CS 495 & 540
Spring 2016 Artificial Intelligence

# Final Exam

- The exam is closed book, closed notes except a one-page crib sheet.
- The total score is 120pts and you have approximately 120 minutes. Use your time wisely. If you get stuck, it is advised that you should try those questions that are most rewarding with respect to the time it takes you to solve.
- I leave plenty of space for each problem. Please write your solution on the exam itself. Two blank sheets are attached at the back of the exam to serve as scratch paper for you. DO NOT detach the sheets.
- If you finish the exam early, please leave the exam on the desk and I will collect it. Good luck and thank you for taking this class.

| | |
|---|---|
| First name | |
| Last name | |

**For staff use only:**

| | | |
|---|---|---|
| Q1. | Warm-Up | /3 |
| Q2. | Heuristics: Application & Design | /16 |
| Q3. | BFS/DFS/Iterative Deepening Revisited | /20 |
| Q4. | CSPs: Job Assignments | /6 |
| Q5. | A Star Search | /15 |
| Q6. | Baysian Network: Independence | /15 |
| Q7. | Hidden Markov Model: Computation | /12 |
| Q8. | Inference: Enumeration and Variable Elimination | /15 |
| Q9. | NLP: Smoothing | /10 |
| Q10. | Limiting Distribution | /8 |
| Q11. | Scratch paper: Do Not Detach | /0 |
| Q12. | Scratch paper: Do Not Detach | /0 |
| | Total | /120 |

# Q1. [3 pts] Warm-Up

Circle the AI mascot that would respesent your behavior in this final exam if not being proctored ...

# Q2. [16 pts] Heuristics: Application & Design

$n$ vehicles occupy squares $(1, 1)$ through $(n, 1)$ of an $n \times n$ grid. The vehicles must be moved 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 just 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) Calculate the size of the state space as a function of $n$.

(b) Calculate the branching factor as a function of $n$

(c) 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 this goal location $(n - i + 1, n)$, assuming no other vehicles are on the grid.

(d) Which of the following heuristics are admissible for the problem of moving all $n$ vehicles to their destinations? **Explainatioin is required. Guessing correctly won't gain any partial credits.**

(i) $\sum_{i=1}^{n} h_i$

(ii) $\max\{h_1, \cdots, h_n\}$

(iii) $\min\{h_1, \cdots, h_n\}$

# Q3. [20 pts] BFS/DFS/Iterative Deepening Revisited

Given a balanced tree with branching factor = $b$ and height = $h$. Let suppose the goal is hidden at level $k$ (**now let us assume** ***root*** **is at level** $h$ **and leaves are at level 0**). It turns out this question is not that easy as you could imagine because there is another factor $t$ (how far the target node is away from the **leftmost** node in level $k$, assuming we label the nodes from 1, 2, ..., t, ... , $b^k$, for level $k$). Let us assume that those search algorithms start from **right most, instead of leftmost**. So, in order to compare Iterative deepening with DFS, you must count the complexity. With the given condition above, please compute (**Please show step by step calculation since this problem has appeared couple times ... I will check the computation seriously**):

(a) Complexity of Iterative Deepening [Hint: you might want to use ceiling function for this and you have to be careful with how many time level 0 till $k-1$ are counted from iteration 1 till $k-1$. And in the very last iteration (the $k$ level), not all nodes in level 1 to $k-1$ are visited]

(b) Compute the complexity of DFS

# Q4. [6 pts] CSPs: Job Assignments

In some exam, there are a total of 6 questions on the exam and each question will cover a topic. Here is the format of the exam:

- q1. Search
- q2. Games
- q3. CSPs
- q4. MDPs
- q5. True/False
- q6. Short Answer

There are 7 people on the course staff: Brad, Donahue, Ferguson, Judy, Kyle, Michael, and Nick. Each of them is responsible to work with Dr. Chiang on one question. (But a question could end up having more than one staff person, or potentially zero staff assigned to it.) However, the staff are pretty quirky and want the following constraints to be satisfied:

(i) Donahue (D) will not work on a question together with Judy (J).

(ii) Kyle (K) must work on either Search, Games or CSPs.

(iii) Michael (M) is very even, so he can only contribute to an even-numbered question.

(iv) Nick (N) must work on a question that's before Michael (M)'s question.

(v) Kyle (K) must work on a question that's before Donahue (D)'s question

(vi) Brad (B) does not like grading exams, so he must work on True/False.

(vii) Judy (J) must work on a question that's after Nick (N)'s question.

