IMPORTANT.
The aim of these practice questions is to help you review.
It does not reflect the length nor the content of the exam
Kindly, do not distribute.

Name: ______________________________

UNI: ______________________________

1. What are the criteria used to evaluate the performance of search algorithms?

2. BFS versus DFS. Please check all that apply:
   - ☐ DFS and BFS use a similar amount of space.
   - ☐ DFS uses less space than BFS.
   - ☐ Both of them have exponential time.

3. The k-Nearest Neighbors approach scales well to high dimensional spaces. Justify briefly.
   - ◯ True
   - ◯ False

4. Classification with *decision trees* is a linear classification method. Justify briefly.
   - ◯ True
   - ◯ False

5. Define *supervised learning* and *unsupervised learning*.

6. How are *genetic algorithms* and *hill climbing* the same, and how are they different? Be specific.

7. "A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. " Tom Mitchell.

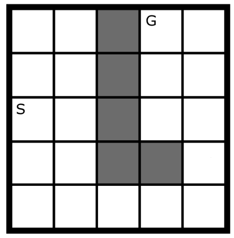
   What are the sets E, T, and P in the case of a **Recommender System**? Please justify your answer and elaborate with examples. Provide 2 or more measures for P.

8. For the learning situation below, say whether KNN or decision trees would be best to use, and why.

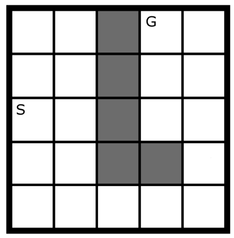
   "You are going to develop a classifier to recommend which children should be assigned to special education classes in kindergarten. The classifier has to be justified to the board of education before it is implemented."

9. Given a maze with the start state S and the goal state G. Number the cells (starting from 1) in the order they are visited by each algorithm using the following guidelines: Gray cells contain obstacles and cannot be used. Moves are considered in the following order: Up, Right, Down, Left. Diagonal moves are not allowed. No state is visited twice.

   (a) Breadth-First search



(b) Depth-First search



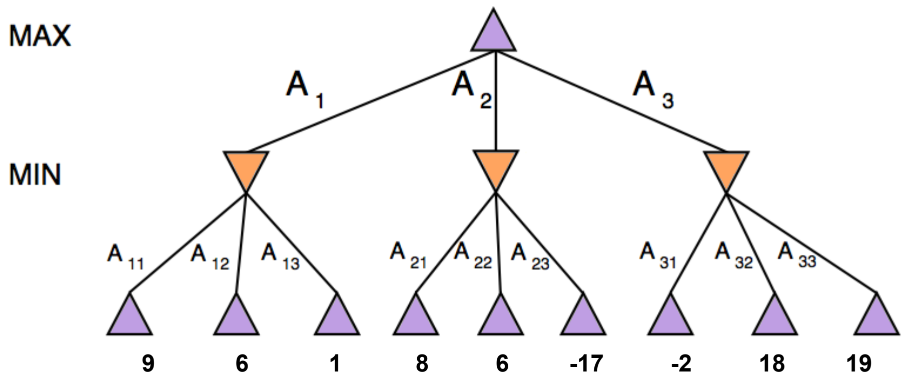
10. A Mars rover has to leave the lander, collect rock samples from three places (in any order) and return to the lander. Assume that it has a navigation module that can take it directly from any place of interest to any other place of interest. So it has primitive actions go-to-lander, go-to-rock-1, go-to-rock-2, and go-to-rock-3. We know the time it takes to traverse between each pair of special locations. Our goal is to find a sequence of actions that will perform this task in the shortest amount of time.

    1. Formulate this problem as a search problem by specifying the state space, initial state, path-cost function, and goal test. Try to be sure that the state space is detailed enough to support solving the problem, but not redundant.

    2. What search technique would be most appropriate? Explain.

    3. One possible heuristic evaluation function for a state would be the amount of time required for the robot to go back to the lander from the location of the state. This is clearly admissible. What would be a more powerful, but still admissible, heuristic for this problem?

