Semester 2 Examinations 2018/2019

Exam Code(s)	1CSD, 1SPE, 2SPE
Exam(s)	MSc in Computer Science (Data Analytics) 1st and 2nd Structured PhD
Module Code(s)	CT5107
Module(s)	Advanced Topics in Machine Learning and Information Retrieval
Paper No.	1
External Examiner(s)	Prof. Pier Luca Lanzi
Internal Examiner(s)	* Prof. Michael Madden * Dr James McDermott

Instructions: Answer 2 questions from each section (4 in total).
All questions carry equal marks.

Duration	2 hours
No. of Pages	5
Discipline(s)	Information Technology
Course Co-ordinator(s)	Dr Conor Hayes (CSD)

Requirements:

Release in Exam Venue	Yes	X	No	
MCQ Answersheet	Yes		No	X
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	None			
Graphic material in colour	Yes		No	X

[PTO]

Section A

1. (a) Draw a diagram showing all nodes and weights in a fully-connected feed-forward neural network that has 2 input nodes, 2 hidden nodes and one output node, including additional nodes/weights for all bias terms. Annotate your diagram with appropriate notation for nodes and weights. Write down a full set of equations for forward propagation of inputs to outputs, following your notation. (You can choose to use either vector or scalar notation.) [6]

(b) For your network from Part (a), provide a set of weight values to generate the XOR function listed below, and provide calculations to demonstrate how the weights map the inputs to the output. Specify what activation function you are using. In this case, are non-zero biases needed? [6]

Input A	Input B	Output A xor B
0	0	0
0	1	1
1	0	1
1	1	0

(c) Explain what both the sigmoidal and ReLU activation functions are. In each case, provide an equation for the function and an equation for its derivative, and draw a sketch of the function. Define all terms used in the equations. [6]

(d) Building on the previous parts of this question, provide a detailed description of the backpropagation algorithm for learning weights of a feed-forward neural network from training data, including all relevant equations, with explanations of all terms used. [7]

2. (a) Explain what a convolution matrix is and how it is applied to a 2D image to produce convolved features. Building on this, describe the architecture of a convolutional neural network, illustrating your answer with a diagram. In addition, identify which problem(s) encountered in classic shallow neural networks are addressed by this approach. [8]

(b) In the context of deep neural networks, explain each of the following terms in detail:

(i) Pooling layer
(ii) Stride length
(iii) Epoch and Batch size
(iv) Softmax layer. [8]

(c) *"Transfer learning has emerged as an important technique in deep learning"*. Explain what transfer learning involves and describe, with examples, three different forms of transfer learning. [6]

(d) One challenge facing deployment of current convolutional neural networks is that they can be "fooled". Discuss, with reference to some examples, what fooling means in this context. Why does this happen? What, if any, are the associated risks/problems? [5]

[PTO]

3. (a) You have received the following email message from a psychologist working with high-dimensional brain image data. Prepare a detailed reply.

"I am working with MRI data from 200 people: this is brain image data where each image has 3D pixels (called voxels) of resolution 100x100x100. For each person, I have their aptitude at performing a task (a number in the range 0-100) and I want to build a model to try to predict aptitude from the MRI images. For related tasks, people are using CART Random Forests. I have heard of CART, but can you explain in detail what random forests are? I believe these are related to ensembles, but what exactly is an ensemble, and how do they improve performance over using a single CART tree? Will this random forest approach work in my case, where I have such a large amount of data from each person, relative to the number of people? Also, would you recommend a deep learning approach instead, and if so why?" [12]

(b) Using a diagram, explain the terms *support vector, margin, slack vector,* and *maximum margin hyperplane* as applied to Linear Support Vector Machines. [6]

(c) What is a *convex optimisation problem* and what makes such problems easier to solve, relative to optimisation problems that are non-convex? Describe how gradient descent may be used to solve convex optimisation problems. In each of these cases, can learning a hypothesis from data be formulated as a convex optimisation problem: (i) Logistic Regression; (ii) Support Vector Machines; (iii) Feed-Forward Neural Networks? [7]

[PTO]

Section B

4. Let's suppose we wish to build an **image-matching** method, and we have decided that each image will be encoded as a 3D colour histogram.

(a) Given a method of extracting histograms, explain why a method to measure the *distance* between a pair of histograms is all we need to complete our image-matching method. [5]

(b) Give the formula for the chi-squared distance $X^2(g, h)$ between two one-dimensional histograms g and h, and explain the purpose of the epsilon (ε) term. [5]

(c) Explain two situations where our method would return a "false positive": an image which is not truly relevant to the query. [5]

(d) Data augmentation is a common idea in training convolutional neural networks. Would this be useful in our colour histogram approach? Explain why or why not. [5]

(e) Describe briefly one realistic application of image-matching. [5]

5. **Deep convolutional neural networks** are very important in modern information retrieval on images.

(a) For a deep convolutional network for face recognition, performance can be improved by adding some non-neural network aspects to our processing pipeline. What are these and why are they needed? [5]

(b) How can a network trained on one image recognition task be changed or adapted to work on another? [5]

(c) In the context of convolutional neural networks, explain Dropout and explain why it is useful. In *Keras*, what code would you write to use Dropout? [5]

(d) In the context of face recognition, explain the terms *embedding* and *triplet training* and how they contribute to the face recognition task. [5]

(e) Given an input image of size 100x100 and 1 channel (i.e. greyscale), and given a first convolutional layer of 16 filters with a 3x3 kernel, stride 1 and padding 1, say what size and shape the output of the convolutional layer will have. Show workings. [5]

[PTO]

6. As discussed in lectures, Shazam is a well-known example of **music information retrieval**.

(a) What are the main difficulties of the Shazam use-case? [5]

(b) Describe an approach (not based on neural networks) by which a system like Shazam could be implemented, mentioning how songs are represented in the database, and how the matching is carried out. You may ignore implementation issues such as user interface, network connection, and databases. [10]

(c) "A convolutional neural network could be used to create a service similar to Shazam." Discuss this statement. Your answer should mention: differences and similarities between this use of convnets and more typical uses on image recognition tasks; possible problems; and possible solutions. [10]

[END]