**CSci 384: Artificial Intelligence Spring, 2017**

**Instructor: Dr. M. E. Kim** **Date: May 4th, 2017**

**Due: 3:15 PM, May 11th (Thr.)**

**Final Exam:**

**Total: 224/ 250 point**

**Instruction:**

**Q3 – Q5.** In order to compute the required probability, you have to define the proper formula and show the ***essential*** computational steps precisely; for instance,

1. Define the variables (if they’re not defined) and what needs to be computed: e.g.) P(C|B)

If the variables are pre-defined, use them only: e.g.) Q5.1. G – large gas.

1. Derive the formulas for computation, step by step:

e.g.) P(C|B) = P(B| C)⋅P(C)

1. Assign the values to the derived formula to complete the computation.

e.g.) P(C|B) = P(B|C)⋅P(C ) = 0.2 ⋅ 0.6 = 0.12

1. For the inference using Variable Elimination, clearly show the factors and the computation with them.

**Q6 .** You have to clearly define the formula of information gain or the entropy and show their ***essential*** computational steps. In the full decision tree, the final classification should be specified in each leaf.

**Q7 – Q8 .** In order to get the hypothesis or do prediction, you have to define the proper formula and show the ***essential*** computational steps.

∙ Any answer without clear computational steps or sufficient description will **NOT** get a full point while an excellent answer would be rewarded.

∙ Save your assignment file under the name of **Final-YourLastNameOnly**: e.g.) Final-Kim.docx

∙ Upload it to Submission section by 3:15 PM, May 12th , 2017.

∙ Hours taken to complete the exam: \_5\_ Hours \_\_\_ Minutes.

∙ Mark the difficulty of the Exam:

Very Easy: \_\_\_\_\_ Easy: \_\_\_\_\_ Moderate: \_\_\_\_\_ Difficult: \_X\_\_ Very Difficult: \_\_\_\_\_

**Q1. [26/ 25] 1st-Order Logic (FOL).** Perfect. +1

Convert the following sentences into 1st-order logic sentences, using the following predicates:

• *P(x)* = *x* is a programmer. • *R(x)* = *x* is red-haired.

• *S(x)* = *x* is smart. • *L(x, y)* = *x* likes *y.*

1. [5] No programmer is smart.

∀x( P(x ) -> ¬S(x) )

1. [5] Not everyone is a programmer.

∃x ( ¬ P(x) )

1. [5] Everyone is not a programmer.

∀x ( ¬P(x) )

1. [5] Some programmers are not smart.

∃x ( P(x) ∧ ¬S(x) )

1. [5] There is some with red-hair whom everyone likes.

∀x ∃y ( R(y) ∧ L(x,y) )

**Q2. [35]** **Inference in the 1st-Order Logic.**

Consider the following statements.

* (A) No software is guaranteed.
* (B) All programs are software.
* (C) **Conclusion:** Thus, no program is guaranteed.

1. [10] Translate the above statements in the 1st –Order Logical sentences using the following predicates. Clearly use the universal and/or existential quantifiers.

• *S(x)* = *x* is a software. • *G(x)* = *x* is guaranteed. • *P(x)* = *x* is a program.

1. ∀x( S(x) -> ¬G(x) )
2. ∀x( P(x) -> S(x) )
3. ∀x( P(x) -> ¬G(x) )
4. [ 5] Negate the conclusion in (1).

∃ x ( P(x) ∧ G(x) )

1. [10] Convert your sentences in (1.A & 1.B) and in (2) to Conjunctive Normal Form (CNF).
2. ∀x(¬S(x) ∨ ¬G(x))
3. ∀x (¬P(x) ∨ S(x))
4. P(K) ∧ G(K)
5. [10] Using ***resolution***, prove the conclusion is either true or false. Show you proof clearly with the substitution.

If P(K) ∧ G(K) is true, then both P(K) and G(K) must both be true. Therefore (¬P(x) ∨ S(x)) is also true, which makes S(x) true as well. We know that In order for (¬S(x) ∨ ¬G(x)) to be true, either G(x) or S(x) has to be true, however since G(x) and S(x) are both true then (¬S(x) ∨ ¬G(x)) is therefor false.

**Q3. [20] Uncertainty.**

A screening test is a low-cost way of checking large groups of people for a disease. A more costly

but accurate test shows that 1% of all people have the disease. The screening test indicates the disease

(test positive(+)) in 90% of those who have it, and in 10% of those who do not have the disease (false

positive(+)).

For the following questions, define the probability to compute and compute it by showing its essential computational steps.

P(a) = People without the disease. - 99

P(b) = People with the disease. - .01

P(c|b) = People with true test positive -.90

P(c|a) = people with false positive test -.1

P(e) = People test positive - (.01 \* .9)+ (.99 \* .1) = 10.8%

1. [10] What percent of people who test positive don’t have the disease (false +)?

