Examiners' commentary 2017–2018

CO3311 Neural networks – Zone B

General remarks

As usual with this course the examination covered all the major topics in the syllabus and allowed candidates to demonstrate their understanding of the applications and the limitations of the variety of networks studied, as well as their ability to perform the necessary calculations to train or develop a specific example.

Of the questions on the examination paper Question 1 proved the most popular with most candidates. It was also the question for which highest average marks were obtained.

After Question 1 , Questions 3, 4 and 5 were most rewarding in terms of marks. These were also the most popular questions after Question 1.

Question 2 was less popular than those mentioned above and less well answered.

Question 6 , the least popular question, was also the least well answered of all the questions.

Comments on specific questions

Question 1

Question1 tested candidates' understanding of the motivations for studying artificial neural networks, the basic terminology used, and the architecture and types of learning associated with them.

Part (a) asked for an account of the three motivations for studying ANNs given in the subject guide. Most candidates were able to score very well on this part by discussing the philosopher's, psychologist's and engineer's rationales.

In part (b) candidates were asked to define the terms ***network***, ***weight***, ***activation function*** and ***bias*** as they relate to ANNs. Whilst there were many good answers, some were spoilt by lack of clarity. Candidates should know these basic definitions!

The term 'architecture' of an artificial neural network has a technical meaning, and part (c) required candidates to describe the features of a neural network that we need to specify when giving an architecture. Good answers gave a definition as covered in the subject guide, whilst poorer answers gave a less precise and rather disorganised account of the requirements.

The final part, part (d), asked candidates to compare and contrast supervised and unsupervised learning. This should be an easy question to answer and most candidates had no difficulty. However, the question also required examples to be given to illustrate where each type is appropriate and the types of network that each type of learning can be applied to. Many candidates ignored these requirements. Consequently, this part proved to be the hardest, extending beyond descriptive bookwork to demonstrate understanding.

Question 2

Perceptron networks were the subject of Question 2, testing candidates on their understanding of the limitations and possibilities of this type of network. This question was not well answered generally.

Part (a) asked for examples and a diagram explaining the limitation of a single Perceptron. Besides issues of separability, good answers mentioned practical issues of the computer's range of numbers and the fact that they have only discrete outputs. However, few answers to this part gained more than half of the available marks.

The fact that Perceptrons can be used to build any Boolean function gives them a universal property. Part (b) required candidates to explain what this means and illustrate by means of the XOR function as an example. Good answers did both of these – explaining the universal property and giving the example of XOR. Poor answers tended to focus on the XOR without regard to the general case.

The final part, part (c) required candidates to show their understanding of how this works in practice, with an example of the design a two-input network of Perceptrons (threshold units) which produces an output of 0 if and only if both of its inputs are between -0.7 and 0.7. Although not trivial, it is easy to design such a network once the problem is broken down into:

- testing for a single input being greater than -0.7
- testing for an input being less than 0.7; and
- then putting these together to test for both and doing this for each of two inputs.

An explanation of how the design achieves its goal was also asked for, but some candidates omitted to provide this. Part (c) was poorly answered on average with very few candidates achieving more than half of the available marks.

Question 3

This question focused on Back Propagation, which continues to be a mainstay of much work on artificial neural networks.

Part (a) looked at the activation functions used during the course. Candidates were asked to list these and to give a diagram showing their form and an expression for each. Good answers used the notation of the subject guide and labelled their diagrams carefully. Weaker answers often lacked labelled axes and gave very brief accounts of the functions. Candidates gained most marks for this part of the question on average.

Back Propagation networks have their difficulties and part (b) required candidates to describe two of these – network paralysis and overfitting. The steps that can be taken to avoid or overcome these problems were also required for full marks. Good answers talked about scaling and about dividing training examples into three sets, training, validation and testing. Weaker answers often omitted validation and were less clear in their explanations.

A straightforward calculation was required for part (c). Given the weights as shown in the given figure, candidates were to calculate the weights after training with the given examples. Whilst most were able to achieve good marks, some answers showed a lack of understanding of the notation and some even seemed unsure of the architecture of the situation presented – a single Backpropagation unit. This is easy to infer from the weights given! As is often the case, there were some slips in candidates' calculations, but marks were awarded for the methods used when these were clear. Despite the straightforwardness of this part of the question, it was the least well answered.

