Written re-exam

MED8

Algorithms, Data Structures and Software Engineering for Media Technology

Tuesday 28 August 2018

Name:_____

Cpr.no.: _____

Study no.: _____

Aalborg University | MSc in Medialogy | Med8

Algorithms, Data Structures and Software Engineering for Media Technology

Extra-ordinary Examination

28 August 2018

Instructions

- You have 3 hours to complete this examination.
- Neither electronic devices nor written material are allowed in the examination room.
- This examination consists of 10 questions. Each question is worth 10 marks. You must obtain at least 50 marks to pass.
- Do not write any answers on this question paper—answers written on the question paper will be ignored by the examiner. Write all your answers on the writing paper provided.
- Do not write your answers in pencil and do not use a pen with red or green ink. Use a pen with blue or black ink.
- Hand in no more than one answer to each question.
- Do not turn over until you are told to do so by the invigilator.

Question 1

For each of the following equations, state whether it is true or false.

a) $5n^3 + 2n\log_2(n^5) + 100 = \Theta(n^3)$

b) $9n^2\log_2(n^2) = \Theta(n\log_2 n)$

c) $4n^2\log_2 n = O(n^3)$

d) $2^n = O(n^2)$

e) $0.0001n^2 = \Omega(n^2)$

f) $4n^2 + 2n = o(n^2)$

g) $3n\log_2(n^2) = o(n^2\log_2 n)$

h) $5n^3 + n^2 + 4n + n\log_2 n + 1 = \omega(n^2\log_2 n)$

i) $4n^2\log_2 n = \omega(n^2)$

j) $n = \Omega\left(\sqrt{n}\log_2 \mathrm{n}\right)$

[1 mark for each correct part]

Question 2

The Master Theorem is stated as follows:

Theorem 4.1 (Master theorem)

Let $a \geq 1$ and $b > 1$ be constants, let $f(n)$ be a function, and let $T(n)$ be defined on the nonnegative integers by the recurrence

$$T(n) = aT(n/b) + f(n) ,$$

where we interpret n/b to mean either $\lfloor n/b \rfloor$ or $\lceil n/b \rceil$. Then $T(n)$ has the following asymptotic bounds:

1. If $f(n) = O(n^{\log_b a - \epsilon})$ for some constant $\epsilon > 0$, then $T(n) = \Theta(n^{\log_b a})$.
2. If $f(n) = \Theta(n^{\log_b a})$, then $T(n) = \Theta(n^{\log_b a} \lg n)$.
3. If $f(n) = \Omega(n^{\log_b a + \epsilon})$ for some constant $\epsilon > 0$, and if $af(n/b) \leq cf(n)$ for some constant $c < 1$ and all sufficiently large n, then $T(n) = \Theta(f(n))$. ■

Given the Master Theorem, as stated above, write down the order of growth in terms of Θ notation for each of the following recurrences.

a) $T(n) = 3T(n/4) + n \log_2 n$
b) $T(n) = T(2n/3) + 1$
c) $T(n) = 9T(n/3) + n$
d) $T(n) = 7T(n/2) + n^2$
e) $T(n) = 2T(n/2) + n$

[2 marks for each correct part]

Question 3

a) Which **one** of the following statements is true?
   A. The POP operation on a stack removes and returns the first element saved in the stack.
   B. The PUSH operation on a stack can lead to stack underflow.
   C. The ENQUEUE operation on a queue can lead to queue overflow.
   D. The DEQUEUE operation on a queue removes and returns the element at the tail of the queue.

b) Which **one** of the following statements is true?
   A. It takes constant time to search for a given key in a doubly-linked list.
   B. It takes linear time to find the minimum key in an unsorted doubly-linked list.
   C. It takes linear time to insert a new element at the head of a doubly-linked list.
   D. It takes quadratic time to delete an element from a doubly-linked list.

c) Which **one** of the following statements is true?
   A. It takes linear time to carry out an inorder tree walk on a binary search tree.
   B. It takes $\Theta(\log_2 n)$ time in the worst case to search for a value in a binary search tree.
   C. It takes $\Theta(n)$ time on average to find the minimum value in a binary search tree.
   D. It takes constant time to insert an element into a binary search tree.

d) Which **one** of the following statements is true?
   A. In the worst case, any comparison sort must make $O(n \log_2 n)$ comparisons.
   B. In the worst case, any comparison sort must make $\Omega(n \log_2 n)$ comparisons.
   C. In the worst case, any comparison sort must make $\omega(n \log_2 n)$ comparisons.
   D. In the worst case, any comparison sort must make $o(n \log_2 n)$ comparisons.

e) Suppose we have an array A of n integers such that the maximum value is *max* and the minimum value is *min* and $k = max - min + 1$. Suppose that we have enough memory to allocate an array of size k. Which **one** of the following algorithms can we use to sort A in linear time?
   A. Counting sort
   B. Bucket sort
   C. Merge sort
   D. Quicksort

[2 marks for each correct part]

