Csci 543. Advanced Artificial Intelligence **Fall, 2017**

Instructor: Dr. M. E. Kim Date: December 5th, 2017

Deadline: December 14th (Thr.) 12:00 PM STRICTLY

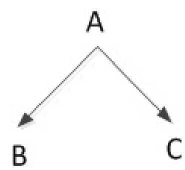
Final Exam (300 points + 50 optional)

Name: ____________________________

1. You have to work on the exam for yourself, independently – do NOT cooperate/collaborate with other students.
2. Your answer should be fully explained; any sloppy answer without computation/explanation would not get a full point.
3. For each question, you should show the formula(s) and the essential computational steps. Any correct answer neither with its correct formula and computational step nor a sufficient explanation will gain NO point.
4. Any plagiarism or cooperation will result in the F grade for the final grade of the course regardless your previous grade of HWs and that of paper.
5. You should TYPE your answer in the file – NO photo image of handwriting is allowed.
6. Save your file in MS-Word, naming 'Final-YourLastName': e.g.) Final-Kim.docx.
7. A strict grading scheme and a late submission policy will be applied to the final exam than to your past assignments: -10% deduction per hour after 12:00 PM.
8. No Extension of deadline will be given due to the moving of the department.
9. Submit it in the blackboard – No Email submission is allowed.

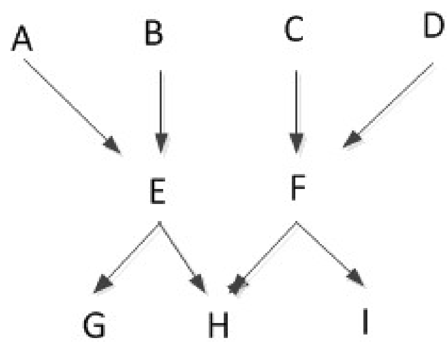
Q1. [10] Bayesian Network

For a given Bayesian network where P(a) = .6, P(b | a) = .8, P(b | ¬a) = .4, P(c | a) = .4 and P(c | ¬a) = .3, compute P(c | b). Note that a, ¬a, b, etc. are propositions: e.g.) a ⇔ A = true, , ¬a ⇔ A = false.



Q2. [20 pt] Conditional Independence

In the given Bayes network, decide the conditional independence of the nodes.



(1) [5] $A \perp B$

(2) [5] $A \perp C \mid G$

(3) [5] $A \perp B \mid E$

(4) [5] $A \perp B \mid G$

Q3. [30] Bayesian Learning.

Consider a medical diagnosis problem in which there are two alternative hypotheses:

- h_1: the patient has a particular form of cancer,
- $h_2 = \neg h_1$: the patient does not have a cancer.

The available data is from a particular lab test with two possible outcomes: +(positive) and - (negative). We have prior knowledge that over the entire population of people only 0.006 have this disease. Furthermore, the lab test is only an imperfect indicator of the disease. The test returns a correct positive result in only 97% of the cases in which the disease is actually present and a correct negative result in only 95% of the cases in which the disease is not present. In other cases, the test returns the opposite result.

(1) [6] Give the following probabilities which summarize the above situation

a) $P(h_1) = P(cancer)$

b) $P(h_2) = P(\neg\ cancer)$

c) $P(+ \mid h_1)$

d) $P(- \mid h_1)$

e) $P(+ \mid h_2)$

f) $P(- \mid h_2)$

(2) [4] Suppose we now observe a new patient for whom the lab test returns a positive result. What is the *maximum a posteriori* (MAP) hypothesis?

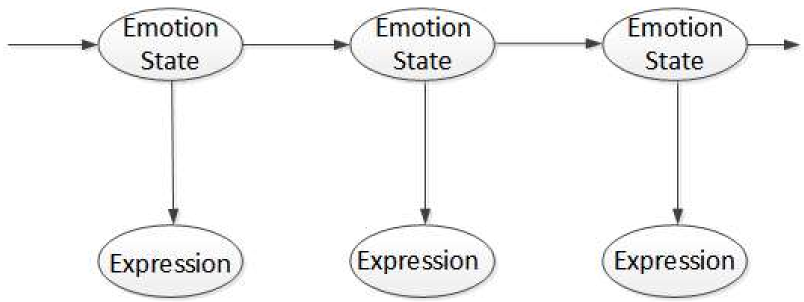
(3) [10] Suppose the doctor decides to order a second laboratory test for the same patient, and suppose the second test returns a positive result as well. What are the posterior probabilities of *cancer* and ¬ *cancer* following these two tests? Assume that the two tests are independent.

