University of Glasgow

Saturday 14th August 2021
(24 hour open online assessment – Indicative duration 2 hours)

DEGREE OF MSc

Deep Learning for MSc (M)
COMPSCI 5103

(Answer ALL 3 questions)

This examination paper is worth a total of 60 marks.

1. Refer to the neural network shown in Figure 1 to answer the following questions. In this diagram, all input values $x_n^{[0]} = 1.0$ and all bias, b, and weight, w, parameters take on the value 0.1.

(a) Give plausible reasons why the network designer has chosen ReLU activation for some neurons in this network while using sigmoid for others. In particular, consider the case if this network was extended to be a much deeper network. [4]

(b) This network contains 9 parameters (6 weights and 3 bias). How many passes of the network would be required to determine the gradients for all these parameters using forward autodiff and reverse autodiff. [2]

(c) Carry out a forward pass on this network to determine the output value $y_1^{[2]}$. Show intermediate values for hidden units. [2]

(d) We define:

$$z_1^{[2]} = w_{1,1}^{[2]} h_1^{[1]} + w_{2,1}^{[2]} h_2^{[1]} + b_1^{[2]}$$

Carry out the start of a backward pass to determine the numerical value of:

$$\frac{\partial y_1^{[2]}}{\partial h_1^{[1]}}$$

(Show your working, employing the $z_1^{[2]}$ term, and also state the final numerical value obtained.) [5]

(e) Continue this backward pass to determine the numerical value of

$$\frac{\partial y_1^{[2]}}{\partial w_{1,1}^{[1]}}$$

showing how this is generated from the previous term you calculated:

$$\frac{\partial y_1^{[2]}}{\partial h_1^{[1]}}$$

(Again show your working employing the intermediate value $z_1^{[1]}$ and the final numerical value obtained.) [4]

(f) Assume these are the weights and biases set at initialization of a gradient descent optimization. What key problem exists with these initial settings in this case? How could it be resolved? [3]

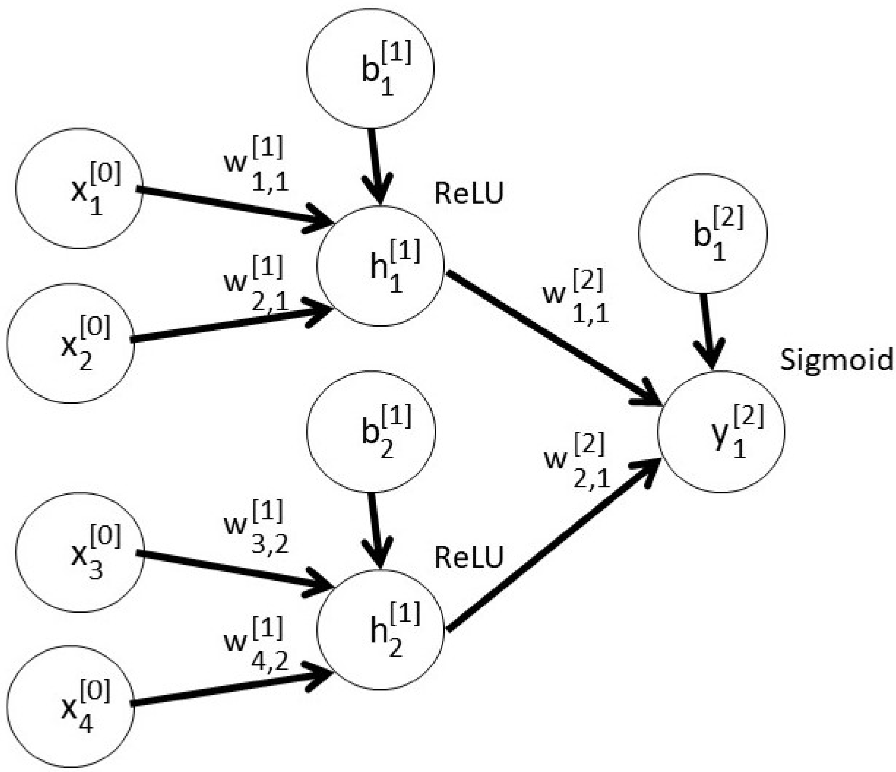

Figure 1: Example feedforward neural network with parameters.

2. You are developing a deep neural network to detect a rare and deadly type of cancer using microscope images of biopsy tissue. Any positive cases detected would undergo further manual checking to confirm them. You have 10000 microscope images of normal tissue (negative samples) and 100 images of the cancer (positive samples). You randomly sample 2000 normal images and 20 cancer images from these data to produce a test set and you intend to employ 4-fold cross validation using the remaining data to optimize hyperparameters using random search.

