CS 188 | Introduction to AI
Fall 1993 | Stuart Russell

Final examination

You have 2 hours 50 min. The exam is open-book, open-notes.
There a total of 100 points available.
Write your answers in blue books. Hand them all in.
Several of the questions on this exam are true/false or multiple choice.
In all the multiple choice questions more than one of the choices may be correct. Give all correct answers. Each multiple choice question will be graded as if it consisted of a set of true/false questions, one for each possible answer.

1. (10 pts.) Definitions Provide brief, *precise* definitions of the following:

(a) Decision-theoretic agent

(b) Intrinsic property

(c) Clause (in CNF)

(d) Conditional probability

(e) Optical flow

2. (10 pts.) Knowledge representation
Represent the following sentences in first-order logic, *using one consistent ontology.* For each predicate, function and constant symbol you use, say what it means in English.

(a) (2) Water is a liquid and Ty Nant is a kind of water.

(b) (3) Each of Earth's oceans contains some water.

(c) (2) A pint of water weighs a pound.

(d) (3) Water has a boiling point. (*Hint: don't say HasBoilingPoint — think what happens to a piece of water above a certain temperature.*)

3. (10 pts.) Logical Inference
Multiple choice: Given the following premises:

$$\begin{array}{l} \forall z \; \neg Q(C, z) \\ \forall x, y \; P(f(x), f(y)) \Rightarrow P(f(y), f(x)) \\ \forall y \; Q(y, A) \Rightarrow P(y, y) \\ Q(A, A) \\ \forall y \; Q(f(B), y) \vee Q(C, y) \end{array}$$

which of the following conclusions is logically justified?

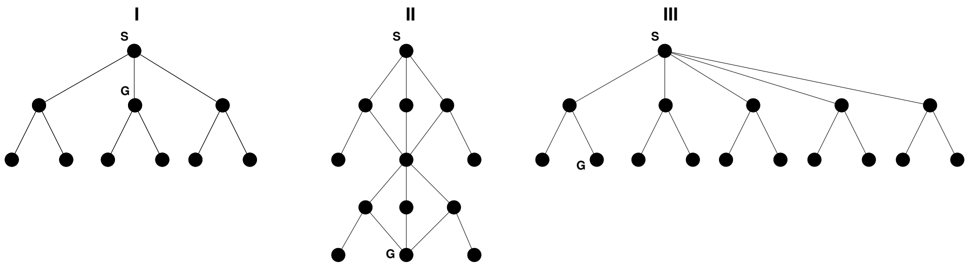
(a) (1) $P(A, A)$

(b) (2) $P(f(A), f(A))$

(c) (3) $P(f(B), f(B))$

(d) (4) The empty clause (i.e., the KB is inconsistent).

4. **(12 pts.) Uninformed search**
Consider the search spaces shown in the following figure, in each, G marks a goal state. Successors are generated left-to-right, and each arc is unidirectional (downward): For each search space (I, II, III), list the MOST efficient



and LEAST efficient algorithm in terms of absolute computation time (NOT O() time).

(a) Breadth-first
(b) Depth-first
(c) Depth-first iterative deepening
(d) Depth-first with repeated-state checking

You may assume that the time for a repeated-state check is small compared to the time to generate a set of successors.

5. **(11 pts.) Heuristic search**
In this question we will consider the problem of colouring an undirected graph (a graph consists of a set of vertices V joined by edges E). Two vertices are adjacent if there is an edge linking them. A colouring of a graph assigns a colour to each vertex. It is valid if no two adjacent vertices are the same colour. In this case the objective is to find the minimal valid colouring — the one using the least number of different colours.

(a) (4) Define this problem formally as a search problem suitable for solution by a standard search algorithm such as breadth-first or A*.
(b) (2) Give a non-trivial, admissible heuristic function for this problem that runs in at most linear time in the size of the graph (that is, $O(|V| + |E|)$).
(c) (2) Show that your heuristic is admissible.
(d) (3) Explain how you would set the problem up for simulated annealing. Does your energy function have local minima?

