COMPUTER SCIENCE TRIPOS Part IB

Tuesday 4 June 2013 1.30 to 4.30

COMPUTER SCIENCE Paper 4

*Answer **five** questions.*

*Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.*

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Artificial Intelligence I

(*a*) Describe the basic algorithm for performing *local search*. Give an example of a search problem for which it is an appropriate solution, and an example of a search problem for which it is not. When would an algorithm such as $A^\star$ search be preferred? [7 marks]

(*b*) Give two examples of a situation that might reduce the effectiveness of the basic local search algorithm. For each example you give, describe in detail a modification to the basic algorithm that might be used to overcome it. [8 marks]

(*c*) It is suggested to you that an application previously addressed as a *constraint satisfaction problem* might alternatively be solved using some variant of local search. Is this a reasonable suggestion? If it is, then outline a way in which it might be achieved. If not, then provide a reasoned discussion explaining why. [5 marks]

2 Artificial Intelligence I

We wish to solve a *supervised learning* problem using a *perceptron* computing the function

$$h(\mathbf{w};\mathbf{x}) = \sigma\left(\mathbf{w}^T\mathbf{x}\right)$$

where $\mathbf{w}$ is a vector of *weights*, $\mathbf{x}$ is a vector of *features* and $\sigma(z) = 1/(1+e^{-z})$. We have a set of m *labelled examples* $\mathbf{s} = ((\mathbf{x}_1, o_1), \ldots, (\mathbf{x}_m, o_m))$ where $o_i \in \{0,1\}$.

(*a*) Derive the *gradient descent training algorithm* for training the perceptron by minimizing the *error function*

$$E(\mathbf{w}) = \sum_{i=1}^{m} \left(o_i - h(\mathbf{w};\mathbf{x}_i)\right)^2 .$$

You may if you wish employ the result

$$\frac{d\sigma(z)}{dz} = \sigma(z)(1-\sigma(z)).$$

[7 marks]

(*b*) We are now told that some training examples are more important than others, and it is thus more important that, after training, there is only a small difference between o_i and $h(\mathbf{w};\mathbf{x}_i)$ for these examples. Derive a new version of the training algorithm that takes this modification into account. [6 marks]

(*c*) Having trained a classifier $h(\mathbf{w}_{\text{opt}};\mathbf{x})$ in part (*a*) using the training data available, a colleague presents you with a second classifier $h'(\mathbf{w}'_{\text{opt}};\mathbf{x}')$. Your colleague has trained this classifier using the same number of examples and the same labels, but a different collection of features, so for their classifier the training data was

$$\mathbf{s}' = ((\mathbf{x}'_1, o_1), \ldots, (\mathbf{x}'_m, o_m)).$$

Devise a way in which you might perform further training in order to *combine* the two classifiers h and h' into a single, possibly more powerful, classifier.

[7 marks]

3 Computer Graphics and Image Processing

(*a*) Explain the difference between an *explicit formula*, a *closed form* and a *parametric form* for a curve in two dimensions. Give examples to illustrate your answer. [3 marks]

(*b*) Explain the term *mathematical continuity* (C_n) when joining two curves. [2 marks]

(*c*) Give the formulation of a cubic Bézier curve in two dimensions, explaining the rôle of the parameter and control points. [4 marks]

(*d*) Consider the joint between two cubic Bézier curves. State and prove constraints on their control points to ensure:

(*i*) C_0 continuity at the joint; [2 marks]

(*ii*) C_1 continuity at the joint; [4 marks]

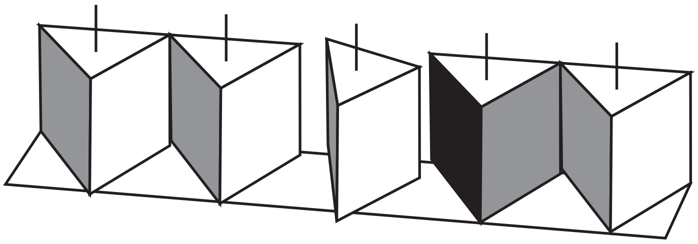
(*iii*) C_2 continuity at the joint. [3 marks]

(*e*) Discuss the implications of requiring C_3 continuity at the joint between two cubic Bézier curves. [2 marks]

4 Computer Graphics and Image Processing

(*a*) Describe, in detail, an algorithm to perform error diffusion on a greyscale image. Your algorithm should take a greyscale image, with eight bits per pixel, and convert it to a black and white image, with one bit per pixel, at the same resolution. [8 marks]

(*b*) An inventor produces a display where each pixel can have one of three values: white, mid-grey, or black. Such a display can be built by, for example, using rotating triangular blocks of painted wood. The figure shows the back view of a row of five pixels with the central pixel turning. From left to right the pixels are showing, to the front side: black, black, turning, white, black.



