UNIVERSITY OF LINCOLN

MODULE ASSESSMENT:	**Mock In-Class Test**
COLLEGE:	**College of Science**
SCHOOL:	**School of Computer Science**
MODULE:	**Advanced Machine Learning**
MODULE CODE:	**CMP9137M**
LEVEL:	**MSc**
DELIVERY TEAM:	**Dr. Heriberto Cuayáhuitl** **Dr. Lei Zhang** **Dr. Junfeng Gao**
RELEASE DATE:	**10 May 2022**
START TIME:	**Your own time**
TIME ALLOWED:	**As needed (3 hours for the real test)**

==========

INSTRUCTIONS TO STUDENTS:

QUESTIONS TO ANSWER:	***All***
MARKING SCHEME:	***As indicated in each question***
MATERIALS PROVIDED:	***Lectures, workshops, and references***
MATERIALS PERMITTED:	***Any during the mock test. Only pen, paper and a non-programmable calculator during the real test***
NOTES TO STUDENTS:	***This mock test will not be marked***

Mock In-Class Test (2021-22)

Instructions: Answer all questions in the spaces provided on the question box – in the order of your preference. If you run out of room for your answers, please use more pages with your name and student number clearly indicated.

Student Name and Number: _____

1. (7 points) Explain why is Machine Learning needed and provide a definition for each of the following Machine Learning paradigms (mentioning an example application per paradigm): Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

2. (6 points) Explain (conceptually) the Dueling DDQN reinforcement learning method.

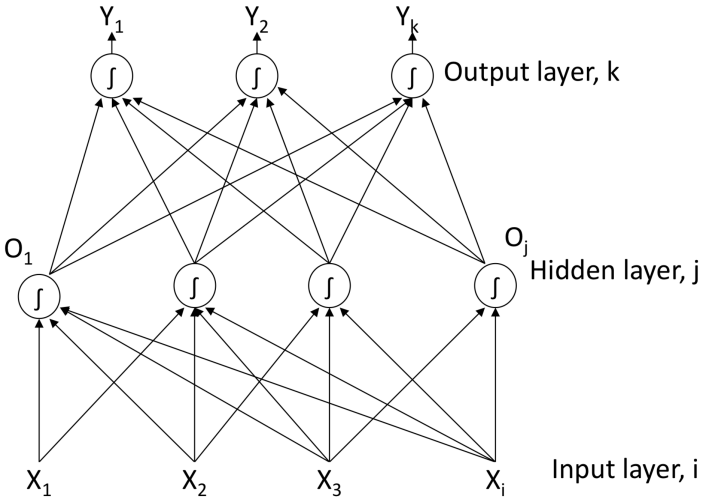
3. (6 points) In Machine Learning, there are two important Errors to dealt with: In-sample Error E_{in} and Out-of- Sample Error E_{out}. Explain the difference between E_{in} and E_{out} and their relationship with 'overfitting' and 'underfitting'.

4. (6 points) A student is working on a machine-learning approach to spam detection. The task is to detect if an email is a spam. He/she has 100 labelled emails, 90% of which are used for training and 10% for validating the model. The student applied 10 different learning algorithms, training each one on the training examples, and recording the accuracy on the validation set. The best performance among the 10 algorithms achieves 90% accuracy. This is assuring that the best algorithm is good enough to be deployed. State whether the above statement is "Correct" or "Wrong" and explain Why.

5. (6 points) Explain the three types of layers in the neural architecture below and write the corresponding equations required for a neural net with 2 hidden layers.

$$Y_k = f(\sum_j w_{jk} \cdot O_j)$$

$$O_j = f(\sum_i w_{ij} \cdot X_i)$$



.

6. (7 points) Explain the concept of Attention in Recurrent Neural Networks and describe how it is calculated.

7. (6 points) Calculate the Scaled-Dot Product Attention ($Softmax(\frac{QK^T}{\sqrt{(d_k)}})V$) of a Transformer network using the information below, and write down your worked out solution:

Consider $E = \{e0, ..., e_{n-1}\}$ as the input word embeddings of the sentence "neural machine translation" with 5 dimensions:

$$e_0 = [0\ 1\ 0\ 0\ 3], e_1 = [1\ 0\ 0\ 1\ 2], e_2 = [0\ 0\ 2\ 0\ 1]$$

Matrices Q, K and V can be calculated as $Q = EW^Q$, $K = EW^K$, $V = EW^V$. Use the following (hypothetical) weight matrices to calculate the Scaled Dot-Product Attention.

