STEVENS INSTITUTE OF TECHNOLOGY

CS 541A– MIDTERM
Artificial Intelligence

07/03/2022

Read the instructions carefully before starting the exam.

Instructions:

- The total time for this exam is 150+10 minutes. No extra time will be allotted.
- The 10 additional minutes are given to upload the solutions to Canvas.
- This exam consists of 5 questions. Questions 1 and 2 are compulsory. You can pick any two questions from 3, 4 and 5.
- If all three questions (3, 4 and 5) are answered, only the first two will be graded.
- The solutions should be submitted on Canvas.
- The submission file can be in any format. For example, you can write the solutions in a notepad, take a picture and upload it.
- Each image (page) of the submission file should be clearly marked with a page number.
- This is an open book exam.

Good luck!

Distribution of Marks

Question	Points	Score
1	30	
2	30	
3/4/5	20	
3/4/5	20	
Total:	100	

1. For each of the following, select either True or False. If False, give either the correct answer or a one line reason for selecting False. For True this step can be omitted.
   Note: Partial credit will be given for the correct answer with incorrect reasoning.

   (a) (2 points) Is the following conditional probability equation true for any two random variables A and B. $P(A|B) = \sum_{a \in A} P(B|A = a)$
   True/False
   Points awarded!

   (b) (2 points) Is the following conditional probability equation true for any two random variables A and B. $P(A|B) = \sum_{b \in B} P(A|B = b)$
   True/False
   Points awarded!

   (c) (2 points) A regression problem can be formulated as a clustering problem by considering each data point's class labels as the cluster label.
   True/False
   False. Unlike clustering, a regression problem has continuous labels.

   (d) (2 points) A classification problem can be formulated as a clustering problem by considering each data point's class labels as the cluster label.
   True/False
   True.

   (e) (2 points) K-Means only works for data which can be visualized in concentric circles.
   True/False
   False. K-Means considers clusters are linearly separable.

   (f) (2 points) An underfit model is better than an overfit model as it gives lower test error.
   True/False
   False. Neither model guarantees lower test error compared to the other one.

   (g) (2 points) Gradient descent computes the gradient of the loss for the complete training data and then updates the model parameters (w).
   True/False
   True.

   (h) (2 points) Smaller learning rate should be prefered over a larger learning rate if we want to arrive at the optimal solution faster.
   True/False
   False. Smaller learning rate may increase the time to get to the optimal solution.

   (i) (2 points) Each of the nodes (excluding the root) in a decision tree represent a decision rule.

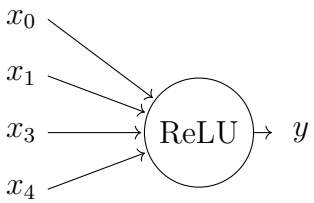
True/False
False. Tereminal/Leaf nodes do not represent decision rule.

(j) (2 points) One reason to prefer a decision tree over a neural network is that it is interpretable?
True/False
True.

(k) (2 points) Backpropogation works by propagating gradient loss from the input layer to the output layer.
True/False
False. Backpropagation propagates gradient loss from output layer to input layer.

(l) (2 points) In neural networks, non-linear activation functions (e.g., ReLU, *tanh*) help learn non-linear decision boundaries?
True/False
True.

For the next three questions consider the following neural network. Given the input $x = [-1, -1, -1, -1]$ is it possible (for any weights) to get the following outputs.



(m) (2 points) $y = -1$?
True/False
False. ReLU outputs values ≥ 0

(n) (2 points) $y = +1$?
True/False
True.

(o) (2 points) The gradient of the output $\frac{\partial y}{\partial w} = \begin{cases} 0 \text{ if } x \geq 0 \\ 1 \text{ otherwise} \end{cases}$.
True/False
False. $\frac{\partial y}{\partial w} = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$

2. Write a short answer (one/two lines at most) for each of the following questions.

(a) (2 points) What are the criteria for a heuristic to be consistent?
$h(A) \leq Cost(A \rightarrow B) + h(B)$. $h(goal) = 0$.

