

Final Examination
CS540: Introduction to Artificial Intelligence

December 2008

LAST NAME: _____

FIRST NAME: _____

Problem	Score	Max Score
1	=====	15
2	=====	15
3	=====	10
4	=====	20
5	=====	10
6	=====	20
7	=====	10
Total	=====	100

Question 1. [15] Probabilistic Reasoning

Consider the age group of women over forty. 1% of women who are screened, have breast cancer. 80% of women who really do have breast cancer will have a positive mammography (meaning the test indicates she has cancer). 9.6% of women who do *not* actually have breast cancer will have a positive mammography (meaning that they are incorrectly diagnosed with cancer). Define two Boolean random variables, M meaning a positive mammography test and $\neg M$ meaning a negative test, and C meaning the woman has breast cancer and $\neg C$ means she does not.

(a) [5] If a woman in this age group gets a positive mammography, what is the probability that she actually has breast cancer? Show the key steps.

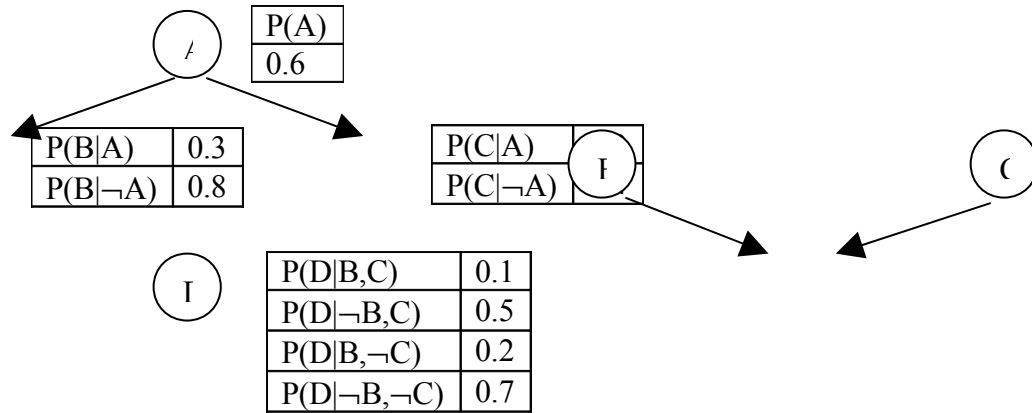
(b) [2] True or False: The "Prior" probability, indicating the percentage of women with breast cancer, is not needed to compute the "Posterior" probability of a woman having breast cancer given a positive mammography.

(c) [6] Say a woman who gets a positive mammography test, M_1 , goes back and gets a second mammography, M_2 , which also is positive. Use the Naive Bayes assumption to compute the probability that she has breast cancer given the results from these two tests.

(d) [2] True or False: $P(C \mid M_1, M_2)$ can be calculated in general given only $P(C)$ and $P(M_1, M_2 \mid C)$.

Question 2. [15] Bayesian Networks

Consider the following Bayesian Network containing four Boolean random variables:



(a) [4] How many independent values are required to store the full joint probability distribution for this problem?

(b) [2] True or False: From the above network it is possible to compute any joint probability of the four variables.

(c) [2] True or False: Based on the topology only of the network, $P(A,C,D) = P(A)P(C)P(D)$

(d) [2] True or False: Based on the topology only of the network, $P(D|C) = P(D|A,C)$

(e) [5] Compute $P(A, B,C,¬D)$ where A means $A = true$, $¬D$ means $D = false$, etc.

Question 3. [10] Support Vector Machines

We want to construct a Support Vector Machine (SVM) that computes the XOR function. Instead of input and output values of 1 and 0, we'll use values 1 and -1, respectively. So, for example, if the input is $[x_1 = -1, x_2 = 1]$ we want the output to be 1.

(a) [5] Using the four possible input vectors and their associated outputs, can a LINEAR SVM be constructed to correctly compute XOR? If it can, show how by drawing the four possible input value pairs in the 2D input space, x_1 , x_2 , and the separator (i.e., decision boundary) computed by the SVM. If it cannot, explain or show why not.

(b) [5] Suppose we re-express the input data using the computed features $[x_1, f(x_1, x_2)]$ instead of the original $[x_1, x_2]$ pair of values, where $f(x_1, x_2)$ is some function of both x_1 and x_2 . Can a LINEAR SVM be constructed to correctly compute XOR using the computed features rather than the raw features? If it can, show how by defining the function $f(x_1, x_2)$, and drawing the four possible input value pairs in the 2D input space, x_1 , your $f(x_1, x_2)$, and the separator (i.e., decision boundary) computed by the SVM. If it cannot, explain or show why not.

