B. Tech Fourth Year End Semester Examination

Department: Computer Science and Engineering

Course Name: Artificial Intelligence **Code: CS 461**

Full Marks-100 **Time: 3 hours**

Answer TWO and THREE questions from Group-A and Group-B, respectively.

Make reasonable assumptions as and whenever necessary. You can answer the questions in any sequence. However, the answers of any particular question should appear together.

Group-A

1. (a). What are the abilities of a knowledge based system? Define the terms *entailment*, *inference*, *sound algorithm* and *complete algorithm* with respect to the logical reasoning.

   (b). Why is propositional logic monotonic? Define resolution theorem in propositional logic. Consider the following propositional logic knowledge base:

   $KB = (B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})) \wedge \neg B_{1,1}$

   Use resolution algorithm to show whether $\alpha = \neg P_{1,2}$ can be concluded or not.

   (c). What is meant by horn clause?

(2+2*4) + (2+2+4) +2

2. (a). For propositional logic, why is entailment with horn clauses more efficient? Consider the following propositional logic knowledge base:

$$
\begin{aligned}
&P \Rightarrow Q\\
&L \wedge M \Rightarrow P\\
&B \wedge L \Rightarrow M\\
&A \wedge P \Rightarrow L\\
&A \wedge B \Rightarrow L\\
&A\\
&B
\end{aligned}
$$

   Show the various steps of backward chain resolution algorithm to derive the query Q (A and B are the facts). Why is backward chain more effective than forward chain?

   (b). Distinguish between propositional logic and first-order logic. Consider the following knowledge base:

   The law says that it is a crime for an American to sell weapons to hostile nations. The country Nono, an enemy of America, has some missiles, and all of its missiles were sold to it by Colonel West, who is American.

Prove using first-order logic forward chain resolution algorithm that *Col. West is a criminal.* Show the detailed knowledge base at each step.

(2+ 6) + (2+10)

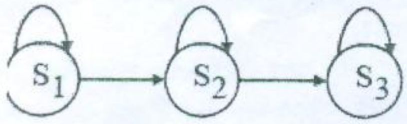
3. (a). Find out all possible unifiers for the following first-order predicate expressions (*x, y, z*: variables; *John*: constant and *Knows*: predicate function):

Knows (John, x) and *Knows(y, z)*

Which one is most general unifier?

(b). Define resolution theorem for first-order predicate logic. Generate the first-order predicate logic expression for the statement "*Everyone who loves all animals is loved by someone*". Convert the expression to conjunctive normal form.

(c). Consider the problem of character recognition as a sequence labelling task under the hidden Markov model. The goal is to recognize an image given a character. The structure of the hidden states is as shown below:



The transition and observation probabilities for character images A and B are shown below:

For A:

Transition probabilities: {{ 0.8, 0.2, 0.0}, {0.0, 0.8, 0.2}, {0.0, 0. 0, 0.1} }

Observation probabilities: { { 0.9, 0.2, 0.0}, { 0.1, 0.8, 0.1}, {0.9, 0.1, 0.0}}

For B:

Transition probabilities: { { 0.8, 0.2, 0.0 }, {0.0, 0.8, 0.2}, { 0.0, 0. 0, 0.1}}

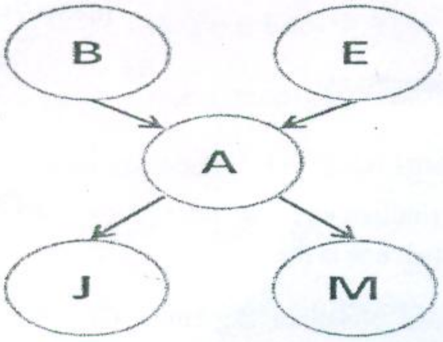
Observation probabilities: { 0.9, 0.1, 0.0}, { 0.0, 0.2, 0.1}, {0.6, 0.4, 0.0}}

Which character model is more suitable for the observation sequence {1,1,2,3}?

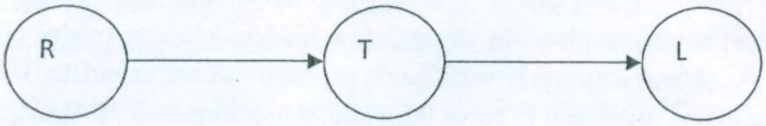
3 + (2+8) +10

Group-B

1. (a). Formulate probabilistic inference. For the following Bayesian network, show all the atomic probabilities required to compute P (+b|+j, +m) following probabilistic inference by enumeration (evidences corresponding to the nodes in the network are a, b, e, j, m for A, B, E, J and M, respectively; all the evidence variables have both positive and negative values). Why is probabilistic inference by enumeration slow?



(b). Consider the following Bayesian network. R, T and L are three random variables representing the events *Rain, Traffic* and *Late for class*.



P(+r)=0.1, P(+t|+r)=0.8, P(+t| -r)=0.1, P(+l|+t)=0.3, P(+l|-t)=0.1, where r, t and l (both positive and negative) denote the evidences of the random variables. Compute P (L| +r).

(c). For the causal chain of figure 4(b), is R independent of L given T?

(2+4+2)+10+2

2. (a). Define inductive learning. Construct a decision tree for the function $f(A, B)=AB + A'B'$, where A and B are the boolean variables and A' and B' are the complements of A and B, respectively. Describe the information gain metric with respect to the choice of attributes in decision tree.

   (b). Distinguish between supervised and unsupervised machine learning algorithms using proper examples. Derive naive Bayes classifier for text classification under conditional independence assumption. What are the disadvantages of maximum likelihood estimates (frequency counts) for naive Bayes classification? How can the problem be eliminated?

(2+4+3) + (2+5+2+2)

3. (a). Show how *learning* and *classification* in text classification are done using multinomial naive Bayes Classifier? Show the training and testing time complexities.

   (b). How document categorization can be done using vector space model? (***hints***: describe in terms of vector representations, term weights (i.e., tf-idf) and cosine similarity measurements).

   (c). Given a document, containing terms with given frequencies: A(3), B(2), C(1). Assume that collection contains 10,000 documents and document frequencies of these terms are: A (50), B (1300), C (250).

   Compute the term *frequency-inverse document frequency* of the document collection.

(6+3)+7+4

4. (a). Consider the weighted term vectors of two documents as:

$D_1 = 2T_1 + 3T_2 + 5T_3 \qquad D_2 = 3T_1 + 7T_2 + 1T_3$

For a query, $Q = 5T_1 + 5T_2 + 2T_3$, compute the similarities using *inner product* and *cosine similarity* metrics. With respect to this problem, which one is the better measurement?

(b). Describe the working principles of K Nearest Neighbor algorithm with respect to document classification. Show its training and testing time complexities. Mention about the shortcomings of this approach and their possible remedies.

(c). Describe the relevance feedback architecture of information retrieval.

(4+1)+(4+3+2+3)+3

5. (a). What are the key issues involved in designing a hidden Markov model (HMM) for sequence labeling tasks (***hints***: mention how different probabilities are calculated)? Formulate forward-backward recursion algorithm to solve the evaluation problem of HMM.

(b). What is meant by part of speech (PoS) tagging? Show how PoS tagging problem can be solved using a second-order HMM without using any additional context during emission probabilities (***hints***: consider the problem of assigning n tags to the sequence of n words; clearly formulate the problem and show how the various issues like sparse data and decoding problems can be addressed).

(3+6) + (1+10)