Modify your algorithm in part (*a*) to handle these three-valued pixels. [4 marks]

(*c*) Describe, in detail, the modifications required to turn the display described above into a colour display. Your display, through use of an appropriate error diffusion algorithm, should be able to display error-diffused versions of 24-bit RGB colour images. [8 marks]

5 Databases

The lectures defined Boyce-Codd Normal Form (BCNF) as follows. A relational schema R is in BCNF if for every functional dependency $\mathbf{X} \rightarrow A$ either

- $A \in \mathbf{X}$, or
- $\mathbf{X}$ is a superkey for R

(*a*) Present a relational schema (with functional dependencies) that is not in BCNF and explain how BCNF is violated. [3 marks]

(*b*) Describe a problem that could be encountered in a database implementing your schema. [3 marks]

(*c*) Decompose your schema into smaller relations that are in BCNF. Justify your answer. [3 marks]

(*d*) Discuss one cost and one benefit involved in the kind of schema normalisation performed in (*c*). [2 marks]

(*e*) Is every BCNF schema free from the problem you described in (*b*)? Explain your answer. [3 marks]

(*f*) Describe a scenario in which the relations you describe in (*c*) are derived directly from an Entity/Relationship (ER) model. [6 marks]

6 Databases

(*a*) Explain how *data duplication* and *data redundancy* are distinct concepts in relational database design. [2 marks]

(*b*) What is meant by the term *on-line transaction processing* (OLTP)? [2 marks]

(*c*) What is meant by the term *on-line analytic processing* (OLAP)? [2 marks]

(*d*) Describe how the task of designing an OLTP system would differ from that of an OLAP system. [4 marks]

(*e*) Suppose that an OLAP database contains data extracted from an OLTP database.

(*i*) How might the presence of NULL values in the OLTP database complicate the task of data extraction? [5 marks]

(*ii*) How might schema migration of the OLTP database complicate the task of data extraction? [5 marks]

7 Economics and Law

(*a*) Explain the criminal offences created by the various sections of the Computer Misuse Act. [12 marks]

(*b*) What offences are likely to be committed in the following circumstances?

(*i*) Alice recruits 10,000 volunteers who run software she has written to protest against the banking industry. The software repeatedly sends transactions to a payment network gateway. The transactions are not valid, so no money is moved. However, the transaction volume is such that the gateway cannot process valid transactions from merchants. [4 marks]

(*ii*) Bob writes software which does the same as Alice's. However, instead of having it run by volunteers, he hacks an adult website and gets its customers to install and run his software in the mistaken belief that it is codec that they need in order to watch the site's content. Otherwise his protest has exactly the same effect as Alice's. [4 marks]

8 Security I

(*a*) In the Galois field GF(2^8) modulo $x^8 + x^4 + x + 1$, calculate

(*i*) the difference 1100 1010 minus 1001 0011; [2 marks]

(*ii*) the product 0100 1011 times 0000 1001. [6 marks]

(*b*) Briefly explain two advantages that arithmetic in GF(2^{128}) has over arithmetic in $\mathbb{Z}_{2^{128}}$ when designing cryptographic algorithms. [6 marks]

(*c*) Given a block cipher E_K and a corresponding decryption function D_K, provide a formula for the decryption of the following modes of operation and state for each whether the E_K or D_K calculations required during decryption can be executed in parallel: CBC, OFB, CTR. [6 marks]

9 Security I

(*a*) While inspecting the discretionary access-control arrangements on a Unix computer, you find the following setup:

Members of group `staff`: `alex`, `benn`, `cloe`

Members of group `gurus`: `cloe`

```
$ ls -ld . * */*
drwxr-xr-x  1 alex staff    32768 Apr  2  2010 .
-rw----r--  1 alex gurus    31359 Jul 24  2011 manual.txt
-r--rw--w-  1 benn gurus     4359 Jul 24  2011 report.txt
-rwsr--r-x  1 benn gurus   141359 Jun  1  2013 microedit
dr--r-xr-x  1 benn staff    32768 Jul 23  2011 src
-rw-r--r--  1 benn staff    81359 Feb 28  2012 src/code.c
-r--rw----  1 cloe gurus      959 Jan 23  2012 src/code.h
```

The file `microedit` is a normal text editor, which allows its users to open, edit and save files.

(*i*) Draw an access control matrix that shows for each of the above five files, whether `alex`, `benn`, or `cloe` are able to obtain the right to read (R) or replace (W) its contents. [12 marks]

	manual.txt	report.txt	microedit	src/code.c	src/code.h
alex					
benn					
cloe					

(*ii*) Which users have at least all the access rights of which other users? [2 marks]

(*b*) Explain briefly *three* mechanisms that the operating system kernel of a desktop computer can use to generate unpredictable numbers for use in cryptographic protocols as soon as it has booted. [6 marks]

END OF PAPER