(b) (2 points) One condition to use minimax search algorithm is that the environment is observable. Give two other conditions.
(Any 2 reasons) 2 Player game. Zero-sum game. Deterministic actions from any state.

(c) (2 points) What is the difference between a search algorithm being complete vs optimal?
Completeness means a solution is found if available. Optimality means the best solution can be found.

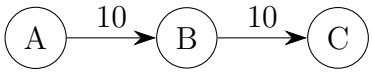
(d) (2 points) Give one reason why Iterative Deepening Search may require smaller memory size compared to BFS?
BFS needs to keep track of all the paths for any given level.

(e) (2 points) For finding a path through a maze, give a reason why we can/cannot use alpha-beta pruning?
Alpha-Beta pruning is used form zero-sum problems.

(f) (2 points) Even though the best-case running time of DFS is better than Backtracking search, give a reason why one might prefer Backtracking search.
If the cost of actions can be negative.

(g) (2 points) A^* search is as good as the heuristic it uses. Give an example of a heuristic for which A^* is not the best search algorithm.
Any heuristic which is not consistent. For example, $h(goal) > 0$

The following graph shows the actual costs of each action and the table shows heuristic values of each state.



s	A	B	C
h(s)	5	?	0

(h) (2 points) Given A as the start state and C as the goal state, give the range of values for which h(B) is admissible.

[2 pts] $0 \leq h(B) \leq 10$
[1 pt] If 0 and 10 are correct but incorrect inequality.

For questions (i) & (j), assume you are asked to code an autonomous robot which automatically moves from an initial state to a goal state. You define the following variables for this problem.
pos_x, pos_y: current position of the robot
$goal\_x, goal\_y$: the position of the goal state
To simplify the problem, you assume that the robot can only move forward and never backward. This means every action will increase either pos_x or pos_y by any value.

(i) (2 points) Complete the function which checks whether the current position is the goal state.

```
def isEnd():
    if ???
        return True
return False
```

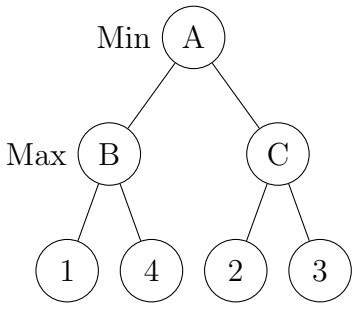
$(pos\_x == goal\_x)$ AND $(pos\_y == goal\_y)$

(j) (2 points) Complete the function which returns True if an action is possible to a new state (i, j).

```
def isSuccessor(i,j):
    if ???
        return True
    return False
```

Any condition where pos_x is less than i or pos_y is less than j
$(pos\_x < i)$ OR $(pos\_y < j)$

For questions (k), (l) & (m) consider the minimax tree below.



(k) (2 points) Run minimax algorithm without pruning and give the values for A, B and C.
A=3,B=4,C=3

(l) (2 points) Alpha-Beta Pruning depends upon the order of the nodes expanded. If the algorithm always expands the left child and then the right child (e.g. from A it expands B and then C), which edge (if any) will be pruned.

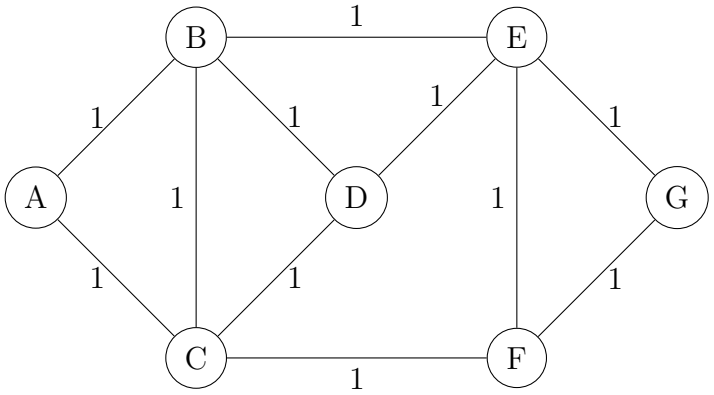
None

(m) (2 points) If the algorithm always expands the right child and then the left child (e.g. from A it expands C and then B), which edge (if any) will be pruned.