(4) [10] What is your prediction for the 3rd test result (d_3), based on the previous two lab test results (d_1, d_2)? Compute the predicted probability $P(d_3 = + \mid d_1 = +, d_2 = +)$, then give the prediction.

Q4. [30] HMM

Andrew lives a simple life. Some days he's angry and some days he's happy. But he hides his emotional state, and so all you can observe is whether he smiles, frowns, laughs, or yells.
We start on day 0 in the Happy state, and there's one transition per day.

$P(S_0\text{=happy})=0.5$



$P(S_{t-1})$	$P(S_t = \text{happy})$
happy	0.8
angry	0.2

$P(S_t)$	$P(E_t = \text{smile})$	$P(E_t = \text{frown})$	$P(E_t = \text{laugh})$
happy	0.4	0.1	0.3
angry	0.1	0.5	0.2

S_t = Emotion State on day t $\in$ {*happy, angry*}
E_t = Observation by Expression on day t $\in$ {*smile, frown, laugh, yell*}

(1) [5] Compute $P(S_2 = angry)$.

(2) [5] Compute $P(E_2 = frown)$.

(3) [5] Compute $P(S_2 = happy \mid E_2 = frown)$.

(4) [5] Compute $P(E_{100} = yell)$.

(5) [10] Assume that $E_1 = frown, E_2 = frown, E_3 = frown, E_4 = frown, E_5 = frown$.
What is the most likely sequence of states?

Q5. [30] Naïve Bayes Model

We have 2 classes of movies: NEW and OLD.
The following training set of 3 Boolean attributes, *x, y, z*, and a class, C, represent each of three features of movie and the class of movie, respectively, where 1 = true and 0 = false.
Suppose you have to predict Class of a movie using a Naïve Bayes Model.

x	*y*	*z*	***Class***
0	1	1	***NEW***
1	0	1	***OLD***
0	1	1	***OLD***
1	1	0	***OLD***
1	0	0	***NEW***
0	0	1	***OLD***
1	1	0	***NEW***

(1) [5] What is P(*OLD* | *x*=1) learned for the training data?

(2) [5] What is P(*OLD* | *x*=1, *y*=0, *z*=1) learned for the training data?

(3) [5] After learning is complete, compute the predicted probability P(OLD | *x*=1, *y*=0, *z*=1) ?

(4) [5] How would a naive Bayes classifier predict Class given the input $<x = 1, y = 0, z = 1>$?
Assume that in case of a tie the classifier prefers to predict *OLD* for Class.

(5) [10] Using the probabilities obtained during the Bayes classifier training, compute the predicted probability P(*OLD* | *x* = 1)?

Q6. [60] Maximum Likelihood (ML) Learning

The table shows the examples of SPAM and those of HAM messages which are consisted of some words whose dictionary size is twelve. Suppose that you've received SPAM messages for the 1st 3 days, then HAM messages for the next 5 days, i.e. a data as a sequence of message is <Spam, Spam, Spam, Ham, Ham, Ham, Ham, Ham >.

SPAM	***HAM***
offer is secret	play golf tomorrow
click secret link	went play golf
secret golf link	secret golf event
	golf is tomorrow
	golf costs money

(1) [10] Compute the ***maximum likelihood*** of *SPAM, i.e.* P(*SPAM*)=θ, using a *log-likelihood.*

(2) [10] In the Bayesian network of this ML parameter learning,
   (A) [7] Draw the BN with the CPT of the required parameters (e.g. θ_1, θ_2, …..). -- You don't have to compute the exact values of parameters yet.

   (B) [3] How many parameters are required?

(3) [10] By ML-learning, compute a parameter value, P("*secret*" | *SPAM*) and P("*secret*" | *HAM*), respectively.