P(a|c) = (P(c|a) \* P(a)) / P(e)

= (.1\*.99) / .108 = .9166667

1. [10] What percent of people who test negative do have the disease (false -)?

P(b|f) = ( (1-P(c|b) ) \* P(b)) /P(f)

= (.1\*.01) / .892 = .00112107

**Q4. [27/ 30] Probabilistic Reasoning**

A student of the AI class notices that people driving SUV’s (S) consume large amounts of gas (G) and are

involved in more accidents than the national average (A). He also noticed that there are two types of people that drive SUVs: people from North Dakota (N) and people with large families (F). After collecting some statistics, he arrives at the following:

1. There used to be 80% chance for people driving SUV’s(S) consume large amounts of gas(G), while
2. it’s only 20% chance of consuming large amounts of gas for people driving non-SUV vehicles.
3. People who drives SUV has a 70% chance of having an accident, while
4. those who drives non-SUV vehicle has only 30% chance of accident.
5. The people from North Dakota takes 2% of national population.
6. 40% of people of the nation used to be with the large family.
7. The probability that North Dakotan with large family drive SUV is 0:8, while
8. the probability that North Dakotan not with large family drive SUV is 0:6.
9. On the other hand, the probability that non North Dakotan with large family drive SUV is 0:5, while
10. the probability that people who are neither from North Dakota nor with large family drive SUV is 0:3.
11. [7/ 10] Using the given variables S, G, A, N and F, draw a ***Bayesian network*** which represent the above information, giving a (conditional) probability for each node.

(a) P(G|S) = .8  or 80%

(b) P(G|¬S) = .2 or 20%

(c) P(A|S) = .7 or 70%

(d) P(A|¬S) = .3 or 30%

(e) P(N) = .02 or 2%

(f) P(F|N,¬N) = .4 or 40%

(g) P(S|N,F) = .8  or 80%

(h) P(S|N,¬F) = .6 or 60%

(i) P(S|¬N,F) = .5 or 50%

(j) P(S|¬N,¬F) = .3 or 30% -3

1. [10] Compute the probability that people drive SUV.

P(S|A) = P(A|S) \* P(S)

P(S) = P (S, N, F) + P(S, N, ~F) + P(S, ~N, F) + P( S, ~N, ~F)

8 \* .02 \* .4 + .6 \* .02 \* .6 + .5 \* .98 \* .4 + .3 \* .98 \* .6 = .386

1. [10] For a driver who had an accident, what is the probability that his vehicle was a SUV?

P(S|A) = P(A|S) \* P(S)

P(S) = P(S, N, F) + P(S, N, ~F) + P(S, ~N, F) + P( S, ~N, ~F)

8 \* .02 \* .4 + .6 \* .02 \* .6 + .5 \* .98 \* .4 + .3 \* .98 \* .6 = .386

P(S|A) - 0.7\* .386, .0.3 \* .614

= .2702, . 1842

= .2702 / .4544, .1842 / .4544>

= .5946, .4054

**Q5. [23/ 25] Bayesian Network and Inference**

In the given Bayesian Network (BN) below,

1. [5] Find the nodes which is conditionally independent of ’C’ given its parent ’B’ and ’D’.

Node E, A are conditionally independent.

1. [5] (a) Find the ***Markov blanket*** of a node ‘B’.

- A, C, D

(b) Indicate what nodes are conditionally independent of B given its Markov blanket found in (a).

* E, F, G nodes

1. [5] Find the irrelevant node(s) to compute P(C | G).

* A is the only node

1. [8/ 10] A) Give the formula of computing the full joint distribution **P(*a, b, c, ¬ d, ¬ e, ¬ f, g*)** where *a, b, … ¬ f, g* are propositions such that A = true, B = true, …. F = false and G=true.