Question 4

Kohonen-Grossberg Networks are covered during this course to exemplify unsupervised learning. They are the main topic of Question 4.

Firstly, in part (a) candidates were asked to explain how Kohonen-Grossberg Networks differ from Perceptron Networks. Not only their function but the activation functions (winner takes all) and learning algorithm are required in a good answer. The concept of layer is also quite different in these networks.

An important part of the usual learning algorithm for such networks includes normalisation. It is important that candidates understand its use, application and limitation. Demonstration of this was the topic of part (b). Good answers explained what normalisation is and how it is not applicable if zero length examples occur in the possible training or testing sets. They also included the point of simplifying the search space and allowing the inner product to be used instead of distance. Poorer answers missed out one or more of these important points. This was the least well answered part of Question 4 .

Part (c) asked for the training algorithm for the Kohonen layer of such a network. Good answers gave formulae as well as explanation and description of the terms in the formulae. The winner takes all nature of the activation function is also a vital point, and was sometimes omitted in weaker answers. This was by far the best answered part of this question.

Although the phrase Kohonen-Grossberg Network is used to describe these networks, often the Grossberg layer is omitted. This may be because it is unable to condition the outputs in a way that is useful. Postprocessing of output is now more commonly done separately. However, we still mention it for completeness and in order to link with early work on these networks. This was the topic of part (d). Most answers were excellent.

Part (e) required the learning algorithm to be put to use in a simple example. Four training examples were given, and this seems to have taxed some candidates. However, the weights and training set were chosen to simplify calculations in the expectation that the full calculations would reduce arithmetic errors and still be easily completed in a reasonable time. Many candidates were not able to complete the calculations successfully.

Question 5

Hopfield networks are the archetypal recurrent network type studied on this course, and form the subject of Question 5.

Starting with some definitions in part (a), candidates were expected to be familiar with the terms such as ***state***, ***predecessor***, ***energy*** and s***ynchronous*** that we meet when using these networks. Most candidates produced reasonable accounts of these ideas. Answers tended to be on the brief side omitting, for example, the formulae for energy or the use of the term state for each unit and for the whole network.

Unlike most other network types considered on this course, the firing scheme used for Hopfield nets is random asynchronous and this is an important feature that ensures stability of the evolution of the network. Some candidates did not state that asynchronicity is required for a network to end up in a sink state when answering part (b).

Another set of constraints on Hopfield networks is on the allowable weights. Good candidates were able to recall, for part (c) that symmetry of the weight table with zeros down the diagonal is required. This was the least well answered part of this question, with many candidates receiving zero marks.

The final two parts, parts (d) and (e) required a straightforward set of calculations to work out the state transition table and from it the state transition diagram of a network with a given weight table. Initial weights in the question were chosen to ease calculations. Many candidates achieved full marks for part (d).

Most candidates completed the exercise without error though there were a number who made simple arithmetic slips. A few candidates strayed from the notation introduced in the course and suffered for this by confusing the output states. Some candidates' state transition diagrams lacked circles around states that transition to themselves, but otherwise most answers followed the table.

Question 6

The final question explored candidates' knowledge of applications and how the nature of these affects the choices of artificial neural networks used on them. Few candidates attempted this question and on the whole those answers were poor, though there were a few very good answers.

Three applications were considered in this question: (i) Predicting the weather at a particular location, (ii) Classifying a new type of plant and (iii) Finding patterns in data on purchases made at a chain of supermarkets.

Four network types are also considered: Perceptron, Backpropagation, Kohonen-Grossberg and Hopfield networks.

Candidates were first asked, in part (a) to describe the sort of data that is likely to be available and in what quantities for the applications. Good answers would give an account of historical data that might be available, and an idea of what data might be easily collected. Another important factor is whether the data is likely to be analogue or digital.

Having considered data availability, candidates were then asked to give the pros and cons of using each of the given network types in each of these applications. Poor answers just gave the candidates' choice of network for each application. Good answers explained how the data listed in part (a) fitted or did not fit with the requirements of each of the network types. Excellent answers went on to explain how data might be enhanced to fit in with the network types.

A simplistic view, and one that is acceptable under exam conditions, is to look at the applications as either time series analysis, classification or correlation and make this the basis of the discussion.

To achieve good marks answering questions such as these, candidates should revise the details of applications. Familiarity with the reading recommended by the guide is essential for answering application-based questions, and it is always helpful to keep up with developments in the field for additional credit.