Final Examination

CS 540: Introduction to Artificial Intelligence

August 6, 2004

LAST (FAMILY) NAME: ________SOLUTION_______________________

FIRST NAME: ____

Problem	**Score**	**Max Score**
1	———–	15
2	———–	20
3	———–	15
4	———–	10
5	———–	20
6	———–	20
Total	———–	100

1. [15] **Neural Networks**

(a) [4] The NAND function of n binary inputs, $x_1, x_2, ..., x_n$ is defined as $\neg(x_1 \wedge x_2 \wedge \cdots \wedge x_n)$. Let True be represented by 1, False by 0, and n is a known, fixed constant. Can this function be represented by a Perceptron? If so, construct a Perceptron that does it; if not, argue why not.

Yes because it must output 0 only when all its inputs are 1s. A Perceptron that achieves this has all weights equal to -1 and the threshold at the single output unit equal to $-n + 0.5$.

(b) [4] The SAME function of two inputs, x_1, x_2, is defined to be 1 if the inputs are both the same (i.e., both 0 or both 1), and 0 otherwise. Can this function be represented by a Perceptron? If so, construct a Perceptron that does it; if not, argue why not.

No because this is not a linearly separable function.

(c) [4] Explain the difference in the Perceptron Learning Algorithm when a sigmoid activation function is used at the output units versus when a linear threshold unit (LTU) is used at the output units.

The weight update function differs in the Perceptron Learning Algorithm differs. For LTUs, $\Delta w_i = \alpha(T - O)x_i$ where T is the desired output of the output unit, O is the actual output, α is the learning rate parameter, and x_i is the value of the ith input unit. For the sigmoid function, $\Delta w_i = \alpha(T - O)O(1 - O)x_i$ where the extra term is the derivative of the sigmoid function.

(d) [3] When is a Support Vector Machine better to use than a Perceptron? Answer by describing characteristics of a specific situation that makes your point.

Perceptrons can only represent linearly separable functions, but SVMs can represent nonlinear functions by mapping the inputs into a higher dimensional space where the data is linearly separable. In addition, by maximizing the margin between the two classes, SVMs are usually more robust to noise and outliers than Perceptrons.

2. [20] **Probabilistic Reasoning**

Acute Pancreatitis (*AP*) is an acute inflammation of the pancreas that has multiple causes. Suppose that the prior probability of *AP* is 1 in 1,000 in a particular population. The blood serum level of amylase (*AMY*) is often one piece of information that is used in diagnosing *AP*. Assume *AMY* can be measured at just two levels, high and low. A high level of amylase is an indicator of *AP*. Suppose that from training data we know $P(\text{AMY=high} \mid \text{AP=true}) = 0.95$ and $P(AMY=low \mid AP=false) = 0.98$.

(a) [5] What is the false positive rate of using the level of amylase for diagnosing *AP*? That is, what percentage of the time would this test lead you to wrongly suspect *AP* in a patient? Write this in terms of $P(\ldots \mid \ldots)$ and then compute the answer.

$$P(AP=false \mid AMY=high) = \frac{P(AMY=high \mid AP=false)P(AP=false)}{P(AMY=high)}$$

$$= \frac{P(AMY=high \mid AP=false)P(AP=false)}{P(AMY=high \mid AP=false)P(AP=false)+P(AMY=high \mid AP=true)P(AP=true)}$$

$$= \frac{(1-0.98)(999/1000)}{(1-0.98)(999/1000)+(0.95)(1/1000)}$$

$$= 0.955$$

So the false positive rate is 95.5%!

(b) [5] Compute $P(AMY=high)$.

This is equal to the denominator in (a), so $P(AMY=high) = 0.021$.

(c) [5] Compute $P(AP=true \mid AMY=high)$.

$$P(AP=true \mid AMY=high) = 1 - P(AP=false \mid AMY=high) = 1 - 0.955 = 0.045$$

(d) [5] Show the full joint probability distribution table for this problem.