B) and **compute it.**

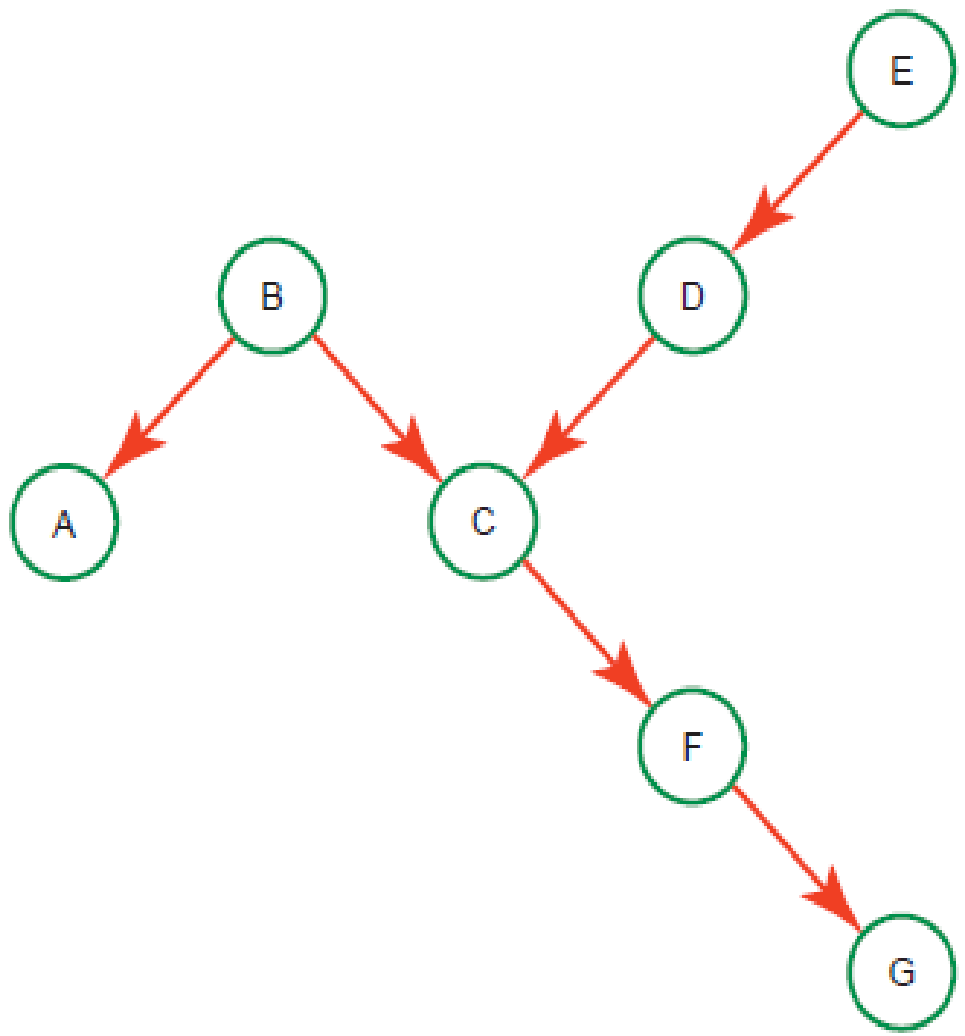
* P(A|B)\*P(B)\*P(C|B,¬D)\*P(¬D| ¬E)\*P(¬E)\*P(¬F|C)\*P(G|¬F)

= .88 \* .7 \* .33 \* 0.16 \*.09 \* .55 \* .96 = 0.0096598656 .001545578

1. [0/ **10, Optional**] Compute the distribution of P(A | g) using ***Variable Elimination,*** showing the created factors and the elimination of variables clearly.

P(E=T|C=T) \* P(C=T|A=F) / (P(E=T|C=T) \* P(C=T|A=F) + P(E=T|C=F) \* P(C=F|A=F))

0.7\*0.4 / (0.7 \* 0.4 + 0.2 \* 0.6) = 0.7



*P(b) = 0.7 P(e) = 0.91*

*P(a | b) = 0.88 P(a | ¬ b) = 0.38*

*P(c | b, d) = 0.93 P(c | b, ¬ d) = 0.33 P(c | ¬b, d ) = 0.53 P(c | ¬ b, ¬ d) = 0.83*

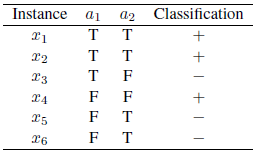
*P(d | e ) = 0.04 P(d | ¬ e) = 0.84*

*P(f | c ) = 0.45 P(f | ¬ c) = 0.85*

*P(g | f ) = 0.26 P(g | ¬ f) = 0.96*

**Q6. [30] Decision Tree Learning**

Consider the following set of training data.



1. [10] What is the entropy of this set of training examples with respect to the target function classification?

Entropy = -3/6 (log3/6) – 3/6 (log3/6) = 1

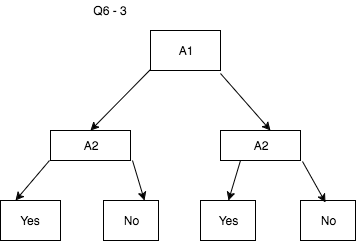
Entropy = 1

1. [10] What is the information gain of *a2* relative to these training examples?

Gain(*a2*) = 1-4/6 \* 1-2/6 \* 1 = 0

Gain of a2 = 0

1. [10] Draw the full decision tree for this data. You should show the computation of information gain of attribute to choose a root of each subtree.