$$W^Q = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} W^K = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} W^V = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 0 & 2 \\ 2 & 0 & 1 \\ 0 & 3 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

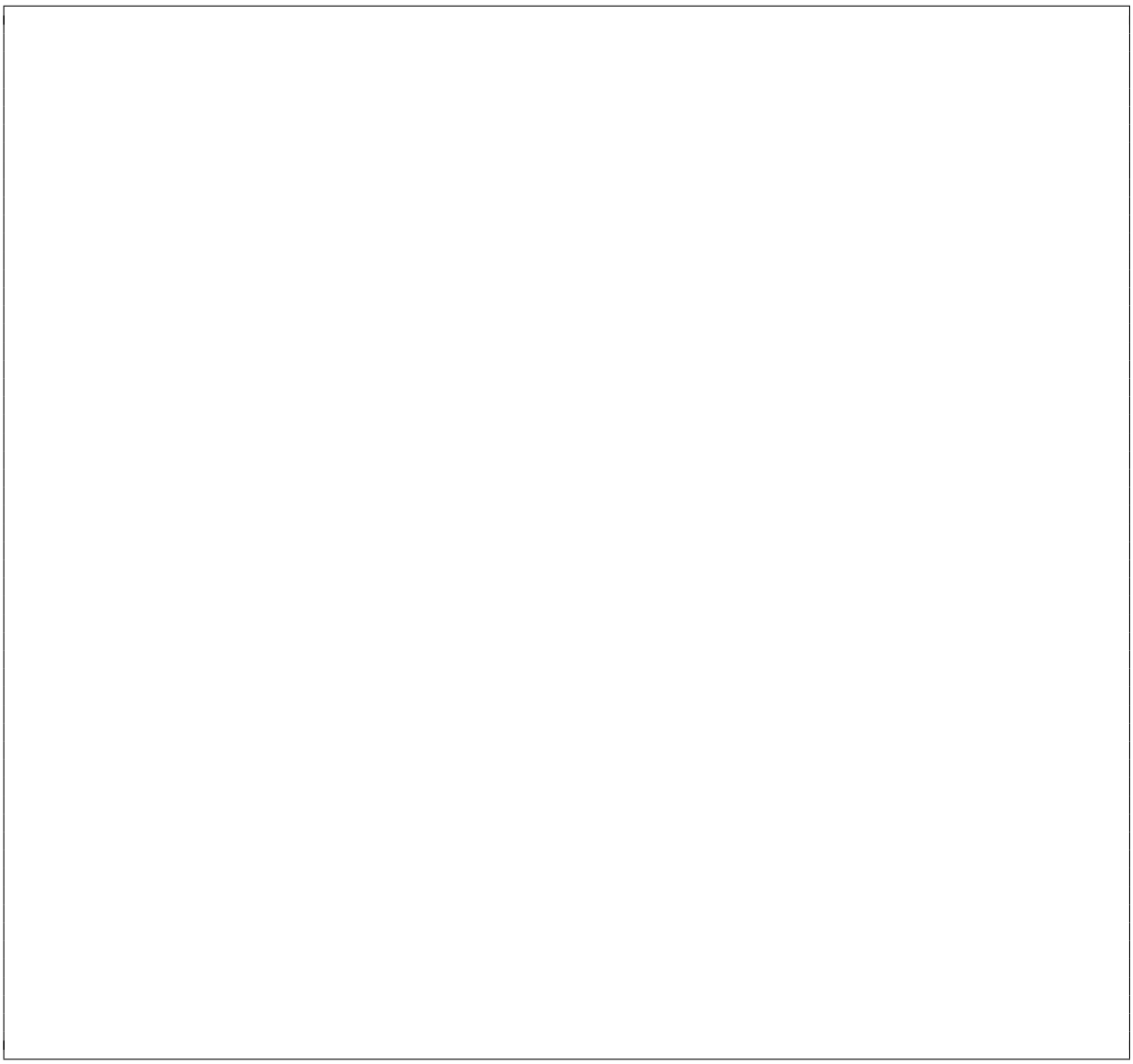
8. (7 points) Given function $f(w,x) = \frac{1}{1+e^{(-w_0x_0+w_1x_1+w_2)}}$ with w0=-3, x0=1, w1=2, x1=3, and w2=-2;

   - write this out as a computational graph and calculate the values that go through each operator/nodes in the forward pass, and
   - calculate all gradients in the backward pass with the backpropagation procedure (you are free to use any loss function).

