IMPORTANT.
The aim of these practice questions is to help you review.
It does not reflect the length nor the content of the exam
Please review all HW written questions.
Please, 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. What is an admissible heuristic?

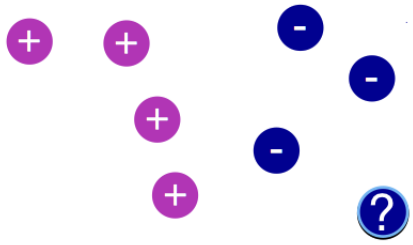
6. Give two examples of admissible heuristics and justify your answer.

7. Uniform-cost search is guaranteed to return an optimal solution.

8. Suppose h_1 is an A^* admissible heuristic. Suppose a set of possible heuristics of the form $h' = a \times h_1$ where a is a real number. Discuss the admissibility of the set of heuristics h'.

9. K-Nearest Neighbors

   1. For what minimal value of k will the point "?" be positive in the following labeled dataset?



   2. For classification with K-Nearest Neighbors, for an example x_q to be classified, assuming $N_k(x_q)$ is the set of the K-nearest neighbors of x_q.

$$\hat{y}_q = sign(\sum_{x_i \in N_k(x_q)} y_i)$$

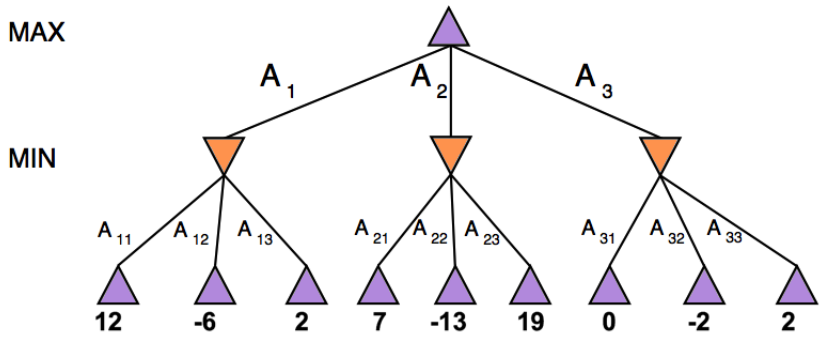
      It would be useful if we can somehow make the nearer neighbors contribute to the class of the query example more than the distant ones. Propose a weighting scheme and then reflect the change in the KNN formula.

10. "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.

11. 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."



12. 1. Using minimax, which of the three possible moves should MAX take at the root node?

2. What is the value of MAX at the root?

3. Using minimax with alpha-beta pruning, compute the value of alpha and beta at each node. What branches are pruned?

13. Naive Bayes & KNN: Compare the two methods and highlight the strengths and weaknesses of each.

14. 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.

15. 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

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.

16. Decision trees

We consider the following set of examples. Each example describes a skin tumor with three feature diagnoses: blood test, shape, and biopsy. The label is binary and represents whether the tumor is malignant or not (y/n).

	blood_test	**biopsy**	**shape**	**label**
e_1	high	positive	irregular	y
e_2	low	uncertain	regular	n
e_3	high	negative		y
e_4	low	positive	irregular	y
e_5	high	positive	regular	y
e_6	low	negative	regular	n
e_7	low	negative		n
e_8	low	positive	regular	n
e_9	low	uncertain	irregular	y
e_{10}	high	negative	regular	y

(a) Replace the missing values in the feature *shape* by the majority value of this feature over **all examples**. What value for *shape* do we give to examples e_3 and e_7? Show your work.

(b) Replace the two missing values as given in (a).

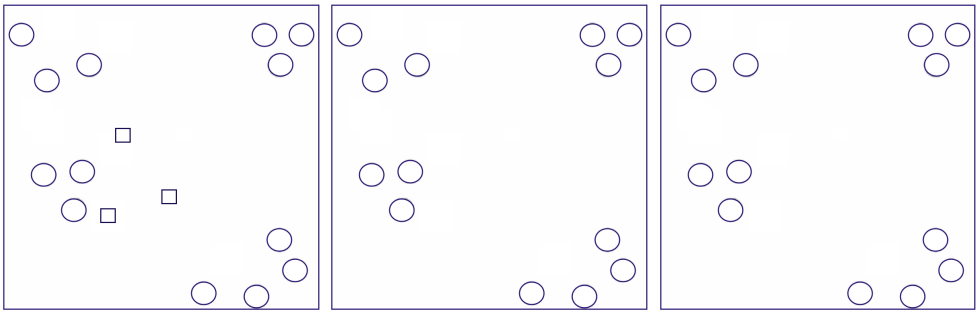
Derive **the first level ONLY of a decision tree** based on the method seen in class, using the Gini index (instead of entropy to ease calculations) and information gain. Draw the tree and detail all your results. Show your work.

