CS 188 Fall 2005 — Introduction to AI — Stuart Russell — Final

You have 2 hours and 50 minutes. The exam is open-book, open-notes. 100 points total. Panic not.

Mark your answers ON THE EXAM ITSELF. Write your name, SID, and section number at the top of each page.
For true/false questions, CIRCLE *True* OR *False*.
For multiple-choice questions, CIRCLE *ALL* CORRECT CHOICES (in some cases, there may be more than one).

If you are not sure of your answer you may wish to provide a *brief* explanation.

For official use only

Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7	Total
/12	/15	/16	/15	/8	/16	/18	/100

1. (12 pts.) Some Easy Questions to Start With

(a) (2) *True/False*: Suppose that variables $X_1, X_2, \ldots, X_k$ have no parents in a given Bayes net that contains n variables in all, where $n > k$. Then the Bayes net asserts that $\mathbf{P}(X_1, X_2, \ldots, X_k) = \mathbf{P}(X_1)\mathbf{P}(X_2)\cdots\mathbf{P}(X_k)$.

(b) (2) *True/False*: In a fully observable, turn-taking, zero-sum game between two perfectly rational players, it does not help the first player to know what move the second player will make.

(c) (2) *True/False*: For any propositional sentences α, β, γ, if at least one of $[\alpha \models \gamma]$ and $[\beta \models \gamma]$ holds then $(\alpha \wedge \beta) \models \gamma$

(d) (2) *True/False*: For any propositional sentences α, β, γ, if $(\alpha \wedge \beta) \models \gamma$ then at least one of $[\alpha \models \gamma]$ and $[\beta \models \gamma]$ holds

(e) (2) *True/False*: If C_1 and C_2 are two first-order clauses that have been standardized apart, then no literal in C_1 is identical to any literal in C_2.

(f) (2) *True/False*: New lexical categories (such as *Noun*, *Verb*, *Preposition*, etc.) are frequently added to English.

2. (15 pts.) Search

n vehicles occupy squares $(1,1)$ through $(n,1)$ (i.e., the bottom row) of an $n \times n$ grid. The vehicles must be moved to the top row but in reverse order; so the vehicle i that starts in $(i,1)$ must end up in $(n-i+1,n)$. On each time step, every one of the n vehicles can move one square up, down, left, or right, or stay put; but if a vehicle stays put, one other adjacent vehicle (but not more than one) can hop over it. Two vehicles cannot occupy the same square.

(a) (4) The size of the state space is roughly
(i) n^2 (ii) n^3 (iii) n^{2n} (iv) n^{n^2}

(b) (3) The branching factor is roughly
(i) 5 (ii) $5n$ (iii) 5^n

(c) (2) Suppose that vehicle i is at (x_i, y_i); write a nontrivial admissible heuristic h_i for the number of moves it will require to get to its goal location $(n-i+1,n)$, assuming there are no other vehicles on the grid.

(d) (2) Which of the following heuristics are admissible for the problem of moving all n vehicles to their destinations?
(i) $\sum_{i=1}^{n} h_i$ (ii) $\max\{h_1, \ldots, h_n\}$ (iii) $\min\{h_1, \ldots, h_n\}$ (iv) None of these

(e) (4) Explain your answer to part (d).

3. (16 pts.) Propositional Logic

Suppose an agent inhabits a world with two states, S and $\neg S$, can do exactly one of two actions, a and b. Action a does nothing and action b flips from one state to the other. Let S^t be the proposition that the agent is in state S at time t, and let a^t be the proposition that the agent does action a at time t (similarly for b^t).

(a) (4) Write a successor-state axiom for S^{t+1}

(b) (4) Convert the sentence in (a) into CNF.

(c) (8) Show a resolution refutation proof that if the agent is in $\neg S$ at time t and does a it will still be in $\neg S$ at time $t+1$.

4. (15 pts.) Pruning in search trees

In the following, a "max" tree consists only of max nodes, whereas an "expectimax" tree consists of a max node at the root with alternating layers of chance and max nodes. At chance nodes, all outcome probabilities are non-zero. The goal is to *find the value of the root* with a bounded-depth search.

(a) (2) Assuming that leaf values are finite but unbounded, is pruning (as in alpha-beta) ever possible in a max tree? Give an example, or explain why not.

(b) (2) Is pruning ever possible in an expectimax tree under the same conditions? Give an example, or explain why not.

(c) (2) If leaf values are constrained to be nonnegative, is pruning ever possible in a max tree? Give an example, or explain why not.

(d) (2) If leaf values are constrained to be nonnegative, is pruning ever possible in an expectimax tree? Give an example, or explain why not.

(e) (2) If leaf values are constrained to be in the range $[0, 1]$, is pruning ever possible in a max tree? Give an example, or explain why not.

(f) (2) If leaf values are constrained to be in the range $[0, 1]$, is pruning ever possible in an expectimax tree? Give an example (qualitatively different from your example in (e), if any), or explain why not.

(g) (3) Consider the the outcomes of a chance node in an expectimax tree. Which of the following evaluation orders is most likely to yield pruning opportunities?

(i) Lowest probability first (ii) Highest probability first (iii) Doesn't make any difference

5. (8 pts.) MDPs

Consider the world in Q.3 as an MDP, with $R(S) = 3$, $R(\neg S) = 2$, and $\gamma = 0.5$. Complete the columns of the following table to perform policy iteration to find the optimal policy.

