Examiners' commentaries 2015–16

CO3311 Neural networks – Zone A

General remarks

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

Question one on Perceptrons, question three on Kohonen-Grossberg networks and question five covering all five network types were the most popular, with most candidates attempting them. Question six proved the least popular question on the paper, despite its link with the coursework.

Comments on specific questions

Question 1

This question had four parts on Perceptrons with each carrying a maximum of five marks except for part d) which required two calculations worth five marks each.

For part a) candidates were asked to explain how a Perceptron may be described as representing 'a line in the plane'. This is mainly bookwork as chapter three of the subject guide covers this in detail. Candidates who had worked conscientiously through the guide should have had little or no difficulty with this. Good answers showed how weights, including the bias, are used to form the activation through the threshold function to split the plane into two. It is essential to relate the equation of the line in the plane with the equation for threshold activation. Good answers also mentioned that the line is for two input units and more inputs required higher dimensions to illustrate what is going on. The question asked for a diagram and marks were deducted if this was omitted. It was disappointing to see some answers that did not mention weights. There were a number of very poor answers that showed candidates had not understood the mathematics behind these simplest of units.

Part b) required the design of a network of threshold units which has four inputs and which outputs a one if and only if an odd number of these inputs are one. There are, of course, many possible solutions to this but good answers gave a design and went on to show that it worked where zero, one, two, three and four of the inputs are one. On the whole this part of the question was not very well answered. Requiring a particular number of inputs to be one requires more than one layer, and this may have been the reason for the difficulty that some had in forming a good answer. Chapter four of the subject guide explains how an arbitrary truth table can be modelled by a feed-forward network and this might have provided the easiest way for candidates to proceed.

Part c) was designed to allow the candidate to show their understanding of the principles used in part b). They were expected to explain how their answer to part b) could be extended to include more inputs. Few were able to verbalise this coherently.

In part d) candidates were given the initial weights and bias of a single two-input threshold unit. The question simply required them to calculate both forward and backward passes. Good answers showed the calculations involved and, noting that no learning rate was specified, chose a sensible value for this. Unfortunately, some candidates used a sigmoid unit while others omitted the bias or failed to update it, despite threshold unit and bias value being clearly given in the question. This use of the sigmoid shows a lack of understanding of the basic differences between backpropagation networks and Perceptron networks.

Question 2

This question asked candidates to compare and contrast two network types: backpropagation networks and Perceptron networks. Overall it was the best answered question: around 50 per cent of candidates attempted it. The question listed five criteria for comparison and each of these carried a maximum of five marks. Not all candidates were able to put their knowledge down on paper in a coherent and logical manner but most attempts were good. Candidates need to know the different types of unit normally used in all of the network types that we deal with. In this case, these are sigmoidal and threshold units. Naming these is not enough. Each different type of neural network has its own training algorithm and full answers would give formulae and details about how these work. Though there were some very good answers, there were some that left out details of the algorithms and thus lost many marks. There were three major points to be noted.

- Some inputs may be discrete while others are continuous, but this is not a feature of the network. Output values are, however, and threshold units can only output zero or one though it is possible to define other types.
- The beauty of Perceptrons is their ease of use in that the learning algorithm is guaranteed to converge if a suitable weight set exists. This is of course countered by the fact that continuous output is not achievable. Backpropagation achieves more power but at the cost of not having guaranteed convergence. Thus the range of applications tends to be much richer for backpropagation than for Perceptron networks.
- Candidates were advised that their answers should describe the effects of these differences and most were able to answer in these terms. It is vital that in a compare and contrast question both similarities and differences between networks are explored. It is not sufficient just to describe each, leaving the comparison to the reader.

Question 3

This question, which was generally reasonably well answered, asked for details of the working of Kohonen-Grossberg networks and tested candidates' knowledge of these, the reasoning behind some of the details and also their ability to work through a very simple example.

Part a) asked for a description of the architecture of such networks and required a diagram for full marks. Good answers included a clearly labelled diagram with a note on each component. The question also required an explanation of the action of the two independent layers. On the whole, candidates seemed to understand the rationale for each layer but were less well able to explain this for the Grossberg layer. The concept of 'winner takes all' is an important feature of the Kohonen layer and answers that omitted to mention this lost marks.

