Quiz 3

Karan Vora (username: kv2154)

Attempt 1

Written: Apr 18, 2023 9:25 PM - Apr 18, 2023 10:04 PM

Submission View

Released: Apr 5, 2023 10:35 PM

Question 1   1.5 / 1.5 points

Consider convolution of a 5x5 input with a 3x3 filter, stride=1, no padding giving a 3x3 output. When writing the convolution as a matrix multiplication using Im2col, what is the size of the corresponding sparse matrix?

- ○ 5x9
- ○ 5x5
- ✓ ○ 9x25
- ○ 9x5

Question 2   1 / 1 point

Which ones of the following statements on Residual Networks are true? (Check all that apply.)

- ✓ ☐ The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.
- ✓ ☐ Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks.
- ✓ ☐ The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.
- ✓ ☐ A ResNet with L layers would have on the order of L^2 skip connections in total.

Question 3   0 / 1 point

Consider convolution of a 5x5 input with a 3x3 filter, stride=1, no padding giving a 3x3 output. When writing the convolution as a matrix multiplication using dense matrix reorganization as

A*X=Y,

where A is a dense matrix, X is a column vector, and Y is the output of convolution (as a column vector), what is the dimension of A and X?

- ○ 4x9
- ⇒ ○ 9x9
- ○ 5x5
- ✗ ○ 25x9

Question 4   0.25 / 1 point

Select all that is true regarding the work and depth of four different methods to do convolution: direct, im2col, FFT, and Winograd.

- ⇒ ✗ ☐ Depth of minibatch of size N is the same as minibatch of size 2N for all the four methods.
- ⇒ ✓ ☐ FFT work and depth are independent of the size of the filter kernel.
- ✗ ☐ Winograd depth is independent of tile size.
- ⇒ ✗ ☐ While work and depth for direct and im2col are same.

Question 5   0 / 1 point

You have an input volume that is 63x63x16, find the output volume when:

1. Convolve it with 32 filters that are each 7x7, using a stride of 2 and no padding _____

2. Convolve it with 32 filters that are each 7x7, using stride of 1 and padding of 3 _____

3. Apply max pooling with a stride of 2 and a filter size of 3 and no padding _____

Answer for blank # 1: 26912   ✗ **(29x29x32, 29*29*32)**
Answer for blank # 2: 127008   ✗ **(63x63x32, 63*63*32)**
Answer for blank # 3: 16384   ✗ **(31x31x16, 31*31*16)**

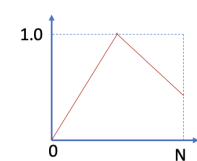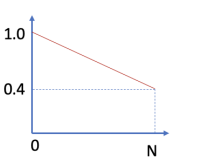**Question 6**   0 / 1 point

According to the Work-depth cost model, what is the work and depth for calculating the sum of 64 floating-point numbers?

- ✗ ○ 16 and 63
- ⇒ ○ 6 and 63
- ○ 6 and 64
- ○ 16 and 64
- ○ 16 and 31

Question 7   2 / 2 points

There are six candidates for assigning dropout probability to different layers of a deep neural network during training. A dropout profile is represented as a function of the depth of neural network layer, with input layer being at depth 0 and last hidden layer at depth N. Correctly identify all the profiles which will result in a dead neural network meaning the network will not learn anything (weights will not get updated) during training. Each profile is shown as a graph with x-axis being the depth of the layer and y axis being the drop-out probability.

✔☐

✔☐

✔☐

✔☐

✔☐

✔☐
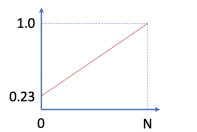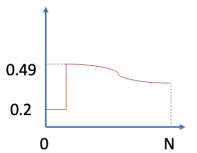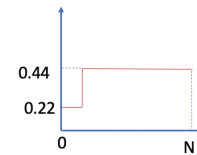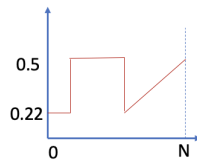
Question 8   1 / 1 point

Suppose your input is a 200 by 200 color (RGB) image, and you are not using a convolutional network. If the first hidden layer has 200 neurons, each one fully connected to the input, how many parameters does this hidden layer have (including the bias parameters)

○ 8,000,000

✔○ 24,000,200

○ 8,000,200

○ 24,002,000

○ 24,000,000

Question 9   0 / 1 point

Consider a "same convolution" with C channel input and a C 5x5 filter. So the output channel depth is also C. Instead of a 5x5 filter if we use a stack of three layers as

1x1, C/2 → 5x5, C/2 → 1x1xC/2,

how much will be the parameter saving given as the ratio between the number of parameters in the single-layer approach to the number of parameters in the 3-layer stack approach?

○ 10

✗ ○ 2.5

⇨ ○ 3.5

○ 20

○ 4.5

Question 10   0 / 2 points

Suppose you have an input volume of dimension nH x nW x nC. Which of the following statements you agree with? (Assume that "1x1 convolutional layer" below always uses a stride of 1 and no padding.)

✗ ☐ You can use a 1x1 convolutional layer to reduce nH, nW, and nC.

✗ ☐ You can use a pooling layer to reduce nH, nW, and nC.

⇨ ✗ ☐ You can use a 1x1 convolutional layer to reduce nC but not nH, nW.

⇨ ✗ ☐ You can use a pooling layer to reduce nH, nW, but not nC.

Question 11   1 / 1 point

What is the output after the following input is subjected to two consecutive maxpooling, first with a 3x3 window and stride 1 and next with a 3x3 window and stride 1.