**Q7. [2/ 45] 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 = *¬h*1: the patient does not has 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 *.*008 have this disease. Further- more, the lab test is only an imperfect indicator of the disease. The test returns a correct positive result in only 98% of the cases in which the disease is actually present and a correct negative result in only 97% of the cases in which the disease is not present. In other cases, the test returns the opposite result.

1. [0/ 10] Compute the following probabilities which summarize the above situation
   1. *P* (*h*1) = *P* (*cancer*)
   2. *P* (*h*2) = *P* (*¬cancer*)
   3. *P* (+*|h*1)
   4. *P* (*−|h*1)
   5. *P* (+*|h*2)
   6. *P* (*−|h*2)
2. [2/ 5] Suppose we now observe a new patient for whom the lab test returns a positive result,

< d1= + >. What is the *maximum a posteriori (MAP)* hypothesis?

P(cancer|+) = P(+|cancer) x P(cancer) /( P(+|cancer) x P(cancer) + P(+|~cancer) x P(~cancer) )

= .98 \* .008 / ( .98 \* .008 + .03 \* .992 )

= .0078 / ( .0078 + .0298 )

= .21

P(~cancer | +) = .79

Since , *hMAP* = *h2* = ¬ cancer.

1. [0/ 5] With the lab test of the positive result in (2), what is the *maximum likelihood (ML)* hypothesis?

HML = arg max P (D | hi)

P(d1=+ | h1) = .98 > P(d1=+ | h2) = .03, hML = h1

1. [0/ 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, < d2= + >. What are the posterior probabilities of *cancer* (*h*1)and *¬cancer* (*h*2) following these two tests? Assume that the two tests are independent.

C – cancer

2+ = two positive test

P(*cancer* |2+) = P(+|*cancer*) \* P(*cancer* |+) /( P(+|*cancer*) \* P(*cancer* |+) + P(+|~*cancer*) \* P(~C|+) )

P(c|2+) = .98 \* .21 / ( .98 \* .21 + .03 \* .79 )

= .21 / (.21 + .02 )

= .91

=

= <.8953, 1047>

1. [0/ 15] What is your prediction for the 3rd test result (d3), based on the previous two lab test results (*d*1*, d*2)? Compute the predicted probability that the 3rd test result is also positive, i.e. *P* (d3= +*|d*1 = +*, d*2 = +)

3+ = three positive test

P(*cancer* |3+) = P(+|*cancer*) x P(*cancer* |+) \* P(*cancer* |+) /( P(+|*cancer*) \* /( P(+|*cancer*) \* P(*cancer* |+) + P(+|~*cancer*) \* P(~C|+) )

* 1. by Full Bayesian learning

Since , the predictive result of 3rd test is + (positive).

* 1. MAP approximation

In (4), P(h1 | d1,d2) > P(h2|d1,d2), hMAP = h1.

Since P(98 > P( the predictive result of 3rd test is + (positive).

* 1. ML approximation

Since P(d1, d2 |h1)= P(+|h1)P(+|h1) = .982 > P(d1,d2 | h2)= P(+|h1)P(+|h2) = .032 , hML = h1.

So, the result is same as MAP prediction.

i.e. P(98 > P( so the predictive result of 3rd test is + (positive).

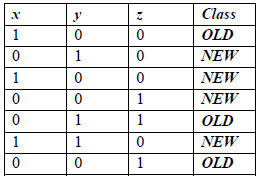
*n*

**Q8. [28/ 40] Naïve Bayesian 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.



1. [5] What is P(*OLD* | *x* = 0) learned for the above training data?

P(*OLD* | *x* = 0) = 1/2

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

P(*OLD* | *x* = 0; *y* = 1; *z* = 0) = 0

1. [10] After learning is complete what would be the predicted probability P(*OLD* | *x* = 0, *y* = 1, *z* = 0)?

P(*OLD* | *x* = 0, *y* = 1, *z* = 0) = 8/35

Prediction probability is = 0.2286

1. [0/ 10] How would a Naïve Bayesian Model predict *Class* given the input < *x*=0, *y*=1, *z*=0 >? Assume that in case of a time the classifier always prefers to predict *OLD* for Class.

In Q8.3, *P(NEW* | *x* =0, *y*=1, *z*=0)*P(OLD* | *x* =0, *y*=1, *z*=0)

Since *P(NEW* | *x* =0, *y*=1, *z*=0) > *P(OLD* | *x* =0, *y*=1, *z*=0), the predicted class *CNB = NEW*.

1. [8/ 10] Using the probabilities obtained during the Bayes classifier training, compute the predicted probability P(*OLD* | x = 0).

P(*OLD* | x = 0) = ½ How? What you’ve computed is equal to (1) which is from the training data, which is 0.5 by counting the # of instances.

You should compute the predictive probability.

and

Thus, < .

Therefore, .