(Bruce), U_1(Hulk), and U_1(Dead)? [2 points]

U1(Bruce) = -20
U1(Hulk) = 20
U1(Dead) = -100

b) If we perform one round of value iteration on this MDP, what are the values for the utilities of each state ie. what are U_2(Bruce), U_2(Hulk), and U_2(Dead)? Use $\gamma = 0.9$. [6 points]

U2(Bruce) = -20 + (0.9)max{-20,(0.7)(20)+(0.3)(-20)} = -20 + (0.9)max{-20,8}
= -20 + (0.9)(8) = -20 + 7.2 = -12.8

U2(Hulk) = 20 + (0.9)max{(0.5)(20)+(0.5)(-20),(0.2)(20)+(0.8)(-100)}
= 20 + (0.9)max{0,-76} = 20

U2(Dead) = -100 + (0.9)max{(1.0)(-100)} = -100 + (0.9)(-100) = -190

c) Now suppose we do policy iteration using an initial policy of π_1(Bruce) = Calm, π_1(Hulk) = Calm and π_1(Dead) = Nothing. Do one round of the Policy Evaluation phase and calculate the utilities of each state. Use $\gamma = 0.9$. [8 points]

U1(Bruce) = -20 + (0.9){(1.0)U1(Bruce)}
=> U1(Bruce) = -20 + (0.9)U1(Bruce)
=>(0.1)U1(Bruce) = -20
=>U1(Bruce) = -200

U1(Hulk) = 20 + (0.9){(0.5)U1(Bruce)+(0.5)U1(Hulk)}
=> U1(Hulk) = 20 + (0.45)U1(Bruce) + (0.45)U1(Hulk)
=>(0.55)U1(Hulk) – (0.45)U1(Bruce) = 20
=>(0.55)U1(Hulk) = 20 + (0.45)U1(Bruce)
=>U1(Hulk) = [20 + (0.45)(-200)]/(0.55) = -127.27

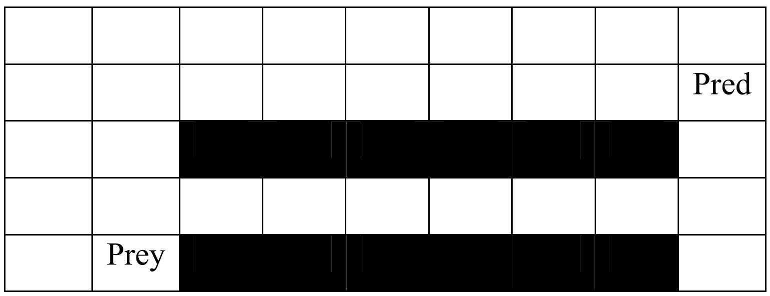
U1(Dead) = -100 + (0.9){(1.0)U1(Dead)
=> U1(Dead) = -100 + (0.9)U1(Dead)
=> (0.1)U1(Dead) = -100
=>U1(Dead) = -1000

2. Explain the difference between the reward for a state and the utility of a state. [2 points]
The reward is simply the goodness of being in state s.
The utility is the long-term total expected reward for the sequence of future states starting at s.

3. In a reinforcement learning algorithm with an ε-greedy (ie. epsilon greedy) exploration scheme, what happens when we make $\varepsilon = 0$? What happens when we make $\varepsilon = 1$? [2 points]
If epsilon is 0, you will always pick the optimal action. There is no exploration. If epsilon is 1, you will always pick randomly. There is no exploitation.

4. The grid world below is the setting for a predator-prey game. At the beginning, the predator is placed at the location marked “Pred” while the prey is placed at the location marked “Prey”. The goal of the predator is to catch the prey by moving into the same square as the prey. The predator can move but the prey remains stationary. The black squares correspond to pits. If the predator falls into a pit, then it will suffer a large negative reward.

We will make our agent the predator and we will learn the optimal policy for the predator using reinforcement learning. In order to formulate the predator-prey game as a reinforcement learning problem, we will first describe the action space. The agent can perform actions of N, E, S, and W in this world. However, these actions are stochastic, meaning with some probability, the agent will succeed and with some probability it will move in a random unintended direction.



a) Define the state space in this problem. How many possible states are there? (To help you, there are 12 black squares). [2 points]

33 states – it is simply the location of the predator

b) Define a reward function for this game, keeping in mind that the predator should capture the prey as quickly as possible. Remember that the pits are undesirable for the predator, so it should avoid them. [2 points]
Here is a sample reward function:
-100 if you fall into a pit
+100 if you catch the prey
-0.1 for all other squares

c) The state with the highest q-value will be the state where the predator catches the prey. What state-action pairs should have the next highest q-values? You can describe this with sentences. [2 points]
The state corresponding to:
- the predator to the square directly north of the prey and the action to go south
- the predator to the square directly west of the prey and the action to go east

d) Without changing the transition model or the reward function, how else can we "encourage" the predator to take the risky path between the pits to catch the prey? [2 points]

Decrease the discount factor. This makes rewards in the future less desirable, so the predator has the incentive to get rewards faster.

e) Suppose the prey could move and the game ends when the predator occupies the same square as the prey. How would you formulate the state space now and how many possible states are there? [2 points]

There are now 33 * 33 states – location of predator and the location of the prey

The End