CS 188 Fall 1997 — Introduction to AI — 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. (12 pts.) True/False
Decide if each of the following is true or false. If you are not sure, you may wish to provide some explanation to follow your answer.

(a) (2) In an MDP, the optimal action can be selected by an agent based only on the current percept (given a fixed policy).

(b) (2) Neural networks can be set up to output only 0/1 values.

(c) (2) α-β search keeps all explored nodes in memory.

(d) (2) Search is a special case of planning.

(e) (2) $(\forall x\, P(x)) \vee (\forall x\, \neg P(x))$ is a valid sentence.

(f) (2) Decision trees with k internal nodes can express any Boolean function of k Boolean attributes.

2. (10 pts.) Logic *True/false*:

(a) (2) The only unsatisfiable clause is the empty clause.

(b) (2) If C_1 and C_2 are clauses in propositional logic, and all the literals in C_1 are contained in C_2, then $C_1 \models C_2$.

(c) (3) If C_1 and C_2 are clauses in propositional logic, and $C_1 \models C_2$, then all the literals in C_1 are contained in C_2.

(d) (3) "There is a tortoise who is older than any human" is a good translation of
$\exists t\, Tortoise(t) \Rightarrow [\forall h\, Human(h) \wedge Age(t) > Age(h)]$

3. (5 pts.) Logic contd.
Let C_2 be the clause $Q(f(a,a))$ where a is a constant symbol. We are interested in finding a clause C_1 such that $C_1 \models C_2$; for example, we have $C_1 = Q(x)$ or $C_1 = Q(f(a,a))$. List *all* the other *distinct* possibilities for C_1. [Note: two clauses are distinct from each other if they are not logically equivalent.]

4. (18+5 pts.) Neural networks

(a) (3) Consider the class of neural nets with inputs $x_1 \ldots x_n$, which are either 0 or 1. A network is specified by giving the weights on the links and the activation function g at each node. Specify a network that computes the *majority* function for 5 input nodes. That is, it should output 1 if at least half the inputs are high, and 0 otherwise.

(b) (3) Draw a decision tree that represents the *disjunction* of five inputs.

(c) (4) Describe *why* and *how* you might apply simulated annealing to train a neural network.

(d) (4) Suppose you are training a neural network in a genuinely nondeterministic domain. The training set consists of N copies of the same example, a fraction $p > 0.5$ of which are positive and a fraction $1-p$ of which are negative. Suppose we use the "absolute error" function:

$$E = \sum_e |T^e - O^e|$$

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 network's output. Suppose also that O_e must be in the range [0,1]. By writing out an expression for the error in terms of O_e, find the value of O_e that minimizes the error.

(e) (4, RELATIVELY HARD) Let us try to find an error function such that the error in the above situation is minimized when $O^e = p$, i.e., we want the trained network to output the probability that the example is positive. Suppose that we try the following kind of error function:

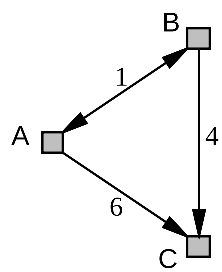
$$E = \sum_e (T^e - O^e)^n$$

for some n. Show that $n = 2$ is the only possibility. [Reminder: $\frac{d(u^n)}{dx} = n\frac{du}{dx}u^{n-1}$]

(f) (5, EXTRA CREDIT) Analyze the case where we allow the error any polynomial function of $(T^e - O^e)$.

5. (20+5 pts.) Search and MDPs

(a) (3) The following diagram represents the state space of a deterministic problem, with each arrow denoting a possible operator (labelled with the step cost). Assume that the successors of a state are generated in alphabetical order, and that there is no repeated-state checking. Show the search tree generated by breadth-first search applied to the problem of starting in A, where C is the goal. Circle the tree node that the search identifies as the solution.



(b) (2) Of the four algorithms breadth-first, depth-first, iterative-deepening, and uniform-cost search, which find a solution in this case?

(c) (2) Of the four algorithms, which find an *optimal* solution in this case?

(d) (3) In MDPs, the values of states are related by the following equation, the Bellman equation (p.503):

$$U(i) = R(i) + \max_a \sum_j M^a_{ij} U(j)$$

where $R(i)$ is the reward associated with being in state i and M^a_{ij} is the probability of reaching state j if action a is executed in state i. Suppose now we wish to associate rewards with *actions* instead of states; i.e., $R(a,i)$ is the reward for doing a in state i. How should the Bellman equation be rewritten to use $R(a,i)$ instead of $R(i)$?