Question 4. [20] Hidden Markov Models

Andy is a three-month old baby. He can be happy, hungry, or having a wet diaper. Initially when he wakes up from his nap at 1pm, he is happy. If he is happy, there is a 50% chance that he will remain happy one hour later, a 25% chance to be hungry by then, and a 25% chance to have a wet diaper. Similarly, if he is hungry, one hour later he will be happy with 25% chance, hungry with 25% chance, and wet diaper with 50% chance. If he has a wet diaper, one hour later he will be happy with 50% chance, hungry with 25% chance, and wet diaper with 25% chance. When he is happy, he smiles 75% of the time and cries 25% of the time; when he is hungry, he smiles 25% and cries 75%; when he has a wet diaper, he smiles 50% and cries 50%.

(a) [5] Draw the HMM that corresponds to the above story. Clearly mark the transition probabilities and output probabilities.

(b) [5] The nanny left a note: “1pm: smile. 2pm: cry. 3pm: smile”. What is the probability that this particular observed sequence happens?

(c) [5] What is the most likely hidden sequence (in terms of happy, hungry, or wet diaper) for the note in (b)?

(d) [5] (This question not related to above) Describe the McGurk effect (“hear with your eyes”) in one sentence. In another sentence, discuss its implication to automatic speech recognition.

Question 5. [10] Clustering

K-means clustering tries to minimize the distortion $\sum_i (x_i - c_i)^2$ where c_i is the center (mean) of the cluster that point x_i is in.

(a) [4] Given a dataset with five points $\{1,4,6,7,8\}$, and $K=2$ clusters whose initial centers are $c_1=0, c_2=9$, run K-means clustering by hand. Show

- i) the final cluster centers,
- ii) the points in the two clusters respectively,
- iii) the distortion.

(b) [4] Repeat (a), but with initial centers at $c_1=0, c_2=6$.

(c) [2] Briefly discuss what property of K-means (a) and (b) show.

Question 6. [20] Game Theory

Consider the game in matrix normal form below:

	x	Y	z
a	0	-1	1
b	1	0	-1
c	-1	1	0

In this question you will derive the optimal mixed strategy for both players. We will refer to the players as the XYZ player and the ABC player. The numbers are from the ABC player's perspective. For all the questions below, you do NOT need to formally prove your answer.

(a) [4] Say player XYZ plays strategy x, y and z with probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$. Give the best pure strategy and expected payoff for the ABC player.

(b) [4] Say player XYZ plays strategy x, y and z with probabilities 0, p, 1-p. Give the best pure strategy and expected payoff for the ABC player. This shows that if player XYZ only mixes between two strategies, then player ABC has an advantage.

(c) [4] Say player XYZ plays strategy x , y and z with probabilities p_x , p_y , p_z . What will the payoff for player ABC be, given that ABC plays a pure strategy?

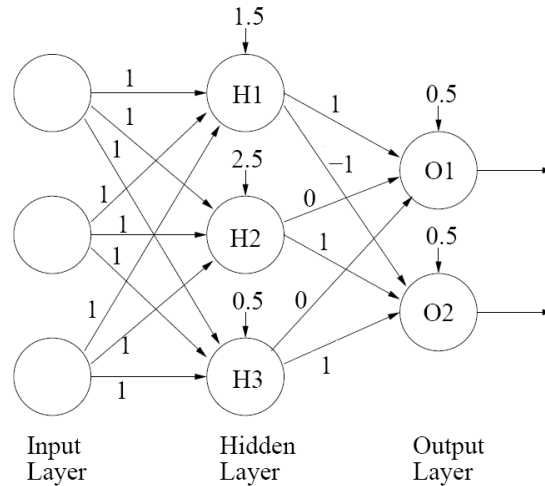
(d) [4] What is the best strategy for player XYZ, so that XYZ can even tell ABC the values of p_x , p_y , p_z , but ABC cannot take advantage of this knowledge?

(e) [3] Give a mixed-strategy Nash equilibrium of this game.

(f) [1] Name an instance of this game in real life.

Question 7. [10] Neural Networks

The following feedforward neural network takes three binary (0 or 1) inputs and produces two binary (0 or 1) outputs. Each node uses a *Linear Threshold Unit* as its activation function with the associated threshold value. Call an input vector that has exactly n inputs equal to "1" as an input vector with "count" n , for $n = 0, 1, 2$ or 3 . Note that for this particular neural network, all input vector with the same "count" have the same output at both output units. In other words, for a given "count," the output of the network is the same no matter which particular input units are the ones with input equal to "1."



(a) [5] For a given input "count" of n , describe what is computed as the output of each hidden unit. Give your answer in terms of n and do not give simply a literal translation of each individual calculation performed.

(b) [5] For a given input "count" of n , give an interpretation of the computed output $O1$, $O2$.