

※ 考生請注意：本試題不可使用計算機。請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

1. (45%) In the following statements, please specify if the statement is **True** or **False**. If the statement is True, explain why it is True. If it is False, give correct answer or explain why.

- (1) A complete binary tree with a height of h can have more nodes than a full binary tree with a height of h .
- (2) Inserting an element to the beginning of an array (i.e., the $A[0]$ element) is more difficult than inserting an element to the beginning of a linked list.
- (3) Adding a constant to every edge weight in an undirected graph can change the set of edges that belong to the minimum spanning tree. Assume unique weights.
- (4) The output of the following Python code is: $[2,4,6,8,10]$

```
print(list(map(lambda x: x+1, range(0,10,2))))
```

- (5) Given the following confusion matrix generated from a classifier, the precision value is $8/20$.

		Answer	
		True	False
Prediction Results	True	8	3
	False	12	11

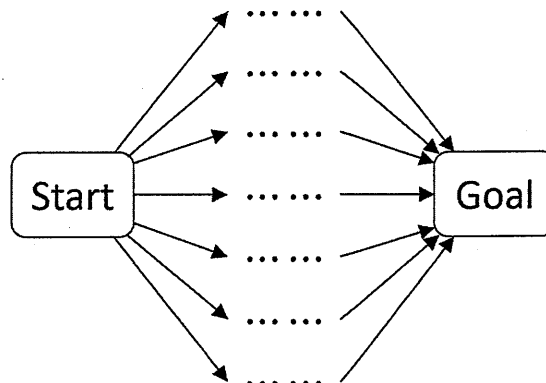
- (6) In database management systems, the term “identifier” is used to uniquely recognize every row in a data table.
- (7) In database management systems, suppose a “relation” contains five “attributes”, such as A, B, C, D, and E, the number of possible primary keys (e.g., BDE) is 10.
- (8) When talking about JSON (JavaScript Object Notation), a possible format structure is: $[id:'1', name:'do1']$
- (9) Network services over Internet, including FTP, WWW, PTT, YouTube live video streaming, and LINE text messages, use TCP as the transport layer protocol.
- (10) About public-key encryption, a message encrypted by the private key can be decrypted by the public key. Besides, a message encrypted by the private key can provide authentication.
- (11) Decision tree is linear model, but random forest is non-linear model.
- (12) Both Area under the ROC Curve (AUC) and Precision are valid performance measures for a classification model.
- (13) A density-based clustering algorithm can generate non-globular clusters.
- (14) Both data formats, *.xml and *.json, are well suited for nested data.
- (15) All of the following statements are reasonable for applying a **log** transformation on data.
 - (a) To help straighten relationships between pairs of variables.
 - (b) Bring data distribution closer to random sampling.
 - (c) To help visualize highly skewed distributions.
 - (d) Perform dimensionality reduction on the data.

- 2. (6%) Answer the following two questions on deadlock.
 - (a) Describe an approach to break an existing deadlock, and a way to prevent a deadlock from happening.
 - (b) Describe a real-life scenario that leads to a deadlock. In addition, in your scenario, you need to specify the processes and the resources they compete for.

- 3. (12%) Describe the differences between the following pairs of terms on operating systems.
 - (a) process vs. processor
 - (b) preemptive vs. non-preemptive
 - (c) CPU-bound vs. I/O-bound
 - (d) fork() vs. exec()

- 4. (8%) Assume you are searching for an element within this list:
[A, B, C, D, E, F, G, H, I, J, K, L, M, N, O].
 - (a) When using sequential search, which element to search will lead to the worst-case performance?
 - (b) When using binary search, which element to search will lead to the worst-case performance?

- 5. (6%) Given an illustrated graph as below, in which the dots indicate some part of the graph that you don't know exactly (note that they are not straight links). There could be many paths connecting the start node to the goal node. And some of the paths could be quite long. Now your task is to develop an BFS-based or DFS-based program that searches for a path and returns the first one whose end nodes are start and goal. It is not necessary to find the optimal path, just any path that connect the start to the goal. Analyze and compare the differences between BFS and DFS as the basis of your program.



6. (16%) Answer the following questions on data science.
- (a) Explain one similarity and two differences between linear regression and logistic regression.
 - (b) Explain three differences between machine learning and deep learning.
 - (c) Explain the idea of regularization in machine learning.
 - (d) Explain two strategies to prevent a decision tree model from overfitting.
7. (7%) Given a dataset, whose scatter plot is shown as below, give two machine learning models that are capable of perfectly modeling the Xs and Os, and justify your answer.

