**CSci 384: Artificial Intelligence Spring, 2016**

**Instructor: Dr. M. E. Kim** **Date: May 5th, 2016**

**Due: 3:00 pm, May 12th (Thr.)**

**Final Exam:**

**Total: 181/ 200**

**Instruction:**

**Q4 - Q6.** In order to compute the required probability, you have to define the proper formula and show the 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.

**Q7 (- Q8).** You have to clearly define the formula of information gain or the entropy and show their computational steps. In the full decision tree, the final classification should be specified in each leaf.

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

• You should work on the exam ***independently.*** ***Any kind of plagiarism will be graded as zero point for the final exam.***

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

• Upload it to Submission section by Noon, May 12th , 2016.

• Hours taken to complete the exam: \_8\_\_ Hours \_5\_\_ Minutes.

• Mark the difficulty of the Exam:

Very Easy: \_\_\_\_\_ Easy: \_\_\_\_\_ Moderate: \_\_\_\_\_ Difficult: \_x\_\_\_ Very Difficult: \_\_\_\_\_

**Q1. [10] Propositional Logic.**

Prove the *soundness* or *unsoundness* of the inference rule, *Modus Tollens,* by applying a sequence of logical inference procedures. Specify the inference rule that is applied at each step. – **Do NOT prove it by truth table enumeration. -- it will get NO point.**

Modus Tollens is ((P -> Q)) ∧  Q) -> P

1. ((P Q) ∧  Q) -> P modus ponens
2. ((PQ) ∧  Q)P modus ponens
3. (((P∧ Q) (Q)))P De Morgan’s Law x2
4. ((P∧ Q) Q) P negation
5. (P Q) ∧ ( P Q)P distribution
6. (PP Q) ∧ ( P PQ) distribution
7. (TQ) ∧(TQ) Law of Excluded Middle
8. T∧T Law of Excluded Middle
9. Therefore, Modus Tollens is sound.

**Q2. 26/ [25] 1st-Order Logic (FOL).**

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)) Perfect! +1*

**Q3. 35/ [35]** **Inference in the 1st-Order Logic.**

Consider the following statements.

* (A) No software is guaranteed.
* (B) All programs are software.
* **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 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))* should be congruent to * x (P(x) ∧ G(x))* thus the negation would be * x (P(x) ∧ G(x))*

1. [10] Convert your sentences in (1.A & 1.B) and in (2) to Conjunctive Normal Form (CNF).

*x(S(x)->G(x))* becomes *x(S(x)* *G(x))*

*x(P(x)->S(x))* becomes *x (P(x)* *S(x))*

The negation of the conclusion is * x (P(x) ∧ G(x))* which becomes *P(K) ∧ G(K)* where K is a skolem constant

1. 10/ [10] Using ***resolution***, prove the conclusion is either true or false. Show you proof clearly with the substitution.

If P(~~x~~ K) ∧ G(~~x~~  K) is true, P(~~x~~  K) must be true and G(~~x~~  K) must be true. If that’s the case, for (P(x) S(x)) to be true, S(x) must be true. For (S(x) G(x)) to be true, either G(x) has to be false or S(x) has to be false. Because we’ve clarified both S(x) and G(x) are true, the conclusion must be false. {x/H}

Good to draw the resolution tree of proof. +1

**Q4. [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) is the percentage of people who do not have the disease, or .99

P(b) is the percentage of people who do have the disease, or .01

P(c|b) is the percentage of people who test positive for the disease who have it, or .90

P(c|a) is the percentage of people who give a false positive test for the disease, or .1

P(e) is the percentage of people who test positive for the disease which is P(b)\*P(c)+P(a)\*P(d) or (.01\*.9)+(.99\*.1) = .108 or 10.8%

P(f) is the percentage of people who test negative for the disease which is ((1-P(c))\*P(b))+((1-P(d))\*P(a)) or (.1\*.01)+(.9\*.99) or .892 or 89.2%

1. [10] What percent of people who test positive don’t have the disease (false +)?
   1. P(a|c) = (P(c|a)\*P(a))/P(e) = (.1\*.99)/.108 = .9166667 or 91 and 2/3%
2. [10] What percent of people who test negative do have the disease (false -)?
   1. P(b|f) = ((1-P(c|b))\*P(b))/P(f) = (.1\*.01)/.892 = .00112107 or .01%

**Q5. 25/ [35] Probabilistic Reasoning**

Suppose you want the diagnostic assistant to be able to reason about the possible causes of a patient’s wheezing and coughing.