Edge B to 1

For next two questions, consider the graph below where A is the start node and G is the end node.

Note: All ties must be broken alphabetically.

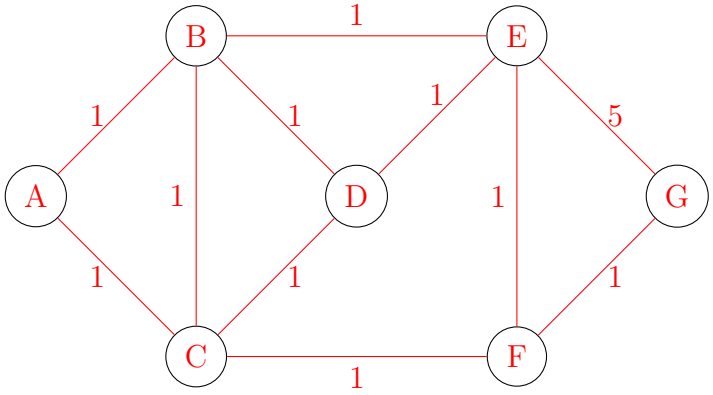


(n) (2 points) Which algorithm will output the following paths.

Algorithm	**B-E-G**	**A-B-C**
Depth First Search		X
Breadth First Search	X	

(o) (2 points) Right now both paths **A-B-E-G** and **A-C-F-G** have the same cost. Draw the graph again with different edge costs such that Uniform Cost Search chooses the path **A-C-F-G**.

Any edge costs where cost of path A-B-E-G is greater than path A-C-F-G.



3. You are given a task to train a model which predicts whether a student uses a bus or walks to school. You choose the amount of rain and whether the student lives far from school as the features for your model. Based on this you take a survey and collect the following data.

Data	Rain	Far	Bus
1	Low	Yes	Yes
2	High	Yes	Yes
3	Low	Yes	Yes
4	Low	No	No
5	High	No	Yes
6	Medium	Yes	Yes
7	High	No	Yes
8	Medium	No	No
9	High	Yes	Yes
10	Medium	No	No

Note: You need to show all the calculations in each of the questions below. You should use log_2 for the calculations.

(a) (2 points) What is the Entropy(Bus)?
$Entropy(Bus) = -1 * \{\frac{7}{10}\log(\frac{7}{10}) + \frac{3}{10}\log(\frac{3}{10})\} = 0.881$
[1 pt] Correct equation and values.
[1 pt] Correct result.

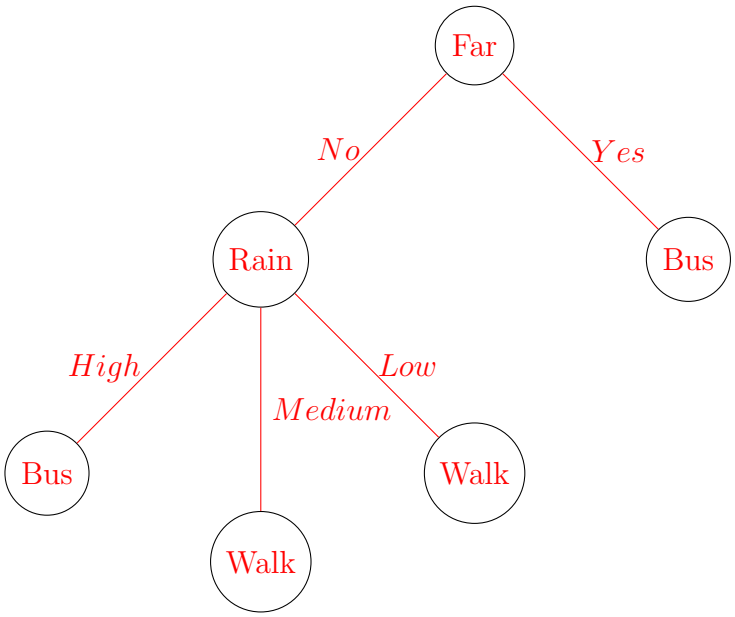
(b) (4 points) What is the Entropy(Bus,Rain)?
$Entropy(Bus, Rain)$
$= \frac{4}{10} * E(Bus, Rain = High) + \frac{3}{10} * E(Bus, Rain = Medium) + \frac{3}{10} * E(Bus, Rain = Low)$
$= 0.4*(-1*\{\frac{4}{4}\log(\frac{4}{4})\})+0.3*(-1*\{\frac{1}{3}\log(\frac{1}{3})+\frac{2}{3}\log(\frac{2}{3})\})+0.3*(-1*\{\frac{2}{3}\log(\frac{2}{3})+\frac{1}{3}\log(\frac{1}{3})\})$
$= 0.0 + 0.275 + 0.275 = 0.550$
[3 pt] Correct equation and values.
[1 pt] Correct result.