11. Consider the following search tree.

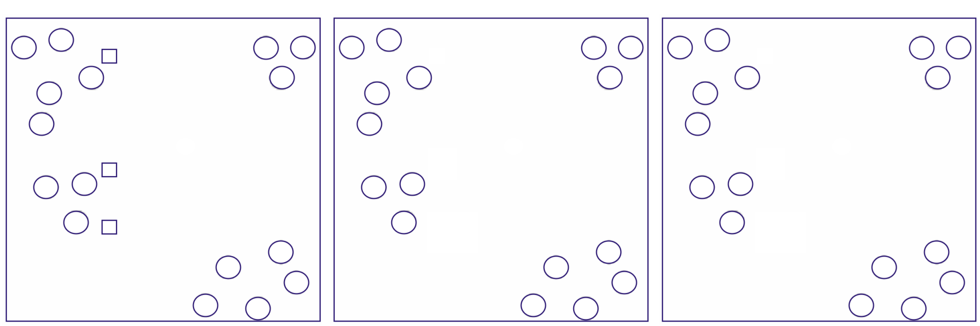


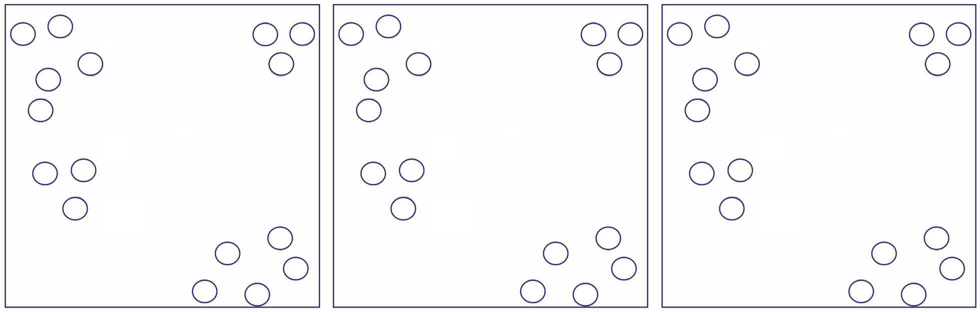
1. Using minimax, which of the three possible moves should MAX take at the root node? and what is the value of MAX at the root?

2. Using minimax with alpha-beta pruning, compute the value of alpha and beta at each node. What branches are pruned? **Show the values of α and β on the tree.**

3. Is there a way to re-order the leaves within the subtrees to further prune the tree? Explain.

12. Run K-means manually on the following dataset. Circles are data points and squares are the initial cluster centers. Use three colors (or other means like numbering the points with corresponding cluster center or drawing boundaries) to show the groupings of points to clusters. (NOTE: It is not necessary to draw the exact location of the squares, but it should be clear from your placement of the squares that you understand how K-means performs quantitatively.) Trace through the first six iterations of the K-means algorithm or until convergence is reached.





13. Compare logistic regression and linear regression. How are they different and how are they alike? Be specific.

14. Decision trees

    Consider the following dataset with $n_\oplus = 3$ positive examples and $n_\ominus = 3$ negative examples.

Example number	a1	a2	label
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	T	-
6	F	T	-

The entropy function at a given node S used for decision trees is given by:

$Entropy(S) = -p_\oplus \ log_2 \ p_\oplus \ - p_\ominus \ log_2 \ p_\ominus$

1. Give a formula for $p_\oplus$ and $p_\ominus$ (the proportion of positive and negative examples respectively).

2. What is the entropy at the root?

3. What is the information gain for feature a_1? for a_2? which of these two features will be chosen at the root?

4. Suppose we used *Example number* as a feature. Note all examples have different numbers. What would be the gain of this feature at the root? Would this feature be picked up first or last by the algorithm to split the data at the root? What would be the decision tree then.