The agent can observe coughing, sore throat, wheezing, and fever from patients.

The agent can ask whether the patient smokes.

The agent can use Bronchitis and Influenza to predict the outcomes of patients.

The following statistics has been reported and will be used for further inference.

1. The 5% of patients were diagnosed for ***influenza***.
2. The 20% of patients were observed for ***smoking*** a cigarette.
3. Whether patients have ***sore throat*** depends on whether they have ***influenza*** or not.
   1. 30% of patients who have influenza have sore throat while 0.1% of patients who don’t have influenza have sore throat.
4. Whether patients have ***fever*** depends on whether they have ***influenza*** or not.
   1. 90% of patients who have influenza have fever throat while 5% of patients who don’t have influenza have fever.
5. Whether patients have bronchitis depends on whether they have influenza and whether they smoke;
   1. 99% of patient have bronchitis if (s)he has influenza and smokes.
   2. 90% of patient have bronchitis if (s)he has influenza but doesn’t smoke.
   3. 70% of patient have bronchitis if (s)he doesn’t have influenza but smokes.
   4. 0.01% of patient have bronchitis if (s)he neither has influenza nor smokes.
6. Whether patients do ***wheezing*** depends on whether they have ***bronchitis*** or not.
   1. 60% of patients who have bronchitis do wheezing while 0.1% of patients who don’t have bronchitis do wheezing.
7. Whether patients do ***coughing*** depends on whether they have ***bronchitis*** or not.

80% of patients who have bronchitis do coughing while only 7% of patients who don’t have bronchitis do coughing.

1. [15] Using the following Boolean variables, define the above statistics (A) – (G) in the probability formulas with their values: *I* – Influenza, S – Smokes, ST – Sore Throat, F - Fever, B – Bronchitis, C – coughing, W – Wheezing.

e.g.) P(W | I ) = 0.02

P(I) = .05

P(S) = .2

P(ST|I) = .3

P(ST|

P(F|I) = .9

P(F|I) = .05

P(B|I∧S) = .99

P(B|I∧ S) = .9

P(B| I∧ S) = .7

P(B| I∧ S) = .0001

P(W|B) = .6

P(W|B) = .001

P(C|B) = .8

P(C|B) = .07

1. [10] Draw a Bayesian Network which represents the above information correctly and give the (conditional) probability table for each node.

|  |  |
| --- | --- |
|  | P(I) |
| T | .05 |
| F | .95 |

|  |  |
| --- | --- |
|  | P(S) |
| T | .2 |
| F | .8 |

|  |  |
| --- | --- |
| B | P(W|B) |
| T | .6 |
| F | .001 |

|  |  |
| --- | --- |
| I | P(ST|I) |
| T | .3 |
| F | .001 |

|  |  |
| --- | --- |
| I | P(F|I) |
| T | .9 |
| F | .05 |

|  |  |
| --- | --- |
| B | P(C|B) |
| T | .8 |
| F | .07 |

|  |  |  |
| --- | --- | --- |
| I | S | P(B|I,S) |
| T | T | .99 |
| T | F | .9 |
| F | T | .7 |
| F | F | .0001 |

1. 0/ [10] Compute the probability that patients who had wheezing and fever had smoked. – The tense of verbs can be ignored.