Question 4

Study the following program and answer the questions that follow it.

```
1  #include<stdio.h>
2  #include<stdlib.h>
3
4  #define N 4
5
6  __global__ void kernel(int *dev_a) {
7      unsigned i = threadIdx.x + blockIdx.x * blockDim.x;
8      dev_a[i] = i;
9  }
10
11 int main(void) {
12     int *dev_a;
13     cudaMalloc((void**)&dev_a, N*N*sizeof(int));
14     kernel<<<N,N>>>(dev_a);
15     int a[N][N];
16     cudaMemcpy(a, dev_a, N*N*sizeof(int), cudaMemcpyDeviceToHost);
17     cudaFree(dev_a);
18     for(int i = 0; i < N; i++) {
19         for(int j = 0; j < N; j++)
20             printf("%3d",a[i][j]);
21         printf("\n");
22     }
23     return EXIT_SUCCESS;
24 }
```

a) What does this program print to the console?
b) In this program, how many thread blocks are there in each grid and how many threads are there in each block?
c) Is the array, a, (defined in line 15) stored in GPU memory or CPU memory?
d) Does the pointer dev_a point at a region of device memory or host memory?
e) In which line is the GPU memory deallocated?

[2 marks for each correct part]

Question 5

a) Write down the three dictionary operations.
b) In the context of hash tables, explain what is meant by the term *collision*.
c) Briefly describe one way in which collisions can be handled in a hash table.
d) What is the average asymptotic running time for dictionary operations in a hash table?
e) Explain the terms *load factor* and *simple uniform hashing*.

[2 marks for each correct part]

Question 6

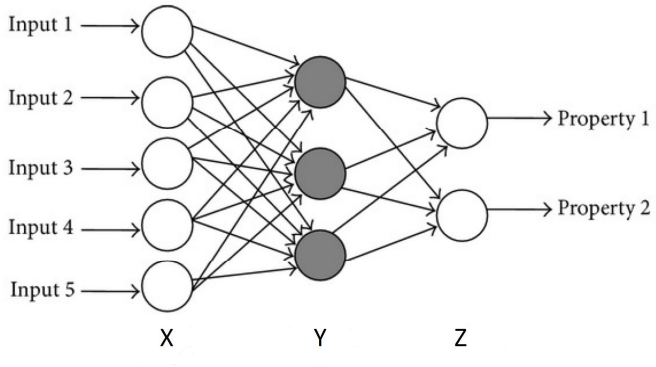
For each of the following statements, say whether it is true or false.

a) In a supervised learning model, a machine learning algorithm is trained with a labelled training set, producing a predictive model which is used to predict a label for previously unseen test data.
b) In k-fold cross validation, the same observation (data point) can appear in the test set in more than one fold.
c) Ensemble methods are used to combine models to increase classification accuracy.
d) Boosting is where the predictions from multiple classifiers are averaged.
e) Bagging emphasises observations misclassified by previous classifiers.

[2 marks for each correct part]

Question 7

a) Consider the following diagram which shows a multi-layer perceptron.



X, Y and Z need to be replaced with the correct labels for their respective layers in the network. Write down these three correct labels. [3 marks]

b) Study the following code and answer the questions that follow it.

```
1   from keras.models import Sequential
2   from keras.layers import Dense
3
4   model = Sequential()
5
6   model.add(Dense(64, input_dim=30, activation='relu'))
7
8   model.add(Dense(64, activation='relu'))
9
10  model.add(Dense(1, activation='sigmoid'))
11
12  model.compile(loss='binary_crossentropy',
13                optimizer='rmsprop',
14                metrics=['accuracy'])
15
16  model.fit(x_train, y_train,
17            epochs=20,
18            batch_size=128)
```

i. What kind of activation function is used in the input layer? Sketch this activation function. [2 marks]

ii. What is the defining characteristic of a *dense* layer? [1 mark]

iii. What kind of loss function does this network use? Name one other loss function that could have been used. [2 marks]

iv. How many nodes are there in the output layer of the network? [1 mark]

v. How many epochs is the network trained over? [1 mark]

Question 8

For each of the following statements, say whether it is true or false.

a) Mutation is an operator in an evolutionary algorithm that is applied to two or more individuals from the initial population (parents) and results in one or more new individuals (children).
b) Crossover is an operator in an evolutionary algorithm that performs a random alteration in a candidate solution and produces one new individual.
c) When working with an evolutionary algorithm, a candidate solution is also called a phenotype and the parameters of the candidate solution form its genotype.
d) The fitness function in an evolutionary algorithm defines what we mean by "improvement" and forms the basis for selection.
e) Mutation of a bit-string can be performed by flipping bits; mutation in integer and float genes can be performed by adding a unit Gaussian distributed random value to the chosen gene.

[2 marks for each correct part]

Question 9

a) Name and briefly describe the three models of cloud computing. Give an example of each model. [3 marks]
b) What is an *Amazon Machine Image* (AMI)? [1 mark]
c) What is *AWS Rekognition*? [2 marks]
d) What is an *AWS Region*? [2 marks]
e) What is *AWS S3*? [2 marks]

Question 10

a) What is *eventual consistency* and how is it used in cloud computing?
b) What is *database sharding*?
c) What is *AWS IAM*?
d) What is a *keypair* and where do you use it in AWS EC2?
e) What is *serverless computing* and what Amazon service can you use for this?

[2 marks for each correct part]

END OF EXAMINATION