(c) (4 points) What is the Entropy(Bus,Far)?
$Entropy(Bus, Far)$
$= \frac{5}{10} * E(Bus, Far = Yes) + \frac{5}{10} * E(Bus, Far = No)$
$= 0.5 * (-1 * \{\frac{5}{5}\log(\frac{5}{5})\}) + 0.5 * (-1 * \{\frac{2}{5}\log(\frac{2}{5}) + \frac{3}{5}\log(\frac{3}{5})\})$
$= 0.0 + 0.48548 = 0.485$
[3 pt] Correct equation and values.
[1 pt] Correct result.

(d) (2 points) Using ID3 algorithm, which feature is selected first for the decision rule. You need to show all the calculations.
Far is selected
$IG(Bus, Rain) = E(Bus) - E(Bus, Rain)$
$= 0.881 - 0.550 = 0.331$

$IG(Bus, Far) = E(Bus) - E(Bus, Far)$
$= 0.881 - 0.485 = 0.396$
[1+1 pt] Correct equations

(e) (5 points) Use ID3 algorithm to select the second feature (if any) for each internal node. Also draw the complete decision tree with the decision rules and the decisions? You need to show all the calculations.
$E(Bus, Far = Yes) = 0$
$IG(Bus, Yes, Rain) = 0$
$E(Bus, Far = No) = 0.971$
$E(Bus, Far = No, Rain) = 0$
$IG(Bus, No, Rain) = E(Bus, No) - E(Bus, No, Rain) = 0.971 - 0 = 0.971$



[2 pt] IG correctly computed.
[2 pt] Correct 2nd feature selection for both children.
[1 pt] Decision tree is correct.

(f) (3 points) Use the decision tree from the previous answer to predict whether the student will take the bus to school in following two scenarios?
i. $p1 = (Rain = High, Far = Yes)$?
[1.5 pt] Bus

ii. $p2 = (Rain = Medium, Far = No)$?
[1.5 pt] No/Walk

4. You are given a task to train a model which predicts lung cancer in patients. You assume that lung cancer only depends upon the number of cigarettes one smokes and the amount of pollutants in the area of residence. You took a survey in which you asked different people about their smoking habits and measured the air pollution. The following is the data you collected.

Data	Average cigarettes per day	Air pollution in ppm
1	1.90	0.97
2	1.76	0.84
3	2.32	1.63
4	5.02	3.02
5	5.74	3.60
6	4.79	3.84

However you forgot to ask the people in your survey whether they were diagnosed with lung cancer or not. What you have planned to do now is to group the data and hope you will find the two groups you are looking for.

Note: You need to show all the calculations in each of the questions below.

(a) (2 points) Given $K = 2$, what are suitable initial cluster centers for your K-Means algorithm?
Any two points are acceptable

(b) (6 points) Compute a single iteration of K-Means using the cluster centers chosen above. You should mention the distance measure you used and show the final cluster assignments.
[1 pt] Distance measure mentioned.
[2 pt] Distance equation based on the distance measure.
[2 pt] Compute distance measure for each point.
[1 pt] Correct cluster assignment.