(e) (5) Can any finite search problem be translated exactly into a Markov decision problem, such that an optimal solution of the latter is also an optimal solution of the former? If so, explain *precisely* how to to translate the problem AND how to translate the solution back; if not, explain *precisely* why not (e.g., give a counterexample).

(f) (5) In this part we will apply the value iteration algorithm to the MDP that corresponds to the above search problem. Assume that each state has an initial value estimate of 0. Copy and complete the following table, showing the value of each state after each iteration and the optimal action choice given those values. Continue the process until it converges.

State	**Iteration**							
	0	1	2	3	4	5	6	...
A	$0, \rightarrow B$							
B	$0, \rightarrow A$							
C	0							

(g) (5, EXTRA CREDIT) Does policy iteration work on this problem? Discuss.

6. (16 pts.) Natural language
Consider the following context-free grammar:

S →NP VP
S →either S or S
NP →Determiner Modifier Noun | Pronoun | ProperNoun
Determiner →a | the | every
Pronoun →she | he | it | him | her
Modifier →Adjective* | Noun*
Adjective →red | green | smelly
Noun →golf | club | green
VP →Verb NP
VP →Copula Adjective
Verb →cleans | breaks | is
Copula →is | seems
ProperNoun →Bill

(a) (6) *Multiple choice*: Which of the following sentences are generated by the grammar?

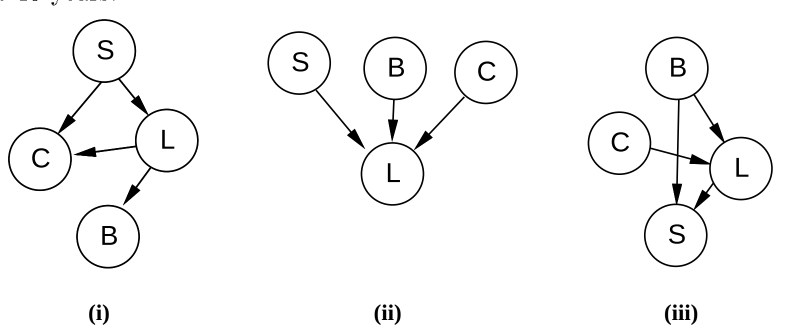
i. either either Bill is smelly or Bill is green or the green is red
ii. every green green green is a green green
iii. the golf club breaks smelly Bill

(b) (4) Show the parse tree for the sentence
"Either the green golf club is smelly or it is red"

(c) (2) How many possible parsings are there for the sentence
"The green green green cleans a green green"? What type of ambiguity is causing this multiplicity?

(d) (4) In English, one can say "either A or B or C or D ..." whereas nested constructions such as "either either A or B or C" and "either A or either B or C" are not usually allowed. Show how to replace the rule "S →either S or S" by one or more new rules to reflect this.

7. (19 pts.) Probabilistic inference
Consider the belief networks shown in the following figure, where $S = Smoking$, $L = LungCancer$, $C = Cough$, $B = BiopsyTest$. ($BiopsyTest$ is true iff the test result is positive for cancer.) All variables are Boolean, none of the CPTs are deterministic, and the population consists only 60-year-olds who are non-smokers or who have smoked for the last 40 years.



(a) (3) *Multiple choice*: Which of the networks are *correct*, given current medical knowledge?

(b) (2) Which network has the fewest parameters?

(c) (3) Write down some reasonable CPT values for node C in network (i).

(d) (3) Using network (i), derive a symbolic expression for $P(B|S)$ in terms of conditional probabilities available in the CPTs, by conditioning on one or more other variables. (Do not plug in any numbers.)

(e) (4) Using network (i), derive a symbolic expression for $P(L|B)$ in terms of conditional probabilities available in the CPTs. (You will need Bayes' Rule as well as conditioning here.)

(f) (4) Suppose that, in a polytree network, X is an ancestor of Y and we wish to compute $P(Y|X)$ in terms of the CPT entries of the network. Explain why conditioning operations suffice for this computation and describe *in general terms* the minimal set of conditioning operations that must be carried out. [Hint: you may find it useful to consider an example such as the following, although the specific details of the example need not appear in your answer:]