Answer THREE questions, at least ONE from EACH of Sections A and B.

SECTION A

1. a. Give your answers to (i)-(iv) briefly.

   i. Give three reasons for why classical logic is appealing for knowledge representation and reasoning.

   ii. Give two problems with the expressibility of classical logic for knowledge representation and reasoning.

   iii. In a multi-agent system, what is common knowledge?

   iv. In a multi-agent system, what is distributed knowledge?

   [8 marks]

   b. i. Give five basic requirements for a truth maintenance system.

   ii. Give a formalization of a truth maintenance system in terms of logical arguments. First define an argument and then explain how each of the five basic requirements given in the previous part of this question is met.

   [15 marks]

   c. Explain how you could use a truth maintenance system as the basis of a diagnostic expert system. Aim to make your explanation as unambiguous and clear as possible.

   [10 marks]

   [Total 33 marks]

2. a. i. Give two reasons why Dempster-Shafer is a useful technique for handling uncertain information.

ii. What is a frame of discernment?

iii. Define a basic probability assignment.

iv. Define a belief function.

v. What is the role of normalization in Dempster's rule of combination and why is it problematical?

[15 marks]

b. Let Ω be a frame of discernment, let m be a basic probability assignment, and let A be a subset of Ω such that $m(A) = 1$.

i. Suppose $\alpha \in A$, what is the value of $m(\{\alpha\})$?

ii. What is the value of $m(\{\alpha\}^c)$?

[6 marks]

c. Let Bel be a belief function. Does the following hold?

$$Bel(A) + Bel(A^c) = 1$$

For your answer, give a proof based on the definition of a basic probability assignment.

[12 marks]

[Total 33 marks]

3. a. i. What is a Bayesian network?

ii. Why is a Bayesian network of particular importance in developing probabilistic reasoning systems?

iii. What is a moral graph?

[9 marks]

b. Let α,β, and γ be random variables. Briefly explain each of the following four basic patterns of relevance between α and β.

i. Marginal and conditional independence.

ii. Marginal and conditional dependence.

iii. Marginal independence and conditional dependence.

iv. Marginal dependence and conditional independence.

[12 marks]

c. Suppose we have the following Bayesian network for the random variables α,β,γ, and δ.

$$\alpha \leftarrow \beta \leftarrow \gamma \leftarrow \delta$$

i. Suppose each of the random variables α,β,γ, and δ, takes a value either true or false. How many joint probability statements do we require to provide a total distribution.

ii. Consider one joint probability statement for this. Rewrite this statement as a product of conditional probability statements using the multiplication theorem.

iii. Simplify the answer to (ii) using the information in the Bayesian network.

[12 marks]

[Total 33 marks]

SECTION B

4. a. Let $\mathbf{s}$ be a training sequence,

$$\mathbf{s} = ((\mathbf{x}_1, o_1), (\mathbf{x}_2, o_2), \ldots, (\mathbf{x}_m, o_m))$$

in which each o_i is a class label taken from a set $\{c_1, c_2, \ldots, c_k\}$ of k possibilities.

i. Define the *entropy* of $\mathbf{s}$.

[4 marks]

ii. If $\mathbf{s}$ is divided into p sequences $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_p$ such that each element of $\mathbf{s}$ appears in exactly one of the p sequences, define the resulting *gain*.

[4 marks]

iii. Briefly explain why the gain can be a problematic quantity in the construction of a decision tree, and describe a method for overcoming this.

[5 marks]

b. Explain how decision trees can deal with missing attributes, both during the construction of the tree, and when classifying new examples.

[20 marks]

[Total 33 marks]

5. a. Let $\mathcal{H}$ be a space of hypotheses $h : X \to \{0, 1\}$. Define the *growth function* $\Pi_{\mathcal{H}}(m)$ and the *Vapnik-Chervonenkis dimension* $\text{VCdim}(\mathcal{H})$ for $\mathcal{H}$.

[8 marks]

b. Give a detailed definition, including definitions of all relevant terms, of what it means to say that $\mathcal{H}$ is *potentially learnable*.

[12 marks]

c. Prove that if $\text{VCdim}(\mathcal{H}) = \infty$ then $\mathcal{H}$ is not potentially learnable.

[13 marks]

[Total 33 marks]

CONTINUED

6. Write an essay, including as much technical detail as you can and illustrating your answer with examples, on the subject of *genetic algorithms* and *genetic programming* as tools for machine learning.

[Total 33 marks]

END OF PAPER