P(S|W∧F) =( P(W|B)\*P(B|I,S)(P|I)(P|S)+P(F|I)P(I))\*P(S) = (.6\*.99\*.05\*.2+.9\*.05)\*.2 = .010188

=( P(W|B)\*P(B|I,S)(P|S)+P(F|I))\*P(S) = (.6\*.7\*.2+.05)\*.2 = .0268

=( P(W|B) +P(F|I))\*P(S) = (.001+.05)\*.2 = .0102

=( P(W|B)+P(F|I)P(I))\*P(S) = (.001+.9\*.05)\*.2 = .0092

P(s | f, w)

= α P(s, f, w) = α ∑*i* ∑*st* ∑*b* ∑*c* P(s, f, w, i, st, b, c)   
= α ∑*i* ∑*b* P(s, f, w, i, b) , discarding irrelevant variables.  
= α ∑*i* ∑*b P(i) P(****s****) P(* ***f*** *| i) P(b| i, s) P(****w****| b)*

= α *P(****s****)* ∑*b P(****w****| b)* ∑*i P(b| i,* ***s****) P(* ***f*** *| i)* *P(i)*

= α *P(****s****)* ∑*b P(****w****| b)* ∑*i P(b| i,* ***s****) P(* ***f*** *| i)P(i)*

= α *P(****s****)* ∑*b P(****w****| b) [ P(b| i,* ***s****) P(* ***f*** *| i)P(i) + P(b| ~i,* ***s****) P(* ***f*** *| ~i)P(~i)]*

= α *P(****s****)* ∑*b P(****w****| b) [ P(b| i,* ***s****)\*.9\*.05 + P(b| ~i,* ***s****)\*.05\*.95 ]*

= α *P(****s****) { P(****w****| b) [ P(b| i,* ***s****)\*.9\*.05 + P(b| ~i,* ***s****)\*.05\*.95 ]*

*+ P(****w****| ~b) [ P(~b| i,* ***s****)\*.9\*.05 + P(~b| ~i,* ***s****)\*.05\*.95 ] }*

= α *.2\* { .6\* [ .99\*.9\*.05 + .7\*.05\*.95 ]*

*+ .001 [ .11\*.9\*.05 + .3\*.05\*.95 ] } when s (i.e. S=true)*

*= .009339α*

*Similarly, when ~s, replace s with ~s in the above formula and their corresponding probabilities*

P(~s | f, w)

= α *P(~****s****) { P(****w****| b) [ P(b| i, ~****s****)\*.9\*.05 + P(b| ~i, ~****s****)\*.05\*.95 ]*

*+ P(****w****| ~b) [ P(~b| i, ~****s****)\*.9\*.05 + P(~b| ~i, ~****s****)\*.05\*.95 ] }*

*= .019484α.*

*Normalize <.009339, 019484.> = <.32401207, .675988 > ≈ < .32, .68 >*

**Q6. [24/ 40] Bayesian Network and Inference**

In the given Bayesian Network (BN) below,



*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*

1. [5] Find the nodes which is conditionally independent of ’C’ given its parent ’B’ and ’D’.
   1. Nodes E and A are conditionally independent of C
2. [5] (a) Find the ***Markov blanket*** of a node ‘B’.
   1. A, C, D

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

~~A~~, E, F, G

1. [5] Find the irrelevant node(s) to compute P(C | G).
   1. A
2. 9/ [10] 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.
   1. P(A|B)\*P(B)\*P(C|B,¬D)\*P(¬D| **~** E)\*P(¬E)\*P(¬F|C)\*P(G|¬F) = .88\*.7\*.33\*~~.96\*~~.09\*.55\*.96 = .00927347097
3. [15] Compute the distribution of P(A | g) using ***Variable Elimination***. Show the created factors and the elimination of variables clearly.

