Queen Mary
University of London

Main Examination period 2023 – May/June – Semester B

MTH767P: Deep Learning and Neural Networks

Duration: 2 hours

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

The exam is intended to be completed within **2 hours**. However, you will have a period of **4 hours** to complete the exam and submit your solutions.

You should attempt ALL questions. Marks available are shown next to the questions.

You are allowed to bring **three A4 sheets of paper** as notes for the exam.

Calculators are not permitted in this examination. The unauthorised use of a calculator constitutes an examination offence.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any unauthorised notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: 1^{st} **N. Otter,** 2^{nd} **K. Papafitsoros**

Question 1 [38 marks]. Consider a convolutional layer with an input feature map of size $500 \times 500 \times 2$ (i.e., with two channels). **Note:** you should motivate your answers, e.g., it is not enough to use formulas that we proved in class. If you do use them, you should briefly explain how one derives them.

(a) What is the size of the output feature map if we consider kernels of size 5×5 for each of the channels, as well as no padding and stride $s = 2$? [**7**]

(b) Suppose that the size of the output feature map is $500 \times 500 \times 3$ (i.e., with 3 channels). What are k, s, p? Here k denotes the size of the kernel along both axes, s is the stride, and p is the padding. [**8**]

(c) Suppose that we want to increase the size of the output feature map with respect to the input, namely, we want the size of the output feature map to be $o \times o \times 2$, with $o > 500$. Give a possible choice of padding, stride and kernel size that would return such an output. [**8**]

(d) Why does maxpooling often perform better than average pooling? Give a brief explanation, and illustrate your argument with an example. [**8**]

(e) Suppose we have an average pooling layer of size 5×5, where the input feature map has size 10×10. Define a convolutional layer that gives exactly the same output as the average pooling layer. Note: you need to provide the kernel, as well as the values for stride and padding.

[**7**]

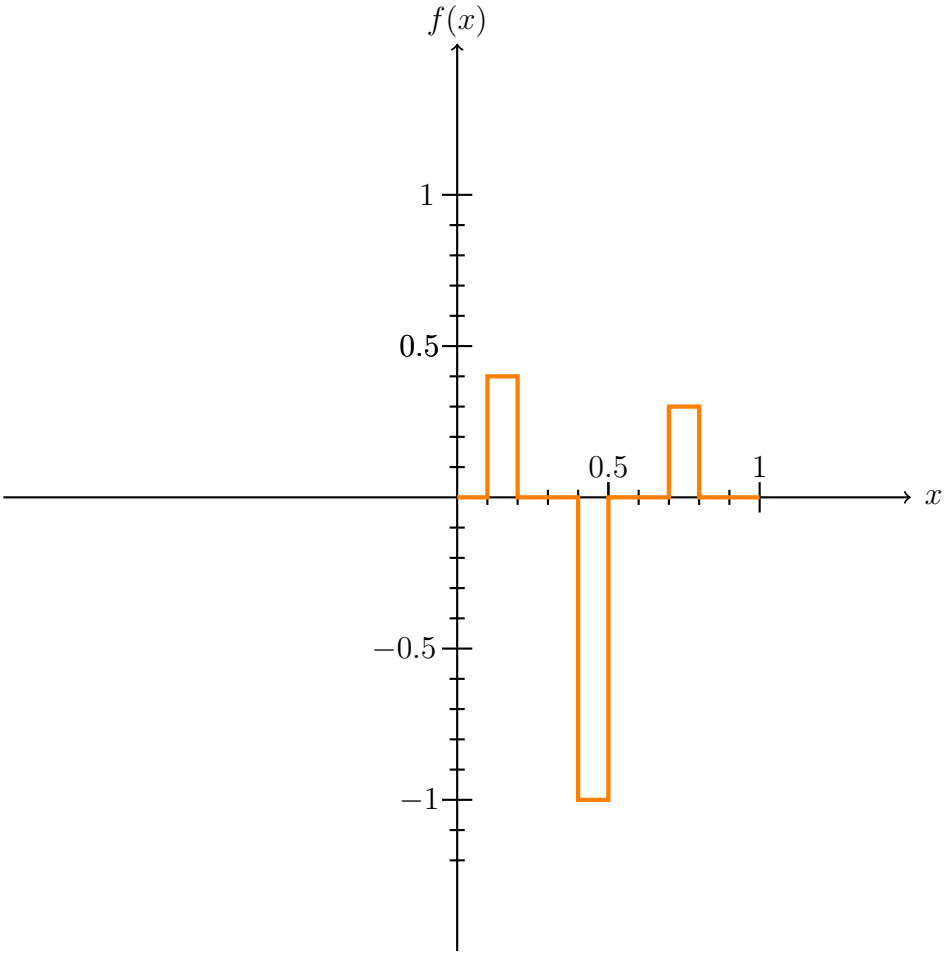
Question 2 [10 marks].
Give an example of a feedforward network with at least one hidden layer that represents the OR function, and specify:

(a) Weights, biases and activation functions for all units (provide a brief motivation for your answers). [**6**]

(b) Input and output. [**2**]

(c) Decision boundary (a drawing is enough). [**2**]

Question 3 [22 marks]. Consider the graph of the function $f\colon [0,1] \to [-1,1]$:



The function f is defined as follows:

$$f(x) = \begin{cases} 0, & \text{if } 0 \leq x < 0.1 \\ 0.4, & \text{if } 0.1 \leq x \leq 0.2 \\ 0, & \text{if } 0.2 < x < 0.4 \\ -1, & \text{if } 0.4 \leq x \leq 0.5 \\ 0, & \text{if } 0.5 < x < 0.7 \\ 0.3, & \text{if } 0.7 \leq x \leq 0.8 \\ 0, & \text{if } 0.8 < x \leq 1 \end{cases}$$

Define a feedforward neural network that represents this function (up to a small error). You should specify:

(a) Weights, biases and activation functions for each unit. [**14**]

(b) Input and output. [**2**]

(c) A brief explanation of why the neural network given can approximate the given function. [**6**]

Question 4 [12 marks]. Explain briefly how dropout is defined, illustrating its definition with an example, and clearly discussing what happens during training and test phases. [**12**]

Question 5 [12 marks].

(a) You are given the words "cat", "snake", "cats", "snakes". Provide a word embedding of these words into the unit square $[0, 1] \times [0, 1]$ and motivate your choice. [**6**]

(b) You are given the sentence "The king sat on the sofa." Write a one-hot encoding of this sentence, assuming that you have a dictionary of size 8. [**6**]

Question 6 [6 marks].
State the No Free Lunch Theorem. Then discuss briefly one implication it has for Deep Learning. [**6**]

End of Paper.