CS410: Artificial Intelligence 2020 Fall
Homework 3: Neural Networks and Convolutional Neural Networks
Due date: 23:59:59 (GMT +08:00), November 24, 2020

1 Single-choice Questions

1. What does a neuron compute?
A. A neuron computes an activation function followed by a linear function ($z = Wx + b$)
B. A neuron computes a linear function ($z = Wx + b$) followed by an activation function
C. A neuron computes a function g that scales the input x linearly ($Wx + b$)
D. A neuron computes the mean of all features before applying the output to an activation function

2. The tanh activation usually works better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data better for the next layer. True/False?
A. True
B. False

3. You are building a binary classifier for recognizing cucumbers ($y = 1$) vs. watermelons ($y = 0$). Which one of these activation functions would you recommend using for the output layer?
A. ReLU
B. Leaky ReLU
C. sigmoid
D. tanh

4. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relative large values, using np.random.randn(..,..)*1000. What will happen?
A. It doesn't matter. So long as you initialize the weights randomly gradient descent is not affected by whether the weights are large or small.
B. This will cause the inputs of the tanh to also be very large, thus causing gradients to also become large. You therefore have to set α to be very small to

prevent divergence; this will slow down learning.
C. This will cause the inputs of the tanh to also be very large, causing the units to be "highly activated" and thus speed up learning compared to if the weights had to start from small values.
D. This will cause the inputs of the tanh to also be very large, thus causing gradients to be close to zero. The optimization algorithm will thus become slow.

5. Which of the following statements is true?
A. The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.
B. The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.

6. During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l, since the gradient depends on it. True/False?
A. True
B. False

7. You have an input volume that is $15 \times 15 \times 8$, and pad it using "pad=2." What is the dimension of the resulting volume (after padding)?
A. $17 \times 17 \times 10$
B. $19 \times 19 \times 8$
C. $19 \times 19 \times 12$
D. $17 \times 17 \times 8$

8. You have an input volume that is $32 \times 32 \times 16$, and apply max pooling with a stride of 2 and a filter size of 2. What is the output volume?
A. $15 \times 15 \times 16$
B. $16 \times 16 \times 8$
C. $16 \times 16 \times 16$
D. $32 \times 32 \times 8$

9. You have an input volume that is $63 \times 63 \times 16$, and convolve it with 32 filters that are each 7×7, and stride of 1. You want to use a "same" convolution. What is the padding?

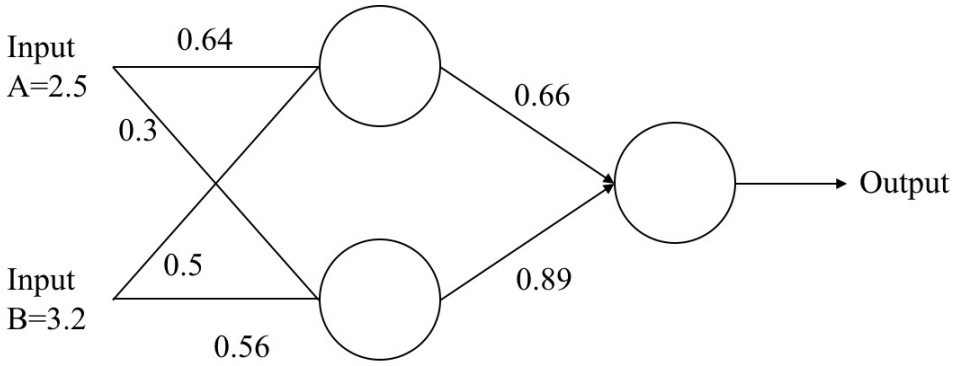
A. 1
B. 2
C. 3
D. 7

10. The "sparsity of connections" is a benefit of using convolutional layers. What does this mean?
A. Regularization causes gradient descent to set many of the parameters to zero.
B. Each filter is connected to every channel in the previous layer.
C. Each activation in the next layer depends on only a small number of activations from the previous layer.
D. Each layer in a convolutional network is connected only to two other layers.

2 Calculation Questions (Please provide the detailed calculation process)

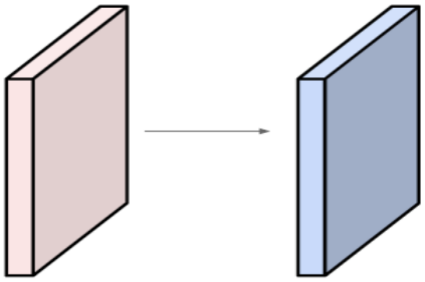
Question 1 (Neural Networks)
1. Retell the calculation process of backpropagation on Page 45 of the Lecture 6 slides.

2. Consider the simple network below:



(1). Assume that the neurons have sigmoid activation function
(2). Perform a forward pass on the network and find the predicted output
(3). Perform a reverse pass (training) once (target = 0.5) with $\eta = 1$
(4). Perform a further forward pass and comment on the result

Question 2 (Convolutional Neural Networks)

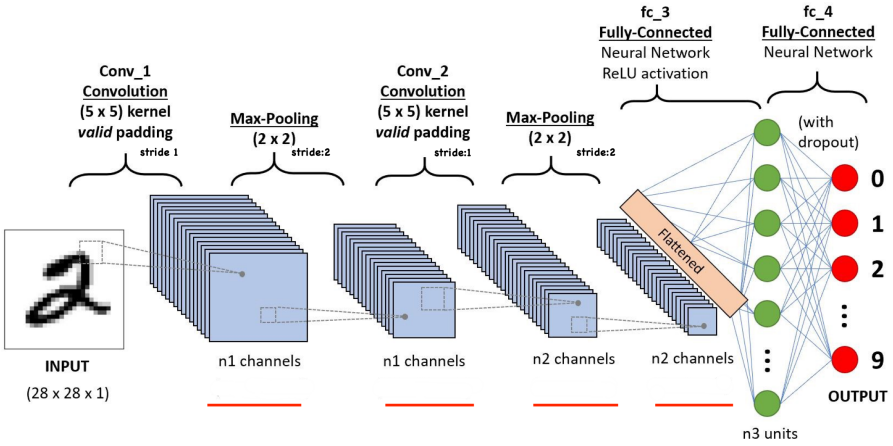


1. Suppose a convolutional layer with input volume $32 \times 32 \times 3$ and 10 5×5 filters with stride 1, pad 2.

(1). What is the output volume size of this layer? Show the computation steps of how to derive the volume size from the form layer.

(2). What is the total number of parameters in this layer?

2. For the following LeNet-5 for the handwritten digit recognition task:



(1). Please calculate the output volume size of each layer (red line) and show the computation steps.

(2). What is the total number of parameters in each layer?

3 Short Answer Questions

1. Try to summarize common activation functions and their characteristics.

2. Summarize methods to avoid overfitting when training the neural network.

3. Describe the motivation of batch normalization and analyze how this method solve the initial problem.