(viii) If Brad (B) is to work with someone, it cannot be with Nick (N).

(ix) Nick (N) cannot work on question 6.

(x) Ferguson (F) cannot work on questions 4, 5, or 6

(xi) Donahue (D) cannot work on question 5.

(xii) Donahue (D) must work on a question before Ferguson (F)'s question.

**(a)** [2 pts] We will model this problem as a constraint satisfaction problem (CSP). Our variables correspond to each of the staff members, J, F, N, D, M, B, K, and the domains are the questions 1, 2, 3, 4, 5, 6. After applying the **unary constraints**, what are the resulting domains of each variable? (The second grid with variables and domains is provided as a back-up in case you mess up on the first one.) **Note: Different from midterm.**

| | | | | | | |
|---|---|---|---|---|---|---|
| B | 1 | 2 | 3 | 4 | 5 | 6 |
| D | 1 | 2 | 3 | 4 | 5 | 6 |
| F | 1 | 2 | 3 | 4 | 5 | 6 |
| J | 1 | 2 | 3 | 4 | 5 | 6 |
| K | 1 | 2 | 3 | 4 | 5 | 6 |
| N | 1 | 2 | 3 | 4 | 5 | 6 |
| M | 1 | 2 | 3 | 4 | 5 | 6 |

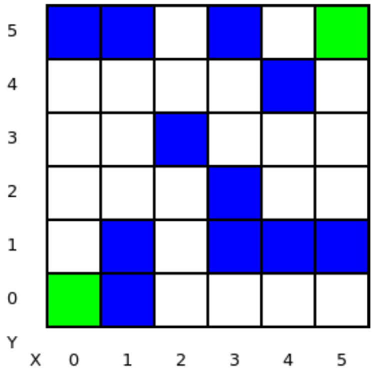
| | | | | | | |
|---|---|---|---|---|---|---|
| B | 1 | 2 | 3 | 4 | 5 | 6 |
| D | 1 | 2 | 3 | 4 | 5 | 6 |
| F | 1 | 2 | 3 | 4 | 5 | 6 |
| J | 1 | 2 | 3 | 4 | 5 | 6 |
| K | 1 | 2 | 3 | 4 | 5 | 6 |
| N | 1 | 2 | 3 | 4 | 5 | 6 |
| M | 1 | 2 | 3 | 4 | 5 | 6 |

**(b)** [4 pts] If we apply the Minimum Remaining Value (MRV) heuristic, please list the first three variables that will be chosen and their associated value. And if the variable has mulitple options, choose the lowest value.

# Q5. [15 pts] A Star Search

In this problem we are given the following configuration that was seen in your homework. Each single step move has the cost of 1. For instance, move from $(0,0)$ to $(0,1)$ has cost 1. And let say if the distance is $k$, then the cost is $k$. Here is some twist:

(I)[3]How many paths (a cell can be visited at most once ) are there to go from (0,0) to (5,5), given the blue cells are the barriers?




(II)[3] So, we want to find the solution using A* search and we have decided to use a hybrid heustric $h = \alpha * h_1 + \beta * h_2$ where $\alpha > 0, \beta > 0$, $\alpha + \beta = 1$, $h_1$ is the Manhattan Distance and $h_2$ is the Euclidean distance. Is this a valid heuristic? Why?

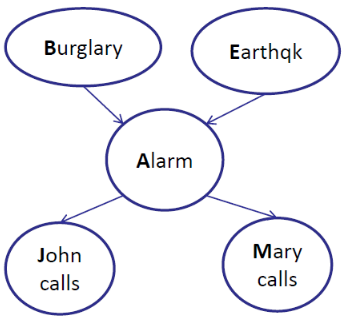
(II)[9] Now let say you decided to use Eclidean Distance only. At $(x, y) = (0, 2)$, your open list should be $[(1,2),(0,3)]$ and your closed list is $[(0,0),(0,1),(0,2)]$. Please simluate the next three moves and the corresponding closed lists and open lists. (e.g. $\{(0,1),[(0,2)],[(0,0),(0,1)]\} \rightarrow \{(0,2),[(1,2),(0,3)],[(0,0),(0,1),(0,2)]\}$

# Q6. [15 pts] Baysian Network: Independence

(a) [9pt] If we observe Alarm = true , are Burglary and Earthquake independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence

| B | P(B) |
|---|---|
| +b | 0.001 |
| -b | 0.999 |

| E | P(E) |
|---|---|
| +e | 0.002 |
| -e | 0.998 |




| A | J | P(J\|A) |
|---|---|---|
| +a | +j | 0.9 |
| +a | -j | 0.1 |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

| A | M | P(M\|A) |
|---|---|---|
| +a | +m | 0.7 |
| +a | -m | 0.3 |
| -a | +m | 0.01 |
| -a | -m | 0.99 |

(b) [6pt] Compute P(+a,-b,+e,-j, -m), P(+a,-b,-e,+j, +m) and P(-a,-b,+e,+j,+m).