Some candidates forgot that normalisation is an essential step in the working of the Kohonen network. It allows one to use the inner product over the distance making the calculations quicker. The omission of this fact seems to suggest that candidates are not looking at questions as a whole. A previous part of the question should have served as a reminder to candidates about the importance of normalisation.

The example given in part c) was designed to make calculations reasonably easy. Some did not remember to normalise the inputs and the units and some did not recognise that neither (1, 1) nor (-1, -1) are normalised.

Part d) threw some candidates. These individuals did not remember that the action of training a Kohonen network is to move the units towards cluster centres. The diagram required was simply one where the units' pre- and post-training were marked and some sort of notation used to show this movement.

The final part (e) was based on bookwork, in that it asked how initial classes are chosen in applications. Valid points to be made include the fact that there is no guaranteed method, that random choices are often made and that even the number of classes is something for experimentation. Most candidates were able to state this.

Question 4

This question concerned one of the iconic applications of Hopfield nets, that is, in finding solutions to optimisation problems. Weights are set up so that the network's energy is minimised by finding the solution of a problem and then from a random starting point allowing the network to minimise energy.

On the whole, part a) of the question was not answered well where candidates did not remember that weights are set up to reflect the distances involved in the travelling salesman problem.

Although most candidates knew the algorithms for producing a state transition table from a set of weights, there was a great variety in the quality of descriptions written for part b). Candidates need to ensure that they are able to explain as well as execute the major algorithms covered in CO3311.

Part c) was well answered, with most candidates able to produce the state transition table and from that a transition diagram. Some candidates forgot to circle states that had self-transitions and a few confused the order of units when producing the tables. Examiners gave marks where the method was correct, with minor reductions for the occasional slip in implementation.

Question 5

This question tested the candidates' knowledge of Perceptrons, backpropagation networks, Kohonen-Grossberg networks, Hopfield networks and Boltzmann machines in terms of algorithms, learning types and limitations.

Each was to be described and a diagram included, but some candidates chose to omit diagrams or to produce generic ones that did not show the details of the specific network type. For example, the 'winner takes all' nature of Kohonen networks is a defining feature of these networks. The question gave candidates the opportunity to show breadth of knowledge of each of the five types, and obtain marks for depth, for example by providing good accounts of the learning outcomes. Many answers gave sufficient detail for good marks but often lacked the detail that is needed for full marks. This was particularly the case in describing the algorithms involved.

Question 6

This question was based on required reading for the course which this year, via coursework, focused on understanding images and the topical issues of self-driving cars. The question covered progress on using neural networks in the automation of driving motor cars so both self-driving and driving aids are valid topics.

Part a) asked for an outline of progress made. Here an historical view was a good way to start. Good answers contained an introduction to the problem, the scale of issues in driving, for example the volume of cars, road capacities and cost of accidents in lives and material. However, the focus must be on artificial neural networks, early attempts at their use and how these have developed into current systems. Excellent answers concluded with some speculation on where the technology is likely to go next.

As part of an examination of artificial neural networks it should have been no surprise that the types of networks used was the focus of part b).

Parts c) and d) asked for the link between network types and particular applications. Many answers did not go into sufficient detail; candidates who have little recall of the applications and the networks used are advised to avoid this type of question. It is important that candidates think of the rationale behind neural network application design during their studies, and specifically how architecture and application relate. It is also important to see an understanding of how the networks contribute to the solution of specific problems, and how this contributes to the overall solution.

The final parts of the question, parts e) and f), asked candidates to write about the broader issues involved in applying artificial neural networks to the problems associated with self-driving cars. The question asked for focus to be on non-technical issues which have to be addressed. These included issues of a social, legal or ethical nature. Comments on how each issue might be resolved were also required. Some candidates were able to give good answers here, but other answers were very limited.

Examiners' commentaries 2015–16

CO3311 Neural networks – Zone B

General remarks

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

Question one on Perceptrons, question three on Kohonen-Grossberg networks and question five covering backpropagation were the most popular, with most candidates attempting them. Question two, comparing Kohonen networks and Perceptrons, proved the least popular question on the paper. Perhaps this is an indication that candidates are not reading broadly enough and so are reluctant to answer general questions that include applications, despite their link with the coursework.