(c) (4 points) Compute the new cluster centers after the previous step.
[2 pt] Mean calculation for each cluster.
[2 pt] New means computed.

(d) (4 points) Given a new data (5.20, 7.51), assign it to the appropriate cluster based on the distance measure you have selected.
[3 pt] Compute distance of new point with new means.
[1 pt] Assigned to the correct cluster.

(e) (2 points) You found that the pollution reading in the previous question was incorrect. Propose a way to detect such noisy inputs in the future. You can assume your K-Means algorithm (in part c) gives the correct final cluster centers as output.
[2 pt] Suitable method to detect outliers given.

(f) (2 points) Using the criteria defined previously, determine which of the following are incorrect readings.
i. $p1 = (2.0, 4.0)$

ii. $p2 = (5.0, 4.0)$

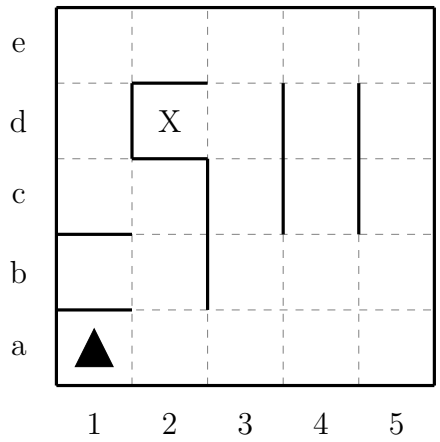
[1+1 pt] Computed the correct clusters based on the outlier detection method proposed.

5. You are given a task to program a vacuum cleaner which automatically cleans dirt from the floor. Each of the thicker lines represent the walls in the building. And X represent one possible location of the dirt.

   To solve the problem, you have decided to map the floor to a 5×5 grid. Each grid location can be represented as a tuple, for example the vacuum's initial location is $(a, 1)$ and the dirt location is $(d, 2)$.

   A vacuum's current state is defined by the cell it is in $(a, 1)$ and one of the four possible directions (N, S, E, W). Consider moving and turning as two separate actions. For example, vacuum is initially stationary in cell $(a, 1)$ pointing towards N. It can move to cell $(a, 2)$ with 2 actions 'turn E' and '*move*'.

   Note: You need to show all the calculations in each of the questions below.



   (a) (3 points) Given an $M \times N$ grid, how many possible states the vacuum can have? Justify your answer.
   M*N*4. The total number of possible positions on the grid times number of the directions of the vacuum.
   [1 pt] Correct answer.
   [2 pt] Correct reason.

   (b) (3 points) What is the cost $Cost(a, 1, N)$ of the vacuum going from its initial state to the dirt location X. Justify your answer.
   Cost = 9. 1 to move East. 2 steps forward. 1 to move North. 3 steps forward. 1 to move West. 1 step forward.
   [1 pt] Correct answer.
   [2 pt] Correct reason.

   (c) (6 points) You plan to implement A^* search to find the dirt location. Give a consistent heuristic which you can use. Show how it is consistent. Also show how the heuristic is admissible.
   For example, Manhattan distance without any walls. Manhattan distance of start to dirt

is less than actual cost and of dirt (goal) is zero.
A consistent heuristic is also admissible.
[2 pt] Correct heuristic.
[2 pt] Correct reason for consistent.
[2 pt] Correct reason for admissible.

(d) (3 points) Using the heuristic given above, compute $h(a, 1)$, which is the heuristic value of the cost for the vacuum to reach the dirt.
Manhattan distance of dirt = 1+3
[2 pt] Correct heuristic used.
[2 pt] Correct result.

(e) (2 points) Using the heuristic given above, compute $h(d, 2)$, which is the heuristic value of the cost for the dirt.
Heuristic of dirt = 0
[2 pt] Correct result.

(f) (3 points) Based on the heuristic values and the cost computed in previous steps, what will be the new cost $Cost'(a, 1, N)$
$Cost'(a, 2, N) = Cost(a, 2, N) - h(a, 2) + h(d, 2) = 9 - 6 + 0 = 3$
[2 pt] Correct equation.
[2 pt] Correct result.