# Written Re-Examination

## MED8 (AAL and ESB)

# Algorithms, Data Structures and Software Engineering for Media Technology

**23 August 2019**

Name:_____

Cpr.no.: _____

Study no.: _____

**Aalborg University | MSc in Medialogy | Med8 (AAL and ESB)**

# Algorithms, Data Structures and Software Engineering for Media Technology

**Re-Examination**

**23 August 2019**

## 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) $2n^3 + n\log_2 n + 90n = \Theta(n^4)$

b) $4n^3 \log_2 n + n^4 = \Theta(n^4)$

c) $\frac{n^3\sqrt{n}}{2}\log_2(n^5) = O(n^4)$

d) $2^{\sqrt{n}} = O(n^2)$

e) $n^{3.1} + 3 = \Omega(n^3)$

f) $n^5 + n^4 \log n = o(n^5)$

g) $\log n = o(\sqrt{n})$

h) $n \log n = \Omega(\log n)$

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

j) $n^4 = \omega(n^3)$

[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) $T(n) = 8T(n/2) + 4n^3\sqrt{n}$
b) $T(n) = 2T(n/4) + \log n$
c) $T(n) = 9T(n/3) + n$
d) $T(n) = 4T(n/4) + 2n$
e) $T(n) = 4T(n/2) + n^3$

[2 marks for each correct part]

## Question 3

The following algorithm, Algo1, takes a single argument, *A*, which must be an array of integers and returns a triple in which each of the three elements is an integer.

```
ALGO1(A)
1   Allocate two arrays, B and S, each of size |A|
2   if |A| == 0 return (0, 0, 0)
3   B[0] = 0, S[0] = A[0], best = 0
4   for i = 1 to |A| − 1
5       if S[i − 1] < 0
6           S[i] = A[i]
7           B[i] = i
8       else
9           S[i] = A[i] + S[i − 1]
10          B[i] = B[i − 1]
11      if S[i] > S[best]
12          best = i
13  return (B[best], best + 1 − B[best], S[best])
```

a) Write down the triple returned by Algo1 when *A* is equal to the following array: [-2, 2, -1, 4, -2, 2]
[6 marks]
b) Which computational problem does Algo1 solve? [2 marks]
c) Using asymptotic notation, write down the worst-case running time of Algo1. [2 marks]

## Question 4

Suppose you have a data structure, D, that supports the standard dictionary operations, INSERT, SEARCH and DELETE.

a) What is the worst-case running time for the SEARCH operation if D is a hash table and under what conditions does this worst case occur?
b) What is the average-case running time for the SEARCH operation if D is a hash table and under what conditions does this average case occur?
c) What is the worst-case running time of the SEARCH operation if D is a singly-linked list and under what conditions does this worst case occur?
d) If a hash table has *m* slots and stores *n* elements, what is its *load factor*?
e) Explain what is meant by the term, *simple unified hashing.*

[2 marks for each correct part]

## Question 5

Study the code below and answer the questions that follow it.

```
1  #include "cuda_runtime.h"
2  #include "device_launch_parameters.h"
3  #include <stdio.h>
4  void addWithCuda(int *c, const int *a, const int *b, unsigned int size);
5
6  __global__ void addKernel(int *c, const int *a, const int *b) {
7      int i = threadIdx.x;
8      c[i] = a[i] + b[i];
9  }
10
11 int main() {
12     const int arraySize = 5;
13     const int a[arraySize] = { 1, 2, 3, 4, 5 };
14     const int b[arraySize] = { 10, 20, 30, 40, 50 };
15     int c[arraySize] = { 0 };
16     addWithCuda(c, a, b, arraySize);
17     printf("%d %d %d %d %d\n",c[0], c[1], c[2], c[3], c[4]);
18     return 0;
19 }
20
21 // Helper function for using CUDA to add vectors in parallel.
22 void addWithCuda(int *c, const int *a, const int *b, unsigned int size){
23     int *dev_a = 0;
24     int *dev_b = 0;
25     int *dev_c = 0;
26     cudaSetDevice(0);
27     cudaMalloc((void**)&dev_c, size * sizeof(int));
28     cudaMalloc((void**)&dev_a, size * sizeof(int));
29     cudaMalloc((void**)&dev_b, size * sizeof(int));
30     cudaMemcpy(dev_a, a, size * sizeof(int), cudaMemcpyHostToDevice);
31     cudaMemcpy(dev_b, b, size * sizeof(int), cudaMemcpyHostToDevice);
32     addKernel<<<1, size>>>(dev_c, dev_a, dev_b);
33     cudaMemcpy(c, dev_c, size * sizeof(int), cudaMemcpyDeviceToHost);
34     cudaFree(dev_c);
35     cudaFree(dev_a);
36     cudaFree(dev_b);
37 }
```

a) In which line or lines of the code above is memory allocated on the device?
b) In which line or lines are arrays copied from the host to the device?
c) In which line or lines are arrays copied from the device to the host?
d) Which lines of code are executed on each thread on the device?
e) How many threads run in parallel on the device?