Comments on specific questions

Question 1

This question stated that a Perceptron unit represents a line in the plane and asked candidates about this correspondence. On the whole this question was not very well answered. Good answers included a diagram showing this correspondence and how the unit splits the plane into two. It was surprising that some candidates were not able to show how the parameters of the line, m and c, were related to bias and the weights on the inputs. Candidates were asked to give the input weights and bias of this Perceptron and explain how these were obtained. Many answers were not completely clear on this latter point. A diagram was to be included and all that was required was one showing a general straight line, how this divides the plane into two and the correspondence between the parameters of the line and the parameters of a unit defined by this line.

The subject guide explains how Perceptrons can be put together to perform tasks that are represented by truth tables. In part b), few candidates were able to apply this to the simple case given of an even number of inputs being one. Many solutions are possible and chapter four of the guide gives a similar example.

In part c) candidates were to explain how this unit could be extended to include more inputs. Clearly the inclusion of five or more inputs means a greatly increased network but the principles are the same. Few were able to express this very well.

Part d), the final part of question one, required a simple calculation of the training of a single threshold unit with two inputs. Although for many this was a straightforward question, some thought that backpropagation was to be used – a fundamental misconception or just possibly a simple mistake. Initial weights were chosen to make the calculation less onerous. One issue that might have thrown some candidates was the lack of specified learning rate. This was intentionally omitted so that candidates could have the opportunity to note the fact and choose a suitable one. No penalty was applied if the rate was not well-chosen, but candidates were expected to state the value they used.

Question 2

This question concerned Kohonen and Perceptron networks, and asked candidates to compare and contrast these, giving five headings under which this might usefully be done. Questions which require comparing and contrasting require more than just the difference between things. They need an exploration of the similarities and the differences between them. The five aspects were given as an aide memoire for candidates to help ensure that key points were made.

The first category, 'Types of unit normally used', showed examiners that some candidates do not know that while Kohonen networks use winner takes all, both use threshold units. Good answers gave a diagram showing the threshold activation.

Training algorithms was the next criterion. Good answers noted that though Perceptrons have a guaranteed learning algorithm when used singly, there are also algorithms to produce Perceptron networks given truth tables when binary inputs are used. The Kohonen network does not have such guaranteed algorithms but then the required outputs are not known. Good answers gave full accounts of the algorithms and explained how the errors reduce. Less good answers left out steps such as normalisation. Good answers also noted the limitations of the algorithms. Kohonen networks always converge except perhaps in some extreme cases, Perceptrons only converge for the particular class of problems called linearly separable ones; good answers explained the significance here.

The two network types also differ in the input and output that are achievable. Some candidates failed to mention that there were no extra restrictions on the inputs. Most candidates were able to describe the fact that while both network types produce binary output, in a Perceptron network more than one output can be at one, in contrast with the Kohonen network's winner takes all output. The need for the Grossberg layer was mentioned in the best answers.

In terms of 'Ease of use' both network types are quite straightforward for classification problems. However the main issue, as most candidates noted, is whether the required outputs are known. Perceptrons are only applicable to supervised learning, while Kohonen networks are only applicable to unsupervised learning.

A range of applications were generally outlined.

This was not a popular question and was not on the whole well answered.

Question 3

This question started with questions on general machine learning and ended in a simple calculation using a Kohonen network. The final part concerned initial choice of units. This popular question had very many good answers but also some very poor ones!

The course covers a number of machine learning systems using networks of units. It is important that candidates understand the variety of alternative strategies available for such learning. Although much of the learning covered was online, batching is also possible. For part a), good answers discussed weight updates and how these were timed in relation to presentation of examples. Most candidates were able to differentiate clearly between these. Similarly, few had difficulty explaining how supervised and unsupervised learning are defined and provided examples for both. The almost trivial distinction between bias and threshold also proved easy for most candidates.

For part b), candidates were asked to illustrate what might go wrong if we omit the normalisation in a Kohonen network example. Few managed to produce a convincing example here. Normalisation makes the calculation less computationally costly because it replaces the distance with the inner product Very good answers explained that normalisation is not always appropriate.