(4) [10] Now, the new message "*golf*" is received. What is the probability that this message is *SPAM*?

(5) [10] The new message "*secret is secret*" is received. What is the probability that this is *SPAM*?

(6) [10] For a new message, "*tomorrow is secret*", what is the probability that the message is *SPAM* and *HAM*, respectively?

Q7. [30] Markov Decision Process

An agent is situated in the 4×2 fully observable environment in the figure. Beginning in the start state at (A, 1), it chooses an action at each time step among {*North, East, South, West*}. For any action, the probability that an action is successful as it's intended is *p* while the probability that an action fails and it is reversed is *1-p*. For example, P(Action=*East*) = *p* if an action is successful as intended while P(Action=*West*) = *1-p* due to its failure. If an action is bounced, the agent stays in the current state. The terminal states are (B, 1) with a reward -100 and (B,4) with a reward 100.
Suppose that p = .7, the reward at each state (except (B,1) and (B,4)) = -5 and a discount factor γ= 0.9.

	1	2	3	4
A				
B	-100			100

(1) [10] Compute the utility value of a location (A, 4) by single action.

(2) [10] By applying Value Iteration algorithm, what is the final utility value of a state (A, 4) after convergence?

(3) [10] What is the optimal policy for each state (A, 1), (A, 2), …. , (B, 2) and (B, C), respectively?

Q8. [50] Decision Making

A used-car buyer can decide to carry out various tests with various costs (e.g., kick the tires, take the car to a qualified mechanic) and then, depending on the outcome of the tests, decide which car to buy.
We will assume that the buyer is deciding whether to buy car A, that there is time to carry out at most one test, and that T_1 is the test of A and costs \$*50*.
A car can be in good shape (quality q^+) or bad shape (quality q^-), and the tests might help to indicate what shape the car is in. Car A costs \$2*,000*, and its market value is \$*2,500* if it is in good shape; if not, \$1,000 in repairs will be needed to make it in good shape. The buyer's estimate is that A has a *60%* chance of being in good shape.

(1) [10] Draw the decision network that represents this problem.

(2) [10] Calculate the expected net gain from buying A, given no test.

(3) [10] Tests can be described by the probability that the car will pass or fail the test given that the car is in good or bad shape.

$$P(pass(A, T_1) \mid q^+(A)) = 0.7; \quad P(pass(A, T_1) \mid q^-(A)) = 0.35.$$

Use Bay'es theorem to calculate the probability that the car will pass (or fail) its test and hence the probability that it is in good (or bad) shape given each possible test outcome.

(4) [10] Calculate the optimal decision given either a pass or a fail, and their expected utility.

(5) [10] Calculate the value of information of the test, and derive an optimal conditional plan for the buyer.

Q9. [40] Decision Tree Learning : Correction in Red.

In electronic commerce applications we want to make predictions about what a user will do. Consider the following made-up data used to predict whether someone will ask for more information (MoreInfo) based on whether they accessed from an educational domain (Edu), whether this is a first visit (Frst), whether they have bought goods from an affiliated company (Bought), and whether they have visited a famous online information store (Visited).

Example	*Bought*	*Edu*	*Frst*	*Visited*	*MoreInfo*
*e*1	false	true	false	false	*true*
*e*2	true	false	true	false	*false*
*e*3	false	false	true	true	*true*
*e*4	false	false	true	false	*false*
*e*5	false	false	false	true	*false*
*e*6	true	false	false	true	*true*
*e*7	true	false	false	false	*true*
*e*8	false	true	true	true	*false*
*e*9	false	true	true	false	*false*
*e*10	true	true	true	false	*true*
*e*11	true	true	false	true	*true*
*e*12	false	false	false	false	*true*

We want to use this data to learn the value of *MoreInfo* as a function of the values of the other variables. Suppose we measure the error of a decision tree as the number of misclassified examples. The optimal decision tree from a class of decision trees is an element of the class with minimal error.

(1) [10] Draw the optimal decision tree that would be learned from this data set to classify the data. You should show the necessary computational steps of information gain to generate the optimal decision tree.

(2) [10] Write the hypothesis which is generated from (1) in the logical expression.

(3) [5] Suppose we have a test data set as follows.