6. **(12 pts.) Probability and decision theory**
Let the current world state be denoted by a random variable W_0, and the current percept be S_0. Let A be a decision variable whose value is the action the agent will do in the current state, and let S_1 and W_1 be the resulting percept and world state. Let U be the utility of the resulting state. Assume initially that the sensor's operation is unaffected by any action the agent may take.

(a) (3) Draw the influence diagram for this decision problem.
(b) (4) Write the name (i.e., $\mathbf{P}(\ldots|\ldots)$) of each distinct conditional probability distribution required for this influence diagram (note that two of the five will usually be identical). Explain in words what each distribution represents.
(c) (3) Suppose that the world is fully accessible. What does this tell you about the conditional probability distribution at the nodes S_0 and S_1?
(d) (2) Suppose that for some values of A the sensor is likely to become damaged. Explain, in words or by redrawing, how the influence diagram in (a) should be changed.

7. **(12+3 pts.) Neural networks**
In this question we will consider neural nets with inputs in the range $[0, 1]$ and with g a step function. A network is defined by the weights on the links and the threshold value of g at each node.

(a) (2) Draw a network to represent the majority function (at least half the inputs high) for 4 input nodes.
(b) (3) Draw a network to represent the "exactly two out of three" function for three inputs.

(c) (3) Describe *why* and *how* you might apply simulated annealing to train a neural network.

(d) (3) Suppose you are training a neural network in a genuinely nondeterministic domain. You give it 100 copies of the same example, 75 of which are positive and 25 of which are negative. Using the standard error function

$$E = \sum_e \frac{1}{2}(T^e - O^e)^2$$

where the sum is taken over the examples in the training set and where T^e is the correct value for the example and O^e is the actual output, calculate the error when the network converges to output 1.0 and output 0.75 respectively. Comment on the result. For what sort of classification task might an output of 1.0 be appropriate given this data?

(e) (3) (Extra credit) Can you design an error function such that the network will output a suitable probability value when trained on data of this sort?

8. (15 pts.) Natural language

Consider the following context-free grammar:

S →NP VP
S →if S then S
NP →Determiner Modifier Noun | Pronoun
Determiner →a | the | three
Pronoun →I | you | me | he | him
Modifier →Adjective* | Noun*
Adjective →large | small
Noun →tennis | racquet | head | kindergarten | table
VP →Verb NP
Verb →clean | break

(a) (3) *Multiple choice*: Which of the following sentences are generated by the grammar?

i. if if you break me then I break you then I clean the small tennis racquet
ii. the large tennis tennis clean me me me
iii. the tennis kindergarten large head clean you

(b) (2) Write down at least one other *English* sentence generated by the grammar above. It should be significantly different from the above sentences, and should be at least six words long. Do not use any of the words from the above sentences; instead, add grammatical rules of your own, of the form (grammatical category) →(specific word)—for instance, Noun → bottle.

(c) (2) Show the parse tree for your sentence.

(d) (2) Fix the rule "S →if S then S" so that it disallows the generation of consecutive ifs but still allows nested conditionals. (Hint: think about what the "condition" part of a conditional sentence can be.)

(e) (4) (open-ended) This grammar allows "noun-noun modification" (e.g., "tennis racquet" or "table tennis"). Not all noun strings in English are considered grammatical; for example, you can say "tennis racquet shop" but not "electron dictionary elbow." Is this a syntactic or semantic distinction? How might it be handled by a definite clause grammar?

(f) (2) In German, noun-noun pairs (triplets etc.) are run together into one word (this also occurs in English very occasionally, as in "bathroom"). Would you need to change a natural language processing system to handle this? If so, how?

9. (8 pts.) Vision/NLP

(Open-ended) What do vision and spoken language understanding[1] have in common? Consider both low-level and high-level aspects of each task, and try to construct a detailed analogy. Are there fundamental differences between the two tasks, beyond the physical differences in the two kinds of signals?

[1] "Understanding" means understanding the content not just recognizing the words