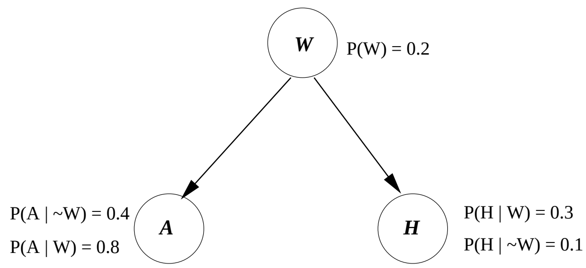
AP	*AMY*	*P(AP, AMY)*
true	high	0.00095
true	low	0.00005
false	high	0.01998
false	low	0.97902

3. [15] **Bayesian Networks**

Consider a situation in which we want to reason about the relationship between the weather and traveling. We'll use 3 Boolean random variables to represent "the Weather is bad" (W), "Airline flights between Madison and Chicago are cancelled" (A), and "automobile travel on Highway 90 between Madison and Chicago is delayed" (H).

Intuitively, we know that for any given weather condition, cancelled flights do not affect highway delays, and highway delays do not affect flight cancellations. From experience we know that if the weather is bad, there is an 80% chance that flights will be cancelled. If the weather is good, there is a 40% chance that flights will be cancelled. Furthermore, bad weather causes highway delays 30% of the time, but in good weather the chance of delays is 10%. There is a 20% chance of bad weather in Madison this time of year.

(a) [4] Create a Bayesian Network for this problem, including both the graph and the conditional probability tables.



(b) [5] Compute $P(\neg A, W, H)$.

$$P(\neg A,\ W,\ H) = P(\neg A \mid W, H)\, P(W, H)$$

$$= P(\neg A \mid W)\, P(H \mid W)\, P(W)$$

$$= (0.2)(0.3)(0.2)$$

$$= 0.012$$

(c) [6] Compute $P(A \mid H)$.

$$P(A \mid H) = \frac{P(A, H)}{P(H)}$$

$$P(A, H) = P(A, H, W) + P(A, H, \neg W)$$

$$= P(A \mid W) P(H \mid W) P(W) + P(A \mid \neg W) P(H \mid \neg W) P(\neg W)$$

$$= (.8)(.3)(.2) + (.4)(.1)(.8)$$

$$= 0.048 + 0.032 = 0.08$$

$$P(H) = P(H,A,W) + P(H,A,\neg W) + P(H,\neg A,W) + P(H,\neg A,\neg W)$$
$$= 0.08 + P(\neg A \mid W)P(H \mid W)P(W) + P(\neg A \mid \neg W)P(H \mid \neg W)P(\neg W)$$
$$= 0.08 + (.2)(.3)(.2) + (.6)(.1)(.8)$$
$$= 0.14$$

So, $P(A \mid H) = \frac{0.08}{0.14} = 0.57$

4. [10] **Situation Calculus**
Consider a domain where playing cards are manipulated in various ways, and one of the actions is to *Flip* a card over, i.e., if a card is face down, then *Flip* results in the card being face up; and if a card is face up, then *Flip* results in it being face down.

(a) [3] We want to use predicates *Card*, *FaceUp*, and *FaceDown* for this domain. Which of these predicates should be fluents?

```
FaceUp and FaceDown should be fluents because they can change
over time.  Card is not a fluent because an object that's a card
will always be a card.
```

(b) [4] Using the given three predicates and one action, write a sentence in the Situation Calculus that describes the act of flipping a card from face down to face up.

$$\forall s,c\ (Card(c) \wedge FaceDown(c,\ s)) \rightarrow FaceUp(c,\ Result(Flip(c),\ s))$$

(c) [3] Would your sentence in (b) require one or more additional frame axioms to completely specify the *Flip* action? Explain why or why not.

```
Yes, it does require a frame axiom in order to say that all other
cards do not change in any way when flipping the given card.
```

5. [20] **Partial-Order Planning**

(a) [4] The POP algorithm makes choices at various stages of its execution. One choice is made when selecting which open precondition to solve next. What are two other choice points?