Example	*Bought*	*Edu*	*Frst*	*Visited*	*MoreInfo*
*e*13	true	true	true	true	true
*e*14	true	true	false	false	true
e 15	true	false	true	true	true
*e*16	false	true	false	true	true

(A) Classify them using the decision tree in (1).

(B) Which example(s) is (or are) false negative or false positive? Consider the classification of 'Yes' as 'positive' while 'No' as 'negative'.

(4) [5] What is the degree of accuracy with the test examples in (3)?

(5) [10] Refine the hypothesis in (2) to handle the misclassified example by generalization or/and by specialization.

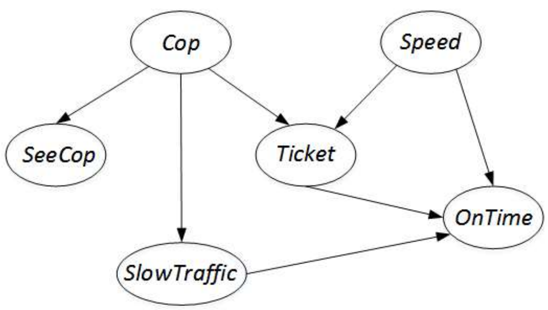
Q10. [50, optional] Decision Making with Bayesian Networks

The following figure shows a simple Bayesian network to help you decide where or not to speed on the highway on your way to your parent's home for Thanksgiving dinner. Each node in the network represents a random Boolean variable. The random variable *Cop* indicates whether there is a highway patrol cop present on the freeway; *SeeCop* indicates whether you have detected a cop or not. The variable *SlowTraffic* indicates whether traffic is moving slower than the posted speed limits and *Speed* indicates whether or not you are speeding. *Ticket* is true when you get a ticket, and *OnTime* indicates that you made it on-time for the festivities. You are given the following probabilities.

P(*Speed* = true) = 0.25, P(*Cop* = true) = 0.1,
P(*SeeCop* = true | *Cop* = true) = 0.6, P(*SeeCop* = true | *Cop* = false) = 0.0,
P(*SlowTraffic* = true | *Cop* = true) = 0.8, P(*SlowTraffic* = true | *Cop* = false) = 0.3,

$$P(\textit{Ticket} = \text{true} \mid \textit{Cop}, \textit{Speed}) = \begin{cases} 0.5 & \text{if } \textit{Cop}, \textit{Speed} \text{ are both true;} \\ 0 & \text{otherwise,} \end{cases}$$

$$P(\textit{OnTime} = \text{true} \mid \textit{Ticket}, \textit{Speed}, \textit{SlowTraffic}) = \begin{cases} 0 & \text{if } \textit{Ticket} \text{ is true} \\ 0.9 & \text{if } \textit{Ticket} \text{ is false, } \textit{Speed} \text{ is true, } \textit{SlowTraffic} \text{ is false,} \\ 0.5 & \text{if } \textit{Ticket} \text{ is false, } \textit{Speed} \text{ is true, } \textit{SlowTraffic} \text{ is true,} \\ 0.3 & \text{if } \textit{Ticket} \text{ is false, } \textit{Speed} \text{ is false, } \textit{SlowTraffic} \text{ is false,} \\ 0.1 & \text{if } \textit{Ticket} \text{ is false, } \textit{Speed} \text{ is false, } \textit{SlowTraffic} \text{ is true,} \end{cases}$$



For each of scenario s_i below,

- s_1: You don't detect a cop and you speed.
- s_2: You don't detect a cop and you do not speed.
- s_3: You don't detect a cop, but traffic is slow, and you speed.
- s_4: You don't detect a cop, but traffic is slow, and you do not speed.

(1) [15] Compute the probability of receiving a ticket for each scenario s_i.
(2) [15] Compute the probability of arriving on time for each scenario s_i.
(3) [10] Suppose the cost of a speeding ticket is \$100, and you lose a \$10 bet to your brother if you arrive late. Using your answers from questions (2-3), compute the *expected utility* in each scenario s_i.
(4) [10] Now determine whether you should speed or not in these two scenarios
(A) a cop is not detected, and
(B) a cop is not detected, but traffic is slow.