Module Code
COM00029H

BEng, BSc, MEng and MMath Degree Examinations 2018–9

DEPARTMENT OF COMPUTER SCIENCE

Introduction to Neural Networks (INNS)

Open Individual Assessment

Issued: 27th February 2019, 12:00 noon
Submission due: 17th April 2019, 12:00 noon
Feedback and marks due: 15th May 2019, 12:00 noon

All students should submit their answers through the electronic submission system: http://www.cs.york.ac.uk/student/assessment/submit/ by 17th April 2019, 12:00 noon. An assessment that has been submitted after this deadline will be marked initially as if it had been handed in on time, but the Board of Examiners will normally apply a lateness penalty.

Your attention is drawn to the section about Academic Misconduct in your Departmental Handbook: https://www.cs.york.ac.uk/student/handbook/.

Any queries on this assessment should be addressed by email to Simon O'Keefe at simon.okeefe@york.ac.uk. Answers that apply to all students will be posted on the VLE.

Rubric:

Answers must not exceed 8 A4 pages in total, with a minimum 11pt font and minimum 2cm margins either side. This limit includes any title page, diagrams, references, and so on. Excess pages will not be marked.

Your answer must address all of the points to gain full marks. It is not necessary to submit code for the neural network, but you must evaluate thoroughly the performance of the network (in terms of accuracy). You must provide justification for the choices you make. If your network does not perform perfectly you will still get credit for attempting to explain the performance. Credit will be given for reference to the appropriate literature. Your answer should show that you have understood the literature you have selected, are able to explain the material, and can critically compare material from different sources.

Your exam number should be on the front cover of your assessment. You should not be otherwise identified anywhere on your submission.

This assessment requires you to create a neural network. You may use any appropriate software to create the neural network, including but not limited to MATLAB and/or its toolboxes. A full justification for your choice of network and evaluation of its performance are more important than the ability of the network to produce "correct" results.

The data for this assessment are taken from the UCI Machine Learning Repository, and may be found on the Assessment page of the VLE site.

The problem is to determine whether a patient is healthy or has breast cancer. Your task is to construct a neural network to map from the data that is given to a binary decision.

The predictors are anthropometric data and parameters which were gathered in routine blood analysis.

For this assessment, you will use the data in the file `dataR2.csv`, a CSV file that contains the raw values of the 10 measurements and the corresponding class label.

[This data was first used in:

Patricio, M., Pereira, J., Crisostomo, J., Matafome, P., Gomes, M., Seica, R., & Caramelo, F. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer, 18(1).

DOI: `10.1186/s12885-017-3877-1`]

Your report must be **no more than 8 A4 pages in length** and must contain the following sections:

1. [20 marks] Discussion of architectures.

   This should

   - describe (briefly) the data you have, and how much there is of it.
   - identify the type of problem
   - identify which classes of architectures would be suitable
   - give a brief discussion of the technical features of the architectures, and the advantages and disadvantages of each
   - state which class of architecture you are going to use and justify your choice, relating the characteristics of the problem to the advantages/disadvantages of the architecture.

   To do this you might need to

   - do some preliminary experiments with simple versions of the architecture to get a feel for what will work
   - do some exploratory data analysis to see what the characteristics of the data are
   - consider the principles involved and relate them to the problem.

2. [30 marks] Creation and application of a neural network.

   Describe the inputs to (and outputs from) the network. You need to describe how the data you started with have been preprocessed. Explain how you investigated the data, including any assumptions you have made. Again, this may include some testing of networks to see what the effects are of different preprocessing choices. Give sufficient detail for someone else to process a new batch of data for use with the trained network.

   State which training algorithm you selected, and explain how you selected the best training algorithm for this problem. For the selected training algorithm, give sufficient detail for someone to implement the training algorithm. This does NOT mean (for example) describe gradient descent in great detail. It DOES mean give any parameters, initialisation, etc, even if they are the toolbox defaults.

   Explain the process you went through in making the selection of the final architecture, for example, the number of neurons or the number of layers to use. If you evaluated a number of networks, give details of what their structures were and how they performed. You may summarise repeated tests of the same structure.

3. [20 marks] Results and evaluation

   Explain the metric or metrics you have used for comparison between networks. – a MSE on its own is not helpful in judging how well something works.

   Give a synopsis of the results obtained from the *final* selected network. Identify anything of interest in the results, such as areas of particularly good or poor performance, or variation between different training runs.

   Relate these results from the network back to the problem given, and reflect on the conclusions that you may draw.

4. [30 marks] Further application

   In the previous section you used a neural network to convert blood test results into a diagnosis.

   Another tool for detection and diagnosis of breast cancer is the mammogram, an X-ray image of the breast. Mammograms are evaluated by human readers, a process which is monotonous, tiring, lengthy, costly and most importantly, prone to errors.

   Assume the availability of a large number of mammograms, both normal and with lesions, and labelled to indicate different types of lesions. Discuss the issues you would need to consider in selecting, constructing, training and evaluating a neural network to process these images and to indicate which images needed further investigation.

End of examination paper