(a) You are considering three different measures of success: i) accuracy, ii) precision & recall and iii) F1 score. Explain whether each measure is appropriate given this scenario and justify your answers. [5]

(b) Explain why you cannot use one of the measures of success given in part (a) as the loss function for the network even though this is what you want to optimize. What loss function should you use in this scenario and what are you hoping by optimising it? [3]

(c) You adjust some common hyperparameters during hyperparameter optimization. One of them generates the learning curves given in Figure 2 as it is adjusted. What is the hyperparameter you are most likely adjusting here and explain whether it is taking a large/medium/small value for the given learning curves A), B) and C). Which of these curves shows the best hyperparameter value and explain why this particular value leads to the most optimal learning. [3]

(d) Similar to part (c), what hyperparameter are you most likely adjusting in Figure 3 and explain whether it is taking a large/medium/small value in the cases of learning curves A), B) and C). Which of these curves shows the best hyperparameter value and explain why this particular value leads to the most optimal learning. [3]

(e) How should you generate your final model from cross-validation and determine its overall final score? Be explicit about the number and size of training and validation sets used, how optimal hyperparameters would be determined, and how the final model would be generated and scored using the data. [6]

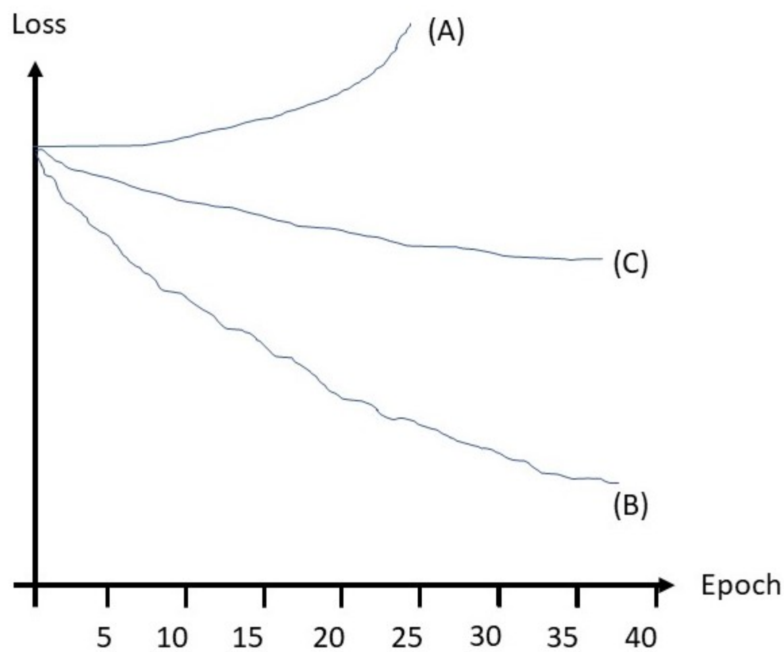

Figure 2: Example learning curves when changing a particular hyperparameter.

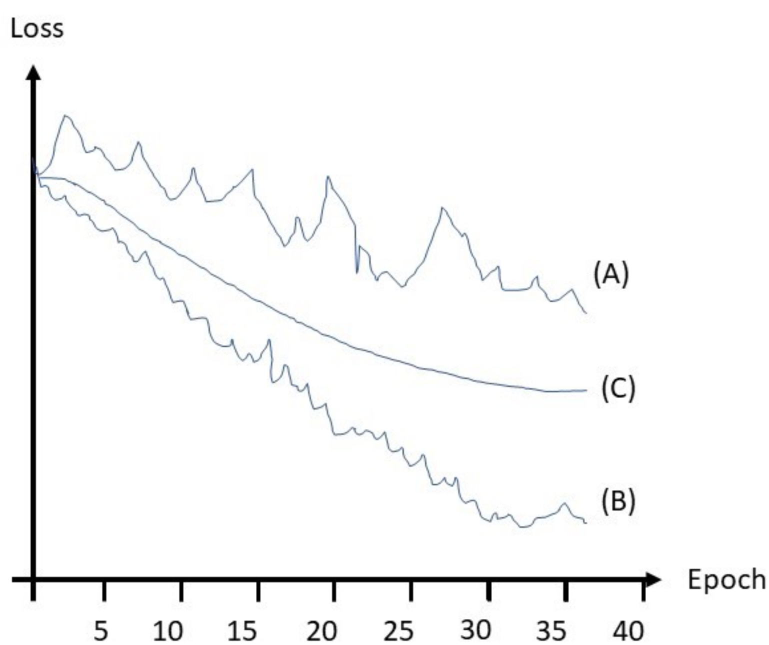

Figure 3: Example learning curves when changing a different hyperparameter.

CONTINUED OVERLEAF

3. A hospital has a large collections of images which could form a training set for a particular diagnostic classification task. However, they have only been able to afford to label 10% of the datapoints. The diagnostic system is intended to analyse thousands of people every day and speed up diagnosis by making decisions on cases it is certain about, and referring cases it is uncertain about to a human specialist. The baseline rate for a positive diagnosis which requires medical intervention is 1% of tests.

(a) Describe a way that the hospital could use the unlabelled data to improve the classification performance. [3]

(b) The manual labelling of the image classes in the training set is subject to uncertainty. How could this be taken into account in the training process, and what advantages would it bring? [4]

(c) Give three examples of visualisation methods that could help a doctor interpret the decision making process of a deep network on a specific image. Explain the limitations of this approach to explaining classification behaviour. [5]

(d) A university is considering using a deep machine learning system to rank applicants for admission based on a set of features which describe the candidate. As a training set they plan to use historical performance data from students accepted in the past, for whom they have the same feature data. It will give students instant feedback about their probability of being accepted, and the success prediction will be shared with staff to help them allocate support for students. Once the system is running, accepted students and their performance will be automatically added to the training data for future classifications.

What are the ethical and practical risks associated with the use of machine learning in such an admission system? [8]