Part c) required a simple calculation. The initial conditions (weights and units) were chosen to make this reasonably straightforward. Although most candidates were able to work efficiently with this, some convinced themselves that the vector (1, 1) is normalised – this is a basic mistake, possibly to be attributed to examination stress. One of the examples given was a unit and examiners hoped candidates would simply note that no learning happens when an example is also a current unit. Few noticed this!

The requirement for part d) was that candidates would show one of the units moving towards the example as it is trained. While many produced a diagram, the quality of these was generally a little disappointing.

Part e) asks about initial choice of number and values of units. Many remembered that no 'good' algorithm is available and that pruning and growing networks are both used. It would have been good to see some acknowledgment that a preliminary investigation of the data set is almost certainly a prerequisite in most cases and this might suggest some ballpark figures.

Question 4

This question was based on an optimisation problem often used in computing: placing eight queens on a chess board in such a way that they cannot take each other. This is a frequently used example in studies of Hopfield networks and is used here to test candidates' understanding of how Hopfield nets can be designed. Sadly, many candidates were not able to explain clearly how the energy function and thus the weights of such a network are determined.

Part b) was much better answered as it required a clear statement of the process of producing a state transition table of a Hopfield network given its weights. It is worth noting, however, that evidence from the answers shows that candidates are more practiced in doing the calculations than they are in explaining why they are done.

Part c) was the best answered part of this question as most attempts produced good transition tables and diagrams. One common error included the ordering of units for the states. Use of three, two, one rather than one, two, three is recommended in the subject guide as it leads to an easier representation of what is going on. Another common error is in the depiction of self-transition. Sink states do not need any special indication but others need self-transition shown explicitly. The example given had two stable states, but few commented on this.

Although not among the most popular questions, more than half of the candidates attempted this question. Some high marks were achieved but so were some mediocre ones.

Question 5

This question covered backpropagation networks. Overall this was a popular question with many achieving very good marks. Part a) required candidates to show their understanding of the overall components of such networks, using words and a diagram. Some candidates chose to use representations other than that recommended in the guide. Although this is acceptable, it tended to lead to loss of marks as candidates using other representations often forgot to include components such as bias, net and activation functions that are explicit in the recommended representation.

For part b), the algorithm for the feedforward stage is straightforward but must include the bias in the calculation of **net**. It is in this stage that the concept of layer is important.

Part c) required the backpropagation stage to be described with corresponding equations. This part was also well answered in general. Again, candidates using other notations than those recommended in the guide often made mistakes using their chosen notation, resulting in loss of marks.

Part d) involved a two-input three-unit backpropagation network (with sigmoidal units). Initial weights were chosen to make calculations easy. Most candidates had little difficulty with this part.

Part e) required candidates to explain why a target of one cannot be achieved by such a network. While most candidates had no difficulty with this, a few were not able to give satisfactory answers to this basic question.

Question 6

This question explored aspects of the reading that candidates are expected to do while studying CO3311. On the whole, the question was not very well answered, with insufficient detail given. Less than 50 per cent of candidates attempted this question, despite the preparation that the coursework should have given them.

Part a) asked for an outline of the progress made in the use of artificial neural networks (ANN) in the automation of image interpretation. Good answers gave an historic perspective introducing the problem and how solutions have improved with time. The focus was of course on ANN.

The types of neural networks that have been used in this application were the topic of part b). An easy two marks for a couple of network types that are used for image interpretation were available, and most candidates were able to gain these.

For part c), the specific problems to which the network types were applied were asked for. Here we are looking for more depth than in part b), that is to what problems in image interpretation does each network type apply. Many candidates struggled to answer this in any real depth.

Knowing that particular network approaches have been used in particular solutions is only part of understanding developments in this field. It is also important to have an appreciation of the motivations and indications that produce these solutions. Candidates answering part d) were not as informed as they should have been about these aspects. Authors of articles usually spend a great deal of effort in explaining what motivates their work and candidates should pay attention to motivation when reading around their subject.

Parts e) and f) concerned non-technical issues involved with the development and use of image interpretation technologies. Candidates should be aware of the social, legal and ethical issues that are inevitably associated with any large technological breakthrough. They were to be listed in part e) and possible solutions explored in part f). On the whole, candidates preferred to write about technical issues and even these were not given the depth that one would hope for in a stage three answer.