7	7	5	5	5
7	7	5	5	7
8	8	5	5	7
8	8	6	6	7
8	8	6	6	6

- ○ 5
- ○ 7
- ✓ ○ 8
- ○ None of the options
- ○ It will be a 3x3 matrix with all identical rows.
- ○ 887

Question 12   0.75 / 1 point

Which of the following do you typically see in a ConvNet? (Check all that apply.)

- ⇒ ✓ ☐ FC layers in the last few layers
- ✗ ☐ Multiple POOL layers followed by a CONV layer
- ✓ ☐ FC layers in the first few layers
- ⇒ ✓ ☐ Multiple CONV layers followed by a POOL layer

Question 13   0 / 1 point

Consider Winograd minimum filtering for computing 4 outputs with a 3-tap filter:

$F(4,3) = A*X=Y$,

where A is a 4x3 matrix. How many multiplications (MULs) are needed to compute Y?

- ⇒ ○ 6
- ○ 18
- ✗ ○ 12
- ○ 10
- ○ 4

Question 14   1 / 1.5 points

Which ones of the following statements on Inception Networks are true? (Check all that apply.)

- ⇒ ✓ ☐ A single inception block allows the network to use a combination of 1x1, 3x3, 5x5 convolutions and pooling.
- ✗ ☐ Inception networks incorporate a variety of network architectures (similar to dropout, which randomly chooses a network architecture on each step) and thus has a similar regularizing effect as dropout.
- ⇒ ✓ ☐ Inception blocks usually use 1x1 convolutions to reduce the input data volume's size before applying 3x3 and 5x5 convolutions.

Question 15   2 / 2 points

Following is a portion of a RestNet.
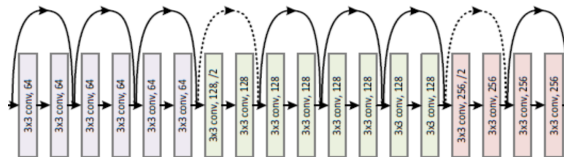
What is the significance of dotted skip connections? (select all that apply)

- ✓ ☐ All of the options are correct.
- ✓ ☐ They may be needed to match the size of the input activation maps so that they can be added properly.
- ✓ ☐ They are needed to match the input channel dimensions e.g., the channel depth of input to skip connection is 64 which needs to be added to an input with 128 depth in the first dotted connection.
- ✓ ☐ They are optional and their presence does not affect the output.
- ✓ ☐ None of the options are correct.

Question 16   0.75 / 1 point

One benefit of using convolutional networks is "parameter sharing". Which of the following statements about parameter sharing in ConvNets are true? (Check all that apply.)

- ⇒ ✓ ☐ It allows a feature detector to be used in multiple locations throughout the whole input image/input volume.
- ⇒ ✓ ☐ It reduces the total number of parameters, thus reducing overfitting.
- ✗ ☐ It allows parameters learned for one task to be shared even for a different task (transfer learning).

✓☐ It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.

Question 17   1 / 1 point

cuDNN provides up to eight different algorithms to perform convolutions, each of which requires different temporary storage (workspace). Shown below is the per-layer breakdowns of memory consumption (left axis, bars) and computation time of forward and backward passes (right axis, points) of AlexNet's convolutional layers on P100-SXM2. There are three different workspace sizes (8, 64, 512 MiB), and a mini-batch of 256.
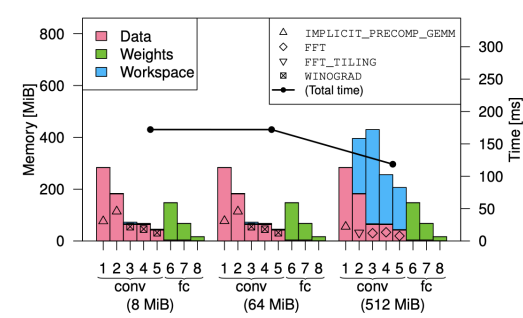
Based on this figure answer the following statements as True/False:

1. FFT_TILING may be faster than GEMM however GEMM is more memory efficient. _____
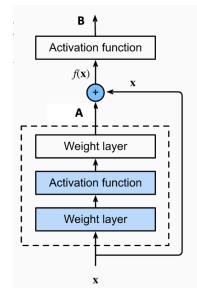2. Winograd requires a larger workspace than FFT. _____

Answer for blank # 1: True   ✓(50 %)
Answer for blank # 2: False   ✓(50 %)

Question 18   2 / 2 points

Consider the following residual block:

Say the activation function is ReLU. What is the value of B? (select all that apply)

✓☐ ReLU(A+2x)

✓☐ ReLU(A+x)

✓☐ ReLU(f(x)-x)

✓☐ ReLU(ReLU(A)+x)

✓☐ ReLU(f(x))

Question 19   1 / 1 point

In P2L paper, an estimator is proposed to identify suitability of a source dataset for transfer learning for a target dataset. The estimator is a function of both the size of the source dataset and a measure of dissimilarity between the source and target dataset. Choose all that is conformant with the observations in the paper.

✓☐ Transferability of a source dataset improves with its size.

✓☐ Larger the dissimilarity between the source and target datasets, lower is the transferability of the source.

✓☐ Given two datasets with same dissimilarity with the target dataset, the one with smaller size is a better candidate for transfer learning.

✓☐ Given two source datasets with same size, the one with larger dissimilarity will be a better candidate for transfer learning.

Question 20   1 / 1 point

Convolutional layers in deep neural networks exhibit "sparsity of connections". What does this mean?

○ Each filter is connected to every channel in the previous layer.

✓○ Each activation in the next layer depends on only a small number of activations from the previous layer.

○ Regularization causes gradient descent to set many of the parameters to zero.

○ Each layer in a convolutional network is connected only to two other layers

---

Attempt Score: 16.25 / 25 - D
Overall Grade (highest attempt): 16.25 / 25 - D

Done