   Use the following derivative formulas:

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -\frac{1}{x^2} \qquad f(x) = e^x \quad \rightarrow \frac{df}{dx} = e^x$$

$$f_b x = b + x \quad \rightarrow \quad \frac{df}{dx} = 1 \qquad f_a x = ax \quad \rightarrow \quad \frac{df}{dx} = a$$



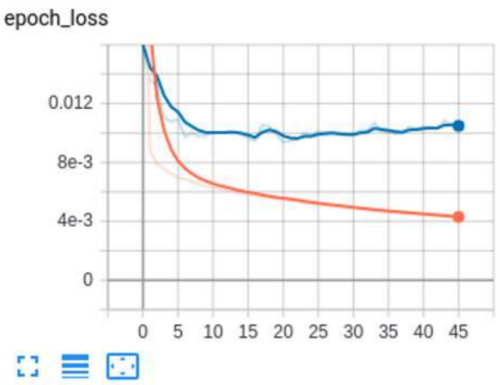
9. (11 points)
   - Provide definitions for the following terms related to convolutional neural networks (CNNs):
     - (i) convolutional layer,
     - (ii) pooling layer,
     - (iii) activation function, and
     - (iv) fully connected layer.
   - Draw a typical CNN classification architecture that contains all the layers above.
   - Compute the spatial size of the output for a 7x7 input that convolutes by a 3x3 filter with stride 2 and padding of 1.

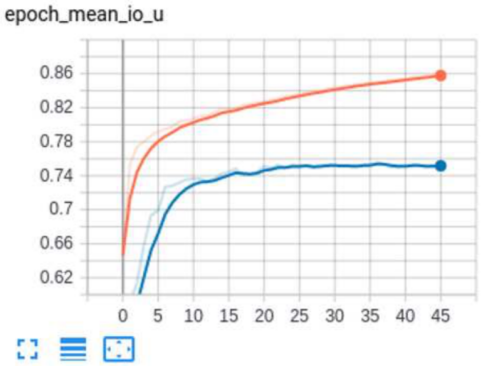
10. (7 points) There are two primary branches for deep CNN-based object detection, one stage and two-stage detector. The Faster-RCNN is a representative two-stage detector which follows a traditional object detection pipeline, and YOLO which is a classic one-stage detector.

    - What are the main differences in terms of learning between the one-stage and two-stage pipelines?
    - Describe the main workflow for both Faster-RCNN and YOLO.

- Consider applying a CNN-based detection network to detect a group of small objects (e.g. birds) in a frame. Which detector (Faster-RCNN or YOLO) would you choose and why?

11. (7 points) Suppose you are training a Fully-Connected Neural Network with a dice-loss function for semantic segmentation on some data (2000 training images, 200 validation images). Observe the training curves below for both loss and IoU (Intersection over Union), where the orange and blue curves represent the training and validation, respectively. Describe what is the issue in the figures, explain any reasons that could lead to this issue during the training and suggest any solution(s) to tackle this issue?





12. (7 points) Which is more important and why – model accuracy or model performance?

13. (7 points) Define the concept of hyperparameters in a machine learning algorithm and elaborate on how to tune the hyperparameters of a Neural Network model.

14. (5 points) Suppose you have an input volume of dimension $n_H \times n_W \times n_C$ for a Convolutional Neural Network. Which of the following statements you agree with? (Assume that the 1x1 convolutional layer below always uses a stride of 1 and no padding.)

    - You can use a 1x1 convolutional layer to reduce n_C but not n_H, n_W.
    - You can use a pooling layer to reduce n_H, n_W, and n_C.
    - You can use a 1x1 convolutional layer to reduce n_H, n_W, and n_C.
    - You can use a pooling layer to reduce n_H, n_W, but not n_C.

15. (5 points) In reality, it is difficult to obtain a large number of training examples and the imbalance samples usually exist in datasets. Explain a solution to deal with imbalance samples and small-size datasets for classification.

16. (6 points) Assume the following simplified Q-function of a reinforcement learning agent with triples (s=state, a=action, q-value): (1,1,-5.4), (1,2,-2.1), (1,3,3.5), (1,4,8.2). What is the value of $a^* = \arg\max_a P(a|s)$. Show your calculations according to the Boltzmann exploration strategy $P(a|s) = \frac{exp^{(Q(s,a)/\tau)}}{\sum_b exp^{(Q(s,b)/\tau}}$ and temperature parameter $\tau = 1.5$.

Best Wishes