THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

DASC7606 Deep Learning

Date: May 18th 2022 **Time: 6:30pm - 8:30pm**

There are three questions in total, the mark value of the question (or part of the question) is indicated before the question (or part of the question). All questions are compulsory.

Candidates are permitted to refer to any printed/handwritten materials in the examination. Internet searching and crowdsourcing from group messages, online forums or social media, etc. are strictly forbidden.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Candidates should provide handwritten answers. Candidates should submit a single PDF file to OLEX.

Question 1 (General questions) (30pts). Provide short answers (max 4-5 lines) for each of the following questions:

a) Explain the concept of L2-regularization and how it can alleviate overfitting.
b) Can the perceptron compute the XOR binary operator? Explain your answer.
c) Can a neural network approximate any function arbitrarily well even when ReLu activation functions are used? Explain your answer.
d) Mention at least three disadvantages of using sigmoid activation functions.
e) What is the main advantage of Adam in comparison with RMSProp?
f) What is the main reason why a "deep" neural network might have worse accuracy than a "shallow" neural network? Which architecture, among the ones we studied in our course, aimed at alleviating such a problem?
g) How do LSTMs alleviate the problem of vanishing gradient?
h) Mention at least one disadvantage of transformers.

Question 2 (NNs) (20pts). Consider a simple neural network with input x_1,x_2,x_3, weights w_1,w_2,w_3, bias b and the sign function as the activation function, where sign(x)=1 if x>0 and 0 otherwise. In particular, the neural network outputs sign ($b+\sum_i w_i x_i$). Determine whether the functions specified in the table below can be computed by the previous neural network. In the positive case, provide the values for the weights w_1,w_2,w_3 and the bias b. For each function that cannot be computed by the aforementioned neural network, provide a proof that this is indeed the case.

x_1	x_2	x_3	f	g	h	l
0	0	0	0	0	0	0
0	0	1	0	1	0	0
0	1	0	1	1	0	0
0	1	1	1	0	1	1
1	0	0	0	1	0	0
1	0	1	0	0	1	1
1	1	0	1	0	0	0
1	1	1	1	1	1	0

Question 3 (CNN) (50pts). Devise a CNN classifier which determines whether a 4 × 4 binary image contains the digit "7" or not. In particular, your CNN should output 1 for the image below and 0 otherwise.

1 1 1 1
0 0 1 0
0 1 0 0
1 0 0 0

Your CNN should consist of a) at most three layers, each of them being a pooling or convolutional layer, with 2x2 filters, stride 1, no padding. For simplicity we use a threshold function f(**x**,**w**,t) as an activation function for the last layer, which outputs 1 if the dot product **xw** >= t, 0 otherwise (**w** can be negative or non-negative). All other activation functions are the identity function.

a. (10pts) Specify the architecture of your CNN as well as the value of its parameters, in the case when the input images contain *exactly* seven "1"s. In particular, for each layer you should specify whether it is a convolutional or pooling layer, the number of filters, the values of the filters, the biases, the value of the vector **w**, the threshold t, as well as the value of any other parameter. We shall refer to such a CNN, as CNN-1.
b. (5pts) Report the value of the tensors for each layer of CNN-1, in the case when the input image is the one shown above.
c. (10pts) Argue that CNN-1 always produces the correct result.
d. (10pts) Is it possible to devise a CNN architecture, with the requirements specified above, so as to be able to classify 4x4 input images with at most seven "1"'s ? In the positive case, devise such a CNN, otherwise argue that it is not possible.
e. (10pts) Is it possible to devise a CNN architecture, with the requirements specified above, so as to be able to classify any 4x4 input image? In the positive case, devise such a CNN, otherwise argue that it is not possible.
f. (5pts) Compute the size of the tensor in the first layer in CNN-1 in the case when padding = 2 and stride = 2.

=== END OF PAPER ===