	π^0	V^{π^0}	π^1	V^{π^1}	π^2
S	a				
$\neg S$	a				

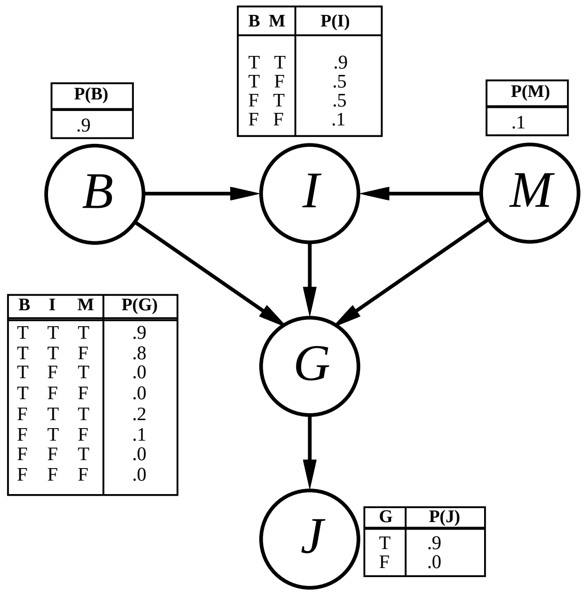

Fig. 1: A simple Bayes net with Boolean variables $B = BrokeElectionLaw$**,** $I = Indicted$**,** $M = PoliticallyMotivatedProsecutor$**,** $G = FoundGuilty$**,** $J = Jailed$**.**

6. (16 pts.) Probabilistic inference

Consider the Bayes net shown in Fig. 1.

(a) (3) Which, if any, of the following are asserted by the network *structure* (ignoring the CPTs for now)?
(i) $\mathbf{P}(B, I, M) = \mathbf{P}(B)\mathbf{P}(I)\mathbf{P}(M)$
(ii) $\mathbf{P}(J|G) = \mathbf{P}(J|G, I)$
(iii) $\mathbf{P}(M|G, B, I) = \mathbf{P}(M|G, B, I, J)$

(b) (2) Calculate the value of $P(b, i, \neg m, g, j)$.

(c) (4) Calculate the probability that someone goes to jail given that they broke the law, have been indicted, and face a politically motivated prosecutor.

(d) (2) A *context-specific* independence has the following form: X is conditionally independent of Y given Z in context $\mathbf{C} = \mathbf{c}$ if $\mathbf{P}(X|Y, Z, \mathbf{C} = \mathbf{c}) = \mathbf{P}(X|Z, \mathbf{C} = \mathbf{c})$. In addition to the usual conditional independences given by the graph structure, what context-specific independences exist in the Bayes net in Fig. 1?

(e) (5) Suppose we want to add the variable $P = PresidentialPardon$ to the network; draw the new network and briefly explain any links you add.

7. (18 pts.) Language and statistical learning

A *probabilistic context-free grammar* (PCFG) is a context-free grammar augmented with a probability value on each rule, so that the PCFG specifies a probability distribution over all allowable strings in the language. Specifically, the rules for each nonterminal are annotated with probabilities that sum to 1; these define how likely it is that each rule is applied to expand that nonterminal; and all such rule choices are made independently. For example, the following is a PCFG for simple verb phrases:

$0.1 : VP \rightarrow Verb$
$0.2 : VP \rightarrow Copula\ Adjective$
$0.5 : VP \rightarrow Verb\ the\ Noun$
$0.2 : VP \rightarrow VP\ Adverb$
$0.5 : Verb \rightarrow is$
$0.5 : Verb \rightarrow shoots$
$0.8 : Copula \rightarrow is$
$0.2 : Copula \rightarrow seems$
$0.5 : Adjective \rightarrow$ **unwell**
$0.5 : Adjective \rightarrow$ **well**
$0.5 : Adverb \rightarrow$ **well**
$0.5 : Adverb \rightarrow$ **badly**
$0.6 : Noun \rightarrow$ **duck**
$0.4 : Noun \rightarrow$ **well**

(a) (3) Which of the following have a nonzero probability of being generated as complete VPs?
(i) shoots the duck well well well (ii) seems the well well (iii) shoots the unwell well badly

(b) (5) What is the probability of generating "is well well"?

(c) (2) What types of ambiguity are exhibited by the phrase in (b)?
(i) lexical (ii) syntactic (iii) referential (iv) none

(d) (2) *True/False*: Given any PCFG, it is possible to calculate the probability that the PCFG generates a string of exactly 10 words.

(e) (1) A PCFG learning algorithm takes as input a set of strings D and outputs the PCFG h that maximizes $\log P(D|h) - C(h)$, where C is a measure of the complexity of h. For a suitably defined C, this could be an example of
(i) Bayesian learning (ii) MAP learning (iii) maximum likelihood learning.

(f) (5, tricky) An *HMM grammar* is essentially a standard HMM whose state variable is N (nonterminal, with values such as Det, $Adjective$, $Noun$ and so on) and whose evidence variable is W (word, with values such as is, $duck$, and so on). The HMM model includes a prior $\mathbf{P}(N_0)$, a transition model $\mathbf{P}(N_{t+1}|N_t)$, and a sensor model $\mathbf{P}(W_t|N_t)$. Show that every HMM grammar can be written as a PCFG. [Hint: start by thinking about how the HMM prior can be represented by PCFG rules for the sentence symbol. You may find it helpful to illustrate for the particular HMM with values A, B for N and values x, y for W.]