# Q7. [12 pts] Hidden Markov Model: Computation

Hidden Markov Model (HMM) basically is a belief propatation process. It starts with an initially uniform distribution for the unseen random variable $X$ by checking the highly correlated observable evidence variable $e$. We know that (I) passage of time phase $B'(X_{t+1}) = P(X_{t+1}|e_{1\cdots t})$ and (II) observation phase $B(X_{t+1}) = P(X_{t+1}|e_{1\cdots t+1})$. Follow the weather HMM with the following modification: $P(+r \rightarrow +r) = 0.7, P(+r \rightarrow -r) = 0.3, P(-r \rightarrow +r) = 0.5, P(-r \rightarrow -r) = 0.5, P(+r \rightarrow +u) = 0.7, P(+r \rightarrow -u) = 0.3, P(-r \rightarrow +u) = 0.4, P(-r \rightarrow -u) = 0.6$. And let assume you live in Utica, therefore your initial belief is $B_0(+r) = 0.35, B_0(-r) = 0.65$. Please compute $B'_i(+r), B'_i(-r), B_i(+r), B_i(-r)$ where $i = 1, 2, 3$.

1.$B'_1(+r) =$

2.$B_1(+r) =$

3.$B'_2(+r) =$

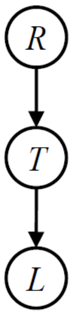
$4.B_2(+r) =$

$5.B_3'(+r) =$

$6.B_3(+r) =$

# Q8. [15 pts] Inference: Enumeration and Variable Elimination

The main advantage of variable elimination over enumeration is the complexity reduction. The general idea is that you marginalinze the joint probability space early such that it costs less in next iteration of join operation. Given the following, please compute $P(L)$ using (I)[5] Enumerationg approach (II)[5] Variable Elimination. Finally, (III)[5]please justify if II is better than I by comparing the cost of generating $P(L)$ via those two different approaches .




$P(R)$

| | |
|---|---|
| +r | 0.1 |
| -r | 0.9 |

$P(T|R)$

| | | |
|---|---|---|
| +r | +t | 0.8 |
| +r | -t | 0.2 |
| -r | +t | 0.1 |
| -r | -t | 0.9 |

$P(L|T)$

| | | |
|---|---|---|
| +t | +l | 0.3 |
| +t | -l | 0.7 |
| -t | +l | 0.1 |
| -t | -l | 0.9 |

# Q9. [10 pts] NLP: Smoothing

One of the common problems in NLP is how well the machine handles unknown words (words not seen in the training data set). We learned about **Laplace Smoothing** which is adding 1 occurence to each word in the vocabulary then recalculate the probability for the predictor. However, this is rather primitive. Let us explore the **Good Turing Smoothing**. An example is as following: you were fishing and you caught 10 carp, 3 perch, 2 whitefish, 1 trout, 1 salmon, 1 eel (total : 18 fish). So, how likely the next fish (next specises) is new? Good Turing will assume it is 3/18 (sum of those you only see once). It is obvious now the probability of seeing a trout in the next catch is definitely less than 1/18. Please describe a method to scale down the probability of seeing a trout from 1/18 to a lower number such that how many trouts you see in previous catches will have an effect in the new probability (prefer Good Turing Smoothing). Please also justify your answer.

# Q10. [8 pts] Limiting Distribution

In problem set 3, we talk about stochastic matrices and some of the convergence property. We know about the spectral gap is 1-$|\lambda|$ where $\lambda$ is the 2nd largest eigenvalue. Given a matrix $G$ that is an $N \times N$ **column-wise stochastic** matrix and a uniform distribution column vector $v$, someone claims that $v$ is the limiting (stationary) distribution of $G$, i.e. $Gv = v$. What extra property can you impose on $G$ such that this is true? Given an example of $G$ (and assume $N = 4$ in the example).

# Q11. [0 pts] Scratch paper: Do Not Detach

**(a)** [0 pts]

# Q12. [0 pts] Scratch paper: Do Not Detach

**(a)** [0 pts]