**Q7. [23/ 35] Decision Tree Learning**

You’re trying to determine whether Andrew finds a particular type of food ***appealing*** based on the food’s temperature, taste and size.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | *Temperature* | *Taste* | *Size* | *Appealing* |
| X1 | Hot | Salty | Large | Yes |
| X2 | Cold | Sweet | Small | Yes |
| X3 | Cold | Sweet | Small | Yes |
| X4 | Cold | Sour | Large | No |
| X5 | Hot | Sour | Large | No |
| X6 | Hot | Salty | Small | Yes |
| X7 | Hot | Sour | Small | No |
| X8 | Cold | Sweet | Large | No |
| X9 | Cold | Sweet | Large | No |
| X10 | Hot | Salty | Small | Yes |

1. [10] What is the initial entropy of ***Appealing***?
   1. –((5/10 \* log(5/10)) + (5/10 \* log(5/10))) = -(-0.5+-0.5) = -(-1) = 1
2. 0/ [10] Assume that ***Taste*** is chosen as the root of the decision tree. What is the information gain associated with this attribute?

H(Edible) = B<5/11, 6/11> = - (5/11⋅log2 (5/11) +6/11⋅log2 (6/11)) = .994

Remainder(Shape) = 5/11 \* B(<2/5, 3/5>) + 6/11 \* B(<3,/6, 3/6>)

= 0.986796 bits

Remainder(Color) = 5/11 \* B(<2/5, 3/5>) + 4/11 \* B(<2/4, 2/4>) + 1/11 \* 0 + 1/11 \* 0

= 0.804978 bits

Remainder(Odor) = 3/11 \* 0 + 3/11 \* 0 + 5/11 \* B(<3/5, 2/5>)

= 0.441341 bits

So,

Gain(Shape) = .994 - .986796 = .007636 ?

Gain(Color) = .994 - .804978 = .18945

Gain(Odor) = .994 - .441341 = .553

Thus, Odor is chosen as the root of the tree.

1. 8/ [10] Draw the full decision tree with the root ‘***Taste***’, learned for this data. You should show the computation of information gain of attribute to choose a root of each subtree.

Compute the inf. Gains for the nodes to choose a root of subtree, Odor=3, etc. -2

1. [5] Suppose you have the following test data X11 - X15. What is the error rate of test data set from the decision tree in (3)? i.e. the percentage of data that are misclassified by the tree.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | *Temperature* | *Taste* | *Size* | *Appealing* |
| X11 | Hot | Sweet | Small | No |
| X12 | Hot | Salty | Small | Yes |
| X13 | Cold | Sour | Small | No |
| X14 | Hot | Sweet | Large | Yes |
| X15 | Cold | Salty | Large | No |

X11 is misclassified

X12 is misclassified

X13 is correct

X14 is misclassified

X15 is correct

So, 3/5 are misclassified. There is a 60% error rate.

**Q8. 18/ [Optional, 20] Decision Tree Learning**

The following dataset will be used to learn a decision tree for predicting whether a strawberry is edible or not based on its shape, color and odor.

|  |  |  |  |
| --- | --- | --- | --- |
| *Shape* | *Color* | *Odor* | ***Edible*** |
| A | R | 1 | No |
| A | P | 1 | No |
| B | P | 1 | No |
| A | R | 2 | Yes |
| B | R | 2 | Yes |
| B | P | 2 | Yes |
| B | P | 3 | Yes |
| A | R | 3 | Yes |
| B | R | 3 | No |
| B | G | 3 | No |
| A | V | 3 | No |

1. [5] What is the entropy H(*Edible* | *Order* = 1 or *Odor* = 2)?

-(1/2log(1/2)+1/2log(1/2)) = 1

1. 3/ [5] Which attribute would the decision tree algorithm choose for the root of the tree?
   1. Odor. Consistently, 1 is not edible and 2 is edible.

No computation of inf. Gains for Odor | Color | Size ? -2

1. [5] Draw the full decision tree that would be learned for this data.
2. [5] Suppose we have a test data set as follows. What will be the error rate of the test data set from the decision tree in (3)? i.e. the percentage of examples that would be misclassified by the tree.

|  |  |  |  |
| --- | --- | --- | --- |
| *Shape* | *Color* | *Odor* | ***Edible*** |
| A | R | 3 | Yes |
| B | R | 3 | Yes |
| A | P | 3 | No |

1 is incorrect

2 is incorrect

3 is correct

The error rate is 67% or 2/3.