CSE 493 G1/ 599 G1
Deep Learning
Winter 2024 Quiz 2

Feb 2, 2024

Full Name: ______________________________

UW Net ID: ______________________________

Question	Score
True/False (5 pts)	
Multiple Choice (8 pts)	
Short Answer (9 pts)	
Total (22 pts)	

Welcome to the CSE 493 G1 Quiz 2!

- The exam is 20 min and is **double-sided**.
- No electronic devices are allowed.

I understand and agree to uphold the University of Washington Honor Code during this exam.

Signature: ____________________ Date: ____________

Good luck!

This page is left blank for scratch work only. DO NOT write your answers here.

1 True / False (5 points) - Recommended 5 Minutes

Fill in the circle next to True or False, or fill in neither. Fill it in completely like this: ●. No explanations are required.

Scoring: Correct answer is worth 1 points.

1.1 ReLU (Rectified Linear Unit) activation functions can suffer from the problem of dead neurons.

◯ True

◯ False

1.2 One downside of layer normalization is that it works differently during training and testing, leading to a common source of errors.

◯ True

◯ False

1.3 Batch normalization is applied to the input of each layer before the activation function.

◯ True

◯ False

1.4 Dropout regularization can be applied only to fully connected layers in a neural network.

◯ True

◯ False

1.5 Saliency maps in convolutional neural networks (CNNs) are primarily used to reduce the dimensionality of the output layer for easier interpretation and visualization.

◯ True

◯ False

2 Multiple Choices (8 points) - Recommended 6 Minutes

Fill in the circle next to the letter(s) of your choice (like this: ●). No explanations are required. **Choose ALL options that apply.**

Each question is worth 4 points and the answer may contain one or more options. Selecting all of the correct options and none of the incorrect options will get full credits. For questions with multiple correct options, each incorrect or missing selection gets a 2-point deduction (up to 4 points).

2.1 During the training of a deep neural network using Stochastic Gradient Descent (SGD), you notice that the model's performance on the validation set starts declining after a certain number of epochs, even though the loss on the training set is still going down and the performance continues to improve. What would you do to improve the model's training (Treat each option as independent of each other)? Select ALL options that apply.

○ A: Decrease the regularization parameter λ.

○ B: Reduce the number of layers in the neural network.

○ C: Increase the number of neurons on each layer of the neural network.

○ D: Increase the dropout rate.

○ E: Use more powerful GPUs.

2.2 Which of the following would you consider to be valid activation functions (elementwise non-linearities) to train a neural net in practice?

○ A: $f(x) = max(0, x)$

○ B: $f(x) = 0.8x + 1$

○ C: $f(x) = \begin{cases} \min(x, .1x) & | \quad x >= 0 \\ \min(x, .1x) & | \quad x < 0 \end{cases}$

○ D: $f(x) = \begin{cases} \max(x, .1x) & | \quad x >= 0 \\ \min(x, .1x) & | \quad x < 0 \end{cases}$

○ E: $f(x) = \begin{cases} \max(x, .111x) & | \quad x >= 0 \\ \min(x, .111x) & | \quad x < 0 \end{cases}$

3 Short Answers (9 points) - Recommended 9 Minutes

Please make sure to write your answer only in the provided space.

1. (2D Convolution) (9 points) We have introduced Convolutional Neural Networks (CNNs) and the concept of 2D convolution in our lecture. Convolutional operations play a pivotal role in the field of deep learning, especially in the processing and understanding of image data. These operations help in extracting important features from the input data by applying filters that capture spatial hierarchies and patterns.

   In this question, you are provided with an input matrix I and a filter F, represented as:

$$I = \begin{array}{|c|c|c|c|} \hline 0 & 2 & 1 & 2 \\ \hline 0 & 3 & 2 & 2 \\ \hline 0 & 2 & 1 & 1 \\ \hline 1 & 0 & 1 & 0 \\ \hline \end{array} \qquad F = \begin{array}{|c|c|} \hline 1 & 0 \\ \hline 0 & 1 \\ \hline \end{array}$$

   In this question, you will apply the filter F to the input matrix I using 2D convolution operations under different padding and stride configurations. Padding adds a border of zeros to the input matrix to control the spatial size of the output feature map. Stride determines the step size the filter takes as it slides over the input matrix.

   Your task is to calculate the output matrices for the following configurations. For each configuration, write your answer in the grid with the right size. **Note: If the grid is too small, you can draw your own on the right empty space of each part.**

   (a) (2 points) Apply F on I with **Padding = 0, Stride = 1**

   (b) (2 points) Apply F on I with **Padding = 0, Stride = 2**

   (c) (2 points) Apply max pool on I with 2×2 filter and **Stride =2**

(d) (3 points) Apply F on I with **Padding = 1, Stride = 1**