Identikey: _____

Artificial Intelligence (F19)
Quiz 5 (QZ5)

First Name: _____

Overview
7.5 Point Quiz on Search (7.5% of final grade) **Last Name: _____**
30 minute closed book quiz

Learning Objective
This assignment satisfies learning objective 1 (LO1) as specified in the syllabus. You will demonstrate conceptual understanding of the core AI topics.

The quiz is worth 7.5 points total but there are 8 points available. This means that you can lose 0.5 points, but still earn a perfect score of 7.5 points on the quiz. If you score above 7.5 points, your total score will be rounded down to 7.5 points.

Questions

1. [Total Points: 2.75] Multiple choice/short answer questions.
    a. [Points: 0.25] Which of the following would you consider to be a deep neural network?

        Circle all that apply
        i. A feed-forward network with one hidden layer
        ii. A feed-forward network with four hidden layers
        iii. A convolutional neural network with 100 hidden layers
        iv. A convolutional neural network with one hidden layer
        Solutions: ii, iii, and iv.
        Convolutional neural networks are deep neural networks.

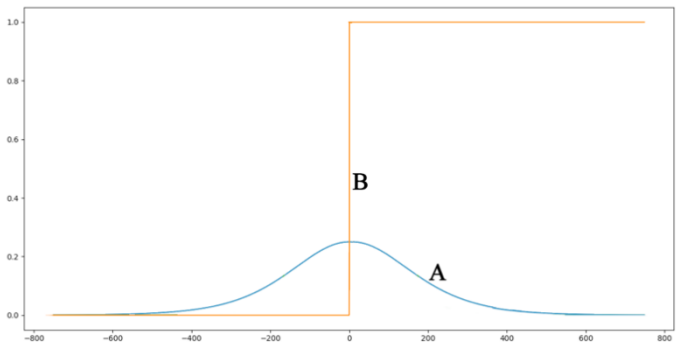
    b. [Points: 0.5] Name the neural network architecture (example: CNN) best suited for each task below:
        i. Learn a compact representation of the input
        Solution: autoencoder
        ii. Learn temporal dependencies in the data
        Solution: RNN/LSTM
        iii. Classify patterns
        Solution: fully connected feed forward network (or MLP)
        iv. Extract features from raw data
        Solution: CNN

Identikey: _____

c. [Points: 0.25] Identify the activation functions A, and B based on their first derivatives shown in the following image.

i. A- Tanh, B- Sigmoid
ii. A- Sigmoid, B- ReLU
iii. A- ReLU, B- Sigmoid
iv. A- ReLU, B- Tanh

Solution: ii)



d. [Points: 1.0] Identify whether you would **typically** (i.e., on average) expect an increase/decrease/no change in bias and variance in each of the following scenarios.

Bias (error): *difference between the model output and target output. How accurate is the model?*
Variance: *variability of a model prediction for a given data point. In other words, does the model overfit?*

i. Adding more units in each hidden layer. Bias: Variance:
Solution: Decreases, Increases
ii. Using Dropout while training. Bias: Variance:
Solution: No Change, Decreases
iii. Getting more training data. Bias: Variance:
Solution: Decreases, Decreases
iv. Using Nested Cross Validation. Bias: Variance:
Solution: No change, Decreases

e. [Points: 0.25] Which of the following network architectures would be most applicable to diagnose broken bones in an X-ray?

i. Fully connected - feed forward
ii. LSTM
iii. CNN
iv. Hopfield network
Solution: iii) Convolutional neural network

f. [Points: 0.25] Which of the following network architectures would you use to learn to translate sentences from English to French?

i. Fully connected - feed forward
ii. LSTM
iii. CNN
iv. Hopfield network
Solution: ii) LSTM

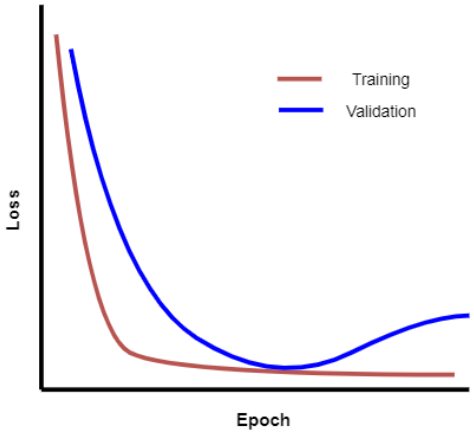
g. [Points: 0.25] Which of the following applies to a CNN?

i. Does not overfit to the data.
ii. Sharing weights between common feature detectors.
iii. Detecting time variant changes.
Solution: ii) Sharing weights between common feature detectors.