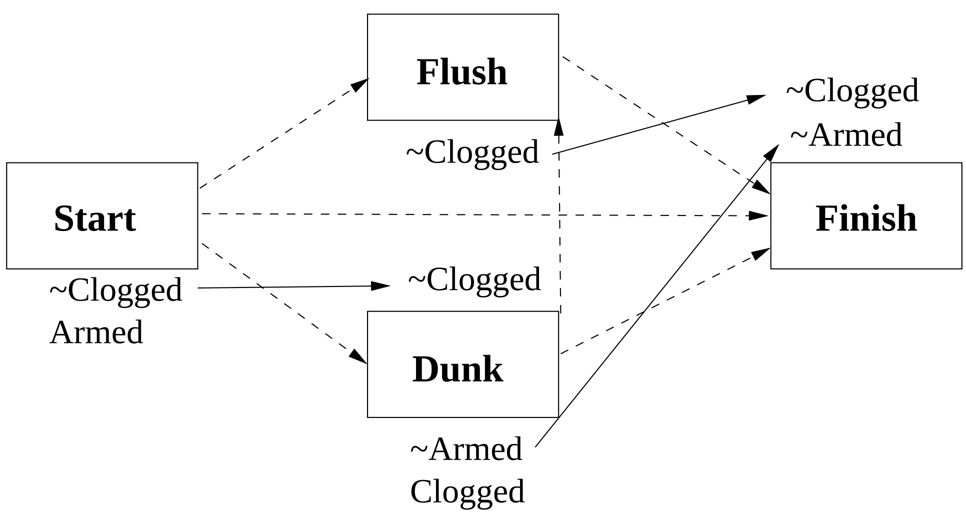
```
Three choice points are (1) whether to use simple establishment
or step addition to satisfy an open precondition, (2) when step
addition is used, which operator should be chosen if many are
applicable to satisfy a given open precondition, and (3) whether
to resolve a threat by promotion or demotion.
```

(b) [16] Suppose there is a bomb in your bathroom. The bomb is armed and the toilet is unclogged. Your goal is to disarm the bomb and have the toilet be unclogged. The only way to disarm the bomb is to dunk it in the toilet, provided the toilet is unclogged. Dunking the bomb in the toilet causes the toilet to be clogged. A clogged toilet can be unclogged by flushing it.

We can formulate this problem using two binary variables, *Clogged* and *Armed*. Initially, we have ¬*Clogged* and *Armed*. The goal is ¬*Clogged* and ¬*Armed*. We'll use the following two STRIPS-type operators:

DUNK	**FLUSH**
Preconditions: ¬*Clogged*	Preconditions: --
Effects: ¬*Armed*, *Clogged*	Effects: ¬*Clogged*

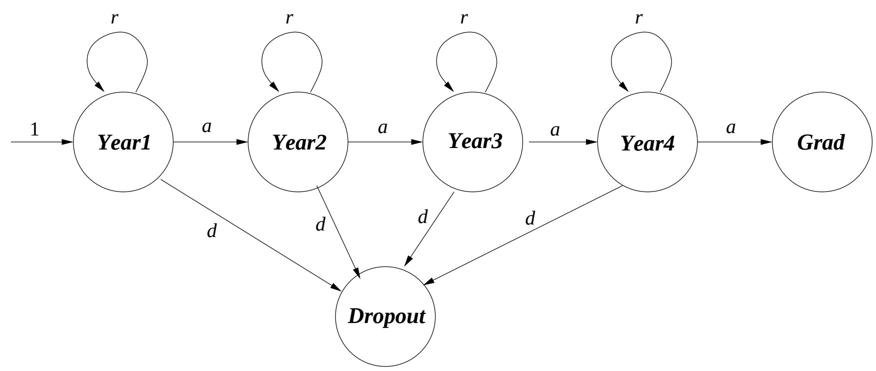
Starting from the initial plan for this problem, apply a sequence of the POP algorithm's plan modification operations that results in a complete plan (i.e., one with no open preconditions and no threats). Draw the final plan using the following notation: Draw steps as boxes with the name of the operator inside. Above the box list the preconditions and below the box list the effects. If there are any threats that had to be resolved, explain what threatened what and how it was resolved.




```
The step Dunk threatened the causal link from Flush to Finish,
and this was fixed by Demotion.
```

