RS228/406

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc (Hons) in Computer Science

Stage 4

SUPPLEMENTAL EXAMINATIONS 2009

ARTIFICIAL INTELLIGENCE 2

Dr. John Kelleher
Prof. B. O'Shea
Dr. I. Arana

Duration: 2 Hours

Answer Question 1 (40 marks) **and**

any 2 Other Questions (30 marks each).

1. (a) In the context of machine learning, explain what is meant by **overfitting** the training data.

(5 marks)

(b) In the context of inductive learning explain what is meant by a **consistent hypothesis**.

(5 marks)

(c) Distinguish between **classification learning** and **regression learning**.

(5 marks)

(d) Briefly describe how a Decision Tree works.

(5 marks)

(e) In the context of inductive logic learning, what is meant by the **extension** of a hypothesis?

(5 marks)

(f) Distinguish between the **generalisation** and **specialization** of a logical predicate.

(5 marks)

(g) In the context of machine learning distinguish between **false negatives** and **false positives**.

(5 marks)

(h) Construct by hand a neural network that computes the XOR function of two inputs. Make sure to specify what sort of units you are using.

(5 marks)

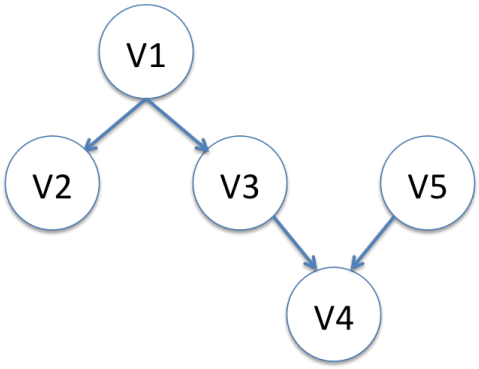

Figure 1: An example Bayesian network.

2. (a) Given that $P(a|b) = 0.5$, $P(a) = 0.3$, $P(b) = 0.4$ calculate $P(b|a)$.

(5 marks)

(b) Express the joint probability distribution for the network in Figure 1 using the chain rule.

(10 marks)

(c) (i) In the context of inference in probabilistic temporal models define, in your own words. what is meant by the term **filtering**.

(5 marks)

(ii) Provide the equation that defines a recursive formulation of filter estimation and explain its components.

(10 marks)

3. (a) Explain what is meant by **inductive learning**.

(5 marks)

(b) Suppose we generate a training set from a decision tree and then apply decision-tree learning to the training set. Is it the case that the learning algorithm will eventually return the correct tree as the training set size goes to infinity? Why or why not?

(5 marks)

(c) The following sets express the mappings between predicates r, p, q, s, $class1$ and $class2$: $r \rightarrow \{a1, a2, a5, a6\}$, $p \rightarrow \{a2, a3, a5, a7\}$, $q \rightarrow \{a1, a2, a6\}$, $s \rightarrow \{(a2, f), (a1, 1), (a6, f)\}$, $class1 \rightarrow \{a2\}$, $class2 \rightarrow \{a2, a6\}$.

(i) Given the above sets give a specialisation of the rule $class1(X) \leftarrow r(X) \wedge p(X)$ such that the rule is only satisfied by $class1$ members.

(5 marks)

(ii) Given the above sets give a rule that will correctly classify only members of $class2$.

(5 marks)

(d) The FOIL inductive logic programming algorithm is considering adding a literal l to a rule r. The extension of the rule r before adding l is the following set of positive and negative examples $\{+, +, +, +, +, +, -, -, -, -\}$. The extension of the rule after the literal is added (i.e. the extension of $r + l$) is $\{+, +, +, +, +, +, -, -\}$. What is the information gain of adding the literal l to the rule r? (Note you do not need to calculate the logs in your answer ,i.e. you can express your answer as an equation containing logs).

(10 marks)

4. (a) Explain why linear threshold perceptrons can represent the AND and OR functions but not the XOR function.

(10 marks)

(b) Describe the perceptron training rule?

(10 marks)

(c) Why does the perceptron learning rule converge toward successful weight values?

(10 marks)