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

October 25, 2024

Full Name: ____

UW Net ID: ____

Question		Score
True/False	(4 pts)	
Multiple Choice	(8 pts)	
Short Answer	(8 pts)	
Total	(20 pts)	

Welcome to the CSE 493 G1 Quiz 2!

- The exam is 30 min and is **double-sided**.
- No electronic devices are allowed.

I agree to uphold the University of Washington Student Conduct Code during this exam.

Signature: ______________________________ Date: __________________

Good luck!

1 True / False (4 points) - Recommended 4 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 You know you have a good training regime if there is no gap between your train accuracy and val accuracy.

○ True

○ False

1.2 One of the reasons that ReLU is better than Sigmoid is that its outputs are zero-centered.

○ True

○ False

1.3 After applying max pooling with spatial extent 2, stride 2, and zero padding on an input with $W \times H \times C$, (where W, C, H are all divisible by 2), the number of values in the output shape is one eighth that of the input.

○ True

○ False

1.4 Saliency maps compute the gradient of class scores with respect to image pixels to understand which pixels are important for the final prediction of the model.

○ True

○ False

2 Multiple Choices (8 points) - Recommended 8 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 Let A_1 and A_2 be two activation maps both of size $W \times H \times C$. They correspond to two elements in a batch of size 2. Let $2 \cdot A_1 = A_2$. Let B_1, B_2 correspond to A_1, A_2 after applying some normalization N. Which of the following are always true regardless of the value of A_1 and A_2?

○ A: If N is layer norm, then $B_1 = B_2$

○ B: If N is batch norm, then $B_1 = B_2$

○ C: If N is instance norm, then $B_1 = B_2$

○ D: The $L2$ norm of the elements in B_1 will be smaller than that of A_1 if N is layer norm.

2.2 Which of the following is true about optimizing models?

○ A: Momentum can cause models to reach the minimum more quickly.

○ B: Momentum can cause models to reach the minimum more slowly.

○ C: Given a constant gradient at each time step, the update to the weights will *decrease* over time when using AdaGrad.

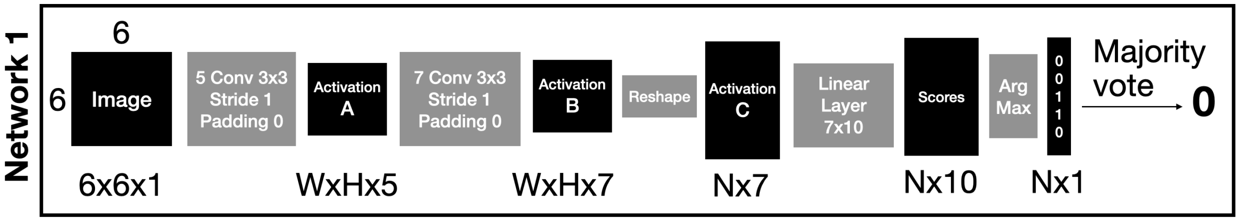
○ D: Given a constant gradient at each time step, the update to the weights will *increase* over time when using SGD without momentum.

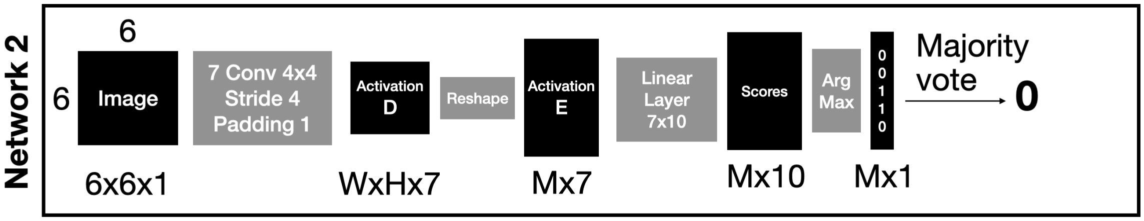
3 Short Answers (8 points) - Recommended 8 Minutes

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

3.1 Using CNNs to classify numbers

You are using CNNs to classify handwritten numbers as a 0 or 1. However, you decide to do something a little different. Instead of totally flattening out the activations after the CNNs, you decide to classify the letter at each location of the activation map and then take a majority vote over all the classifications. You try this out with two networks which you fully train using cross-entropy loss.





1. What is the shape of Activation A? (2 pnts)

2. What is the shape of Activation C? (2 pnts)

3. What is the shape of Activation E? (2 pnts)

4. You fully train both networks with the same data, such that they have very little train and test error. You can tell from the training curves that neither model is underfit or overfit. You then attempt to classify the two images below. Both networks correctly classify Image A. However, only Network 1 successfully classifies Image B. Network 2 classifies it as a 0, with a 75% vote for 0 and a 25% vote for 1. What about the architecture of Network 2 made this happen? Why is this not the case for Network 1? (2 pnts)

Images of 1's

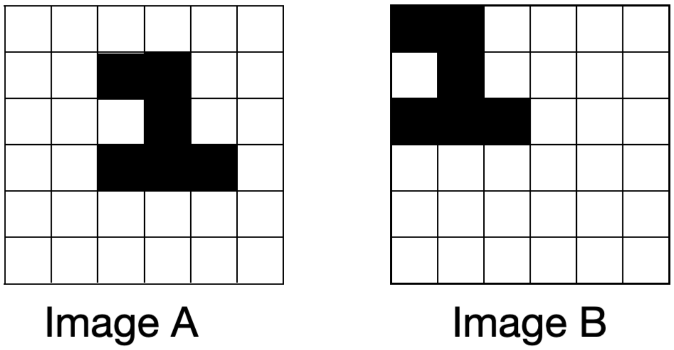


Image A

Image B