15. Naive Bayes & logistic regression: Compare the two methods and highlight the strengths and weaknesses of each.

16. The idea of boosting is to train weak learners on weighted training examples. Check all that apply.
    - ◯ Give large weights to easy examples to get rid of them
    - ◯ Use any classifier as far as its accuracy is slightly worse than random

◯ The classification output is a majority voting of all weak classifiers outputs

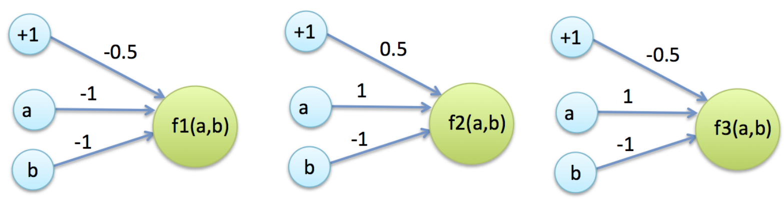
17. Explain how to find the optimal K in K-means.

18. With two hidden layers, Neural Networks can find any arbitrary decision boundary.

    Your answer: A. True B. False

19. Given the following boolean function. Consider perceptron model, the neuron calculates a weighted sum of inputs. Then, it applies a threshold to the result: if the sum is larger than zero, the output is 1. Otherwise, the output is zero. Consider the function f below. Among the perceptrons below, which one(s) represent f? Justify your answer.

a	b	$f(a,b)$
1	1	0
0	0	0
1	0	1
0	1	0



◯ $f1$
◯ $f2$
◯ $f3$

20. Which of the following does the Naive Bayes classifier assume?

    ◯ All the attributes are independent.
    ◯ All the attributes are conditionally independent given the output label.
    ◯ All the attributes are jointly dependent to each other.

21. Naive Bayes

Consider the tennis dataset:

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

1. Predict the class Play for the new example:
   (Outlook = sunny, Temperature = cool, Humidity = high, Windy =strong)
   No need to calculate all probabilities, calculate only the ones you need to make this prediction.

$$y_{new} = \text{argmax}_{y \in \{yes,no\}} p(y) * p(outlook = sunny|y) * p(temp = cool|y) *$$
$$p(humidity = high|y) * p(windy = strong|y)$$
$$p(play = yes) = 9/14 = 0.64$$
$$p(play = no) = 5/14 = 0.36$$

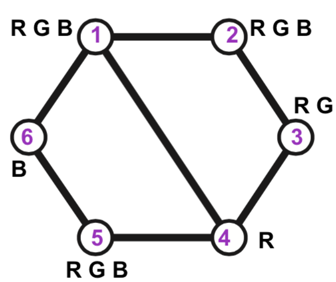
   Conditional probabilities:

$$p(wind = strong|play = yes) = 3/9 = 0.33$$
$$p(wind = strong|play = no) = 3/5 = 0.6$$
$$p(yes) * p(sunny|yes) * p(cool|yes) * p(high|yes) * p(strong|yes) = 0.0053$$
$$p(no) * p(sunny|no) * p(cool|no) * p(high|no) * p(strong|no) = 0.0206$$

$$y_{new} = no$$

22. CSP

Consider the following constraint graph for a graph coloring problem (the constraints indicate that connected nodes cannot have the same color). The domains are {R, G, B}.



(a) What are the variable domains after a full constraint propagation?

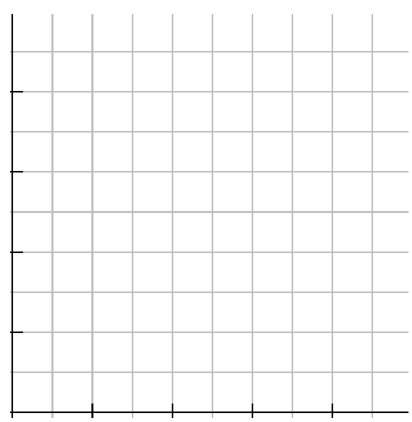
(b) Does a full constraint propagation solve the full CSP or do we still need to search for a solution?

23. **Linear Regression**

Consider the small dataset below. We would like to perform a linear regression $y = \beta_0 + \beta_1 x$.

Example	x	y
1	1	2
2	3	5.2
3	4	6.8
4	5	8.4
5	9	14.8
sum	22	37.2
mean	4.4	7.44

(a) **Without** using the formula derived in class, find β_0 and β_1. (Hint: plot the examples and report the calculation here).



(b) Solve for β_0 and β_1 using the formulas seen in class:

(c) What would be the predicted y of a new example with $x = 6$?