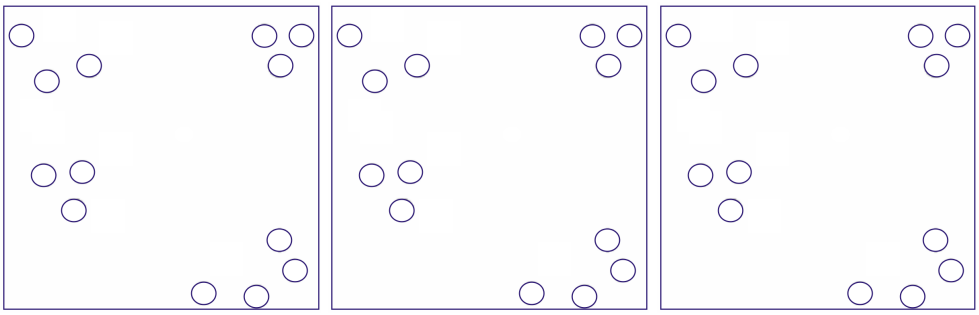
$$Gini = 1 - p_{\oplus}^2 - p_{\ominus}^2$$

$$Gain(S, A) = Gini(S) - \sum_{v \in \text{Values(A)}} \frac{|S_v|}{|S|} Gini(S_v)$$

(c) Features are not alike. A biopsy comes with a high cost (invasive and expensive procedure). How can one modify the information gain to include the cost of a given feature A? Explain your approach.

17. K-Means Algorithm 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.



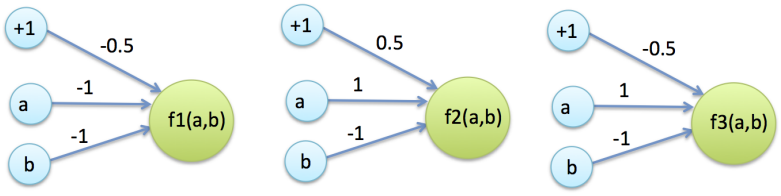


18. The idea of boosting is to train weak learners on weighted training examples. Check all that apply.
    - ◯ Give large weights to easy examples to get rid of them
    - ◯ Use any classifier as far as its accuracy is slightly worse than random
    - ◯ The classification output is a majority voting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

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

20. 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



- ◯ $f1$
- ◯ $f2$
- ◯ $f3$

21. Modus ponens/tollens?

$$\frac{\text{Linda is an excellent swimmer} \rightarrow \text{she can work as a lifeguard} \qquad \text{Linda is an excellent swimmer}}{\text{Linda can work as a lifeguard}}$$

( ) Modus ponens ( ) Modus Tollens

22. The CNF of $p \vee (\neg p \wedge q \wedge r)$ is:

[ ] $(p \vee q) \wedge (p \vee r)$

[ ] $(p \vee q) \wedge (\neg p \vee r)$

[ ] $(p \vee q) \vee (p \vee r)$

23. Given the following KB: $KB = \{p \vee \neg q, q \vee \neg r, q \vee r\}$ Using resolution for propositional logic, does $KB \models p$?

( ) Yes

( ) No

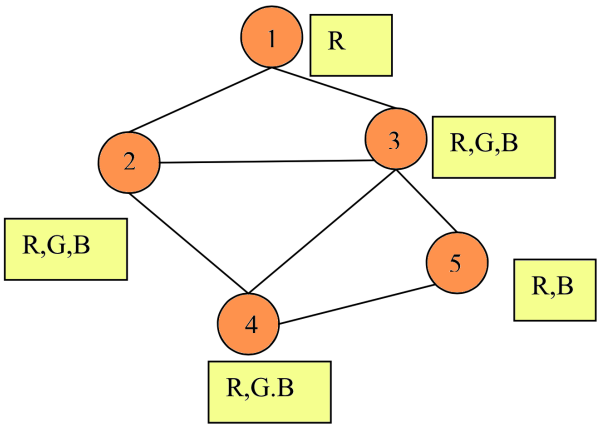
24. In checking arc consistency of $X \rightarrow Y$, deleting one value from domain(X) may enable further reduction in the domains of other variables Z such that $Z \rightarrow X$.

◯ True

◯ False

25. **CSPs**

Consider the following constraint graph for a graph coloring problem (the constraints indicate that connected nodes cannot have the same color). The domains are shown in the boxes next to each variable node.



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