2. [Total Points: 1.0] Hyper-parameter Tuning

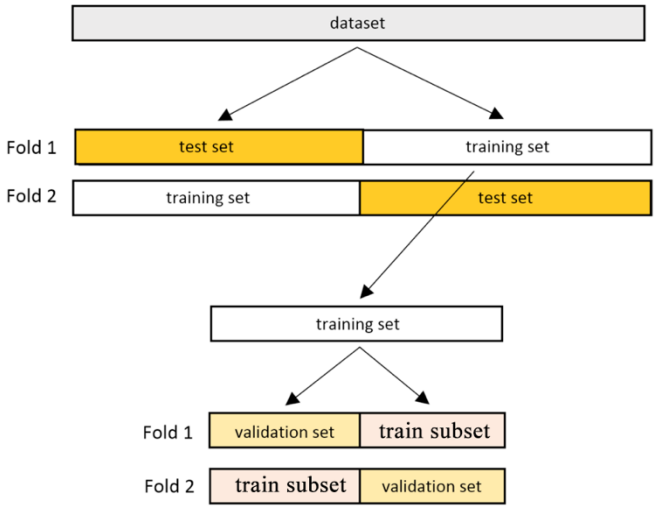
a. [Points: 0.5] Observe the plot of validation and training loss across epochs. Does the scenario illustrate overfit or underfit? *Justify your response*.

Solution: The plot shows overfitting. Since we can see that the training loss is still dropping over epochs but the validation loss has started to increase we can conclude the model has started to learn features in the training data that don't generalize to unseen data.



b. [Points: 0.5] Graphically illustrate the concept of nested cross-validation using two folds.

Solution:



Identikey: _____

3. [Total Points: 0.75] Gradient Descent and Normalization
    a. [Points: 0.25] Order the three types of Gradient Descent techniques based on the number of examples processed before the weight update (increasing order – smallest first).
        i. Stochastic, Mini-Batch, Batch
        ii. Mini-Batch, Batch, Stochastic
        iii. Stochastic, Batch, Mini-Batch
        Solution: i) Stochastic < Mini-Batch < Batch

    b. [Points: 0.5] Suppose stochastic gradient descent in a deep neural network is taking excessively long to converge. Which of the following techniques could help find better parameter values? **Circle all that apply.**
        i. Initialize the weights to 0.
        ii. Trying mini-batch gradient descent instead of stochastic gradient descent.
        iii. Tuning the learning rate.
        iv. Normalizing the inputs.
        v. Normalizing the hidden layer outputs.
        Solution: ii) , iv) v) but iii) is acceptable as well.

4. [Total Points: 1] CNNs
    a. [Points: 1.0] Consider the following image obtained after convolving a 4 x 4 image with a 2x2 filter using stride = 1. What is the result of applying the ReLU activation function followed by 2x2 max pooling feature with stride 1? Show all steps.

4	-1	2
6	-8	5
2	-4	2

Result after ReLU → → Result after Max Pooling on ReLU

Solution:

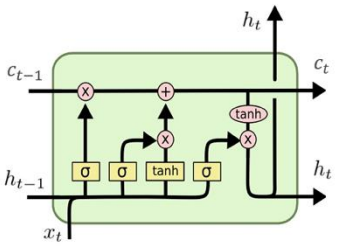
Result after ReLU → → Result after Max Pooling on ReLU

4	0	2
6	0	5
2	0	2

6	5
6	5

Identikey: _____

5. [Total Points: 1.5] Consider the following diagram & equations of an LSTM cell and answer the following questions:



$$i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1}\right)$$
$$f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1}\right)$$
$$o_t = \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1}\right)$$
$$\tilde{c}_t = \tanh\left(W^{(c)}x_t + U^{(c)}h_{t-1}\right)$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t$$
$$h_t = o_t \circ \tanh\left(c_t\right)$$

   a. [Points: 0.5] During the cell update step, the input gate returns 0.5, and forget gate returns 0.75. The old cell state is 0.3 and the current candidate input is 1. What is the new cell value? (No need to compute the result, just plug in the values in the appropriate equation)

      Solution: $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t = 0.75 * 0.3 + 0.5 * 1 = 0.725$

   b. [Points: 0.5] Suppose you want the memory cell state to only consider new inputs. What values should the input gate and forget gate take?

      Input Gate: _____ Forget Gate: _____

      Solution: Input Gate = 1, Forget Gate = 0

   c. [Points: 0.5] The hidden state information h_t is sent in two directions (see top right and bottom right arrows). Where does each go to?

      Solution: To the next time step and to the output.

6. [Total Points: 1.0] Designing Deep Learning Architectures

   a. You are tasked with classifying videos as funny or not. You can use any combination of a CNN, an LSTM, and a fully connected feed forward network (FFN). How would you use them under the conditions discussed below. *Illustrate your answers with simple block diagrams (e.g. a FFN can be shown as a box with the word FFN written on it) or via written descriptions.*

      i. [Points: 0.5] You have a one image frame per video but you have a million videos?

      Solution: Diagram showing CNN to FNN output. There is sufficient data to train a CNN to extract features.

      ii. [Points 0.5] You only have 1000 images.

      Solution: There are insufficient instances to train a CNN. So use transfer learning to extract features and then connect the extracted features in a FFN.