6. [20] **Markov Models and Hidden Markov Models**

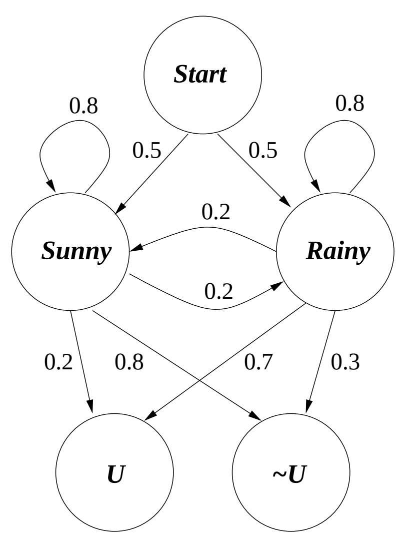
(a) [5] A high school student has three possible outcomes for each school year: dropping out with probability *d*, repeating that year with probability *r*, and advancing to the next year with probability *a*. There are 4 years in high school, after which the student graduates. Draw a *Markov Model* for this situation, including defining the states, show all possible transitions with probabilities where possible, and specify any constraints on the parameters in this model.



In addition, $a + r + d = 1$.

(b) [15] Consider a simplified version of the "weather" *Hidden Markov Model* (HMM) presented in the paper by E. Fosler-Lussier. For this problem, assume the weather can be either sunny (S) or rainy (R) each day, each with prior probability 0.5. From one day to the next, the weather stays the same with probability 0.8 and changes with probability 0.2. On a sunny day, the probability that you will observe the caretaker carrying an umbrella (U) is 0.2, and on a rainy day it is 0.7. The prior probability that the caretaker carries an umbrella on any given day is 0.5. The weather outside is hidden to you and only the presence or absence of an umbrella each day is observed.

(i) [5] Draw the HMM. Include hidden states, observable values, π vector, and label all arcs with probabilities.



The hidden states are *Sunny* and *Rainy*, the observable states (values) are U and $\neg U$. The π vector is shown with the probabilities on the arcs out of the *Start* state.

(ii) [5] If on the first day you observe no umbrella, what is the chance that the weather that day is rainy? In other words, compute $P(w_1 = R \mid u_1 = \neg U)$. (Hint: Use the product rule.)

$$P(w_1{=}R \mid u_1{=}\neg U) = \frac{P(R,\ \neg U)}{P(\neg U)}$$

$$= \frac{P(\neg U \mid R)\,P(R)}{P(\neg U \mid R)\,P(R) + P(\neg U \mid S)\,P(S)}$$

$$= \frac{(0.3)(0.5)}{(0.3)(0.5) + (0.8)(0.5)}$$

$$= 0.27$$

`Note: Many people used` $P(\neg U)$ `= 0.5 as the denominator though, as in the homework, this is not correct. Because of confusion about this in the homework without time to go over it in class, I accepted use of this value for this question.`

(iii) [5] What is the probability of seeing no umbrella on the first day, and seeing an umbrella on the second day, and the first day is rainy, and the second day is sunny? That is, compute $P(u_1 = \neg U,\ u_2 = U,\ w_1 = R,\ w_2 = S)$.

$$P(u_1 = \neg U,\ u_2 = U,\ w_1 = R,\ w_2 = S) = P(S \mid R)\,P(R)\,P(\neg U \mid R)\,P(U \mid S)$$

$$= (0.2)(0.5)(0.3)(0.2)$$

$$= 0.006$$