[2 marks for each correct part]

## Question 6

a) Explain briefly the difference between supervised and unsupervised learning. [2 marks]

b) Which of the following tasks are examples of supervised learning: *classification*, *regression*, *clustering*, *dimensionality reduction*. [2 marks]

c) Assume the following:
1. The probability of someone having meningitis is 1/50000.
2. The probability of someone having a stiff neck is 1/20.
3. 50% of people with meningitis suffer from a stiff neck.
If a person has a stiff neck, what is the probability that he or she has meningitis? [4 marks]

d) In which of the following cases is a Gaussian Naive Bayes classification algorithm a good choice:
i. When the naive assumptions actually match the data.
ii. For well-separated data when model complexity is less important.
iii. For low-dimensional data, when model complexity is important.
iv. For high-dimensional data, when model complexity is less important.
[2 marks]

## Question 7

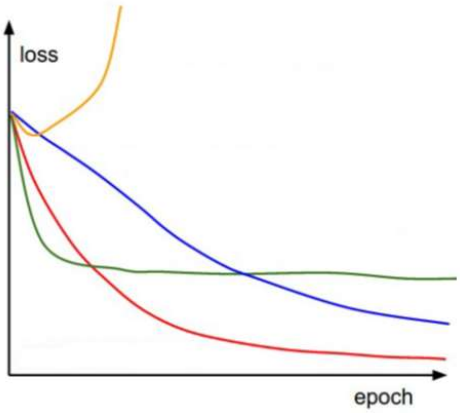
a) With an example, briefly explain the *kernel trick* and when it is used. [4 marks]

b) Briefly describe two ways in which multiple two-class SVMs can be combined to give a multi-class SVM. [4 marks]

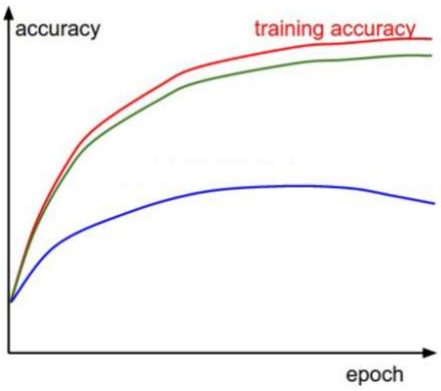
c) Briefly explain what is meant by the *curse of dimensionality.* [2 marks]

## Question 8

a) In a convolutional neural network (CNN), explain what purpose is served by a *pooling layer.* What are the most common types of pooling layer? [3 marks]
b) In a CNN, explain what a *fully connected layer* is and what it is typically used for. [3 marks]
c) In the graph below, which curve represents the highest learning rate and which the lowest learning rate? [2 marks]




d) In the graph below, which curve represents a situation where there is strong overfitting? [2 marks]




## Question 9

a) For each of the following design patterns, state whether it is *creational*, *structural* or *behavioral*:
i. Factory method (107)
ii. Singleton (127)
iii. Decorator (175)
[3 marks]
b) In software engineering, explain what is meant by *programming to an interface.* What are some of the advantages of this strategy? [3 marks]
c) Explain the difference between *inheritance* and *composition*. Which of the two is sometimes called "white-box reuse" and why? Which of the two is generally preferable and why? [4 marks]

## Question 10

a) In the Agile Manifesto, which of the following are valued more highly:
- Individuals and interactions
- Comprehensive documentation
- Processes and tools
- Customer collaboration
- Working software
- Responding to change
- Contract negotiation
- Following a plan

[2 marks]

b) Extreme programming structures the software development process as a system of nested planning/feedback loops operating at different time scales, ranging from seconds to months. Give 4 examples of these feedback loops and state the approximate timescale on which each operates. [4 marks]

c) In the SCRUM methodology,
i. Who is responsible for the product backlog?
ii. How long, typically, is a sprint?
iii. How often does the SCRUM team meet and for how long?
iv. Who is responsible for making sure that the SCRUM framework is followed?
[4 marks]

END OF EXAMINATION