

Week3-Quiz 1

10 May 2020 00:53

Browser tabs: Inbox (5,615) · sri... [1] WhatsApp Coursera Respons... Shallow Neural Ne... deep-learning-co... Neural Networks 29200515-7fd15- 29200515-7fd15- + -

Address bar: coursera.org/learn/neural-networks-deep-learning/exam/7llnl/shallow-neural-networks/attempt

Navigation: Apps YouTube Gmail In-Browser Tutorial... GitHub · yanshengj... ai Telugu Movies 202... Google Images Krish Naik - YouTube Sci-Hub: removing... research papers Optical Character R...

Page Header: Shallow Neural Networks Graded Quiz · 30 min Due May 25, 11:59 AM +05

Shallow Neural Networks

TOTAL POINTS 10

1. Which of the following are true? (Check all that apply.) 1 point
- ☒ X is a matrix in which each column is one training example.
 - ☒ $a_4^{(2)}$ is the activation output by the 4th neuron of the 2nd layer
 - ☒ $a^{(2)}$ denotes the activation vector of the 2nd layer.
 - ☐ $a_4^{(2)}$ is the activation output of the 2nd layer for the 4th training example
 - ☐ $a^{(2)(12)}$ denotes activation vector of the 12th layer on the 2nd training example.
 - ☐ X is a matrix in which each row is one training example.
 - ☒ $a^{(2)(12)}$ denotes the activation vector of the 2nd layer for the 12th training example.
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? 1 point
- ☒ True
- ☐ False

Windows taskbar: Windows Search File Explorer Edge Chrome Visual Studio Code

System tray: Network Volume ENG 00:58 10-05-2020

Browser tabs: Inbox (5,615) · sri... [1] WhatsApp Coursera Respons... Shallow Neural Ne... deep-learning-co... Neural Networks 29200515-7fd15- 29200515-7fd15- + -

Address bar: coursera.org/learn/neural-networks-deep-learning/exam/7llnl/shallow-neural-networks/attempt

Navigation: Apps YouTube Gmail In-Browser Tutorial... GitHub · yanshengj... ai Telugu Movies 202... Google Images Krish Naik - YouTube Sci-Hub: removing... research papers Optical Character R...

Page Header: Shallow Neural Networks Graded Quiz · 30 min Due May 25, 11:59 AM +05

3. Which of these is a correct vectorized implementation of forward propagation for layer l , where $1 \leq l \leq L$? 1 point
- ☐ $Z^{(l)} = W^{(l)} A^{(l)} + b^{(l)}$
 $A^{(l+1)} = g^{(l+1)}(Z^{(l)})$
- ☐ $Z^{(l)} = W^{(l-1)} A^{(l)} + b^{(l-1)}$
 $A^{(l)} = g^{(l)}(Z^{(l)})$
- ☒ $Z^{(l)} = W^{(l)} A^{(l-1)} + b^{(l)}$
 $A^{(l)} = g^{(l)}(Z^{(l)})$
- ☐ $Z^{(l)} = W^{(l)} A^{(l)} + b^{(l)}$
 $A^{(l+1)} = g^{(l)}(Z^{(l)})$
4. 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? 1 point
- ☐ ReLU
- ☐ Leaky ReLU
- ☒ sigmoid
- ☐ tanh
5. Consider the following code: 1 point

Windows taskbar: Windows Search File Explorer Edge Chrome Visual Studio Code

System tray: Network Volume ENG 00:59 10-05-2020

Inbox (5,615) ×(1) WhatsApp ×Coursera Response ×Shallow Neural Networks ×deep-learning-co... ×Neural Networks ×29200515-7fdd15... ×29200515-7fdd15... ×+ - □ ×

coursera.org/learn/neural-networks-deep-learning/exam/7lln/shallow-neural-networks/attempt

Apps YouTube Gmail In-Browser Tutorial... GitHub - yanshengj... ai Telugu Movies 202... Google Images Krish Naik - YouTube Sci-Hub: removing... research papers Optical Character R...

← Shallow Neural NetworksGraded Quiz • 30 minDue May 25, 11:59 AM +05

(3)

6. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true?

1 point

☒ Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.

☐ Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have "broken symmetry".

☐ Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in lecture.

☐ The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

7. Logistic regression's weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to "break symmetry". True/False?

1 point

☐ True

☒ False

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

☐ The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

7. Logistic regression's weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to "break symmetry", True/False? 1 point

☐ True

☒ False

8. 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? 1 point

☐ 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.

☐ 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.

☐ 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.

☒ 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.

9. Consider the following 1 hidden layer neural network: 1 point

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

9. Consider the following 1 hidden layer neural network: 1 point

Which of the following statements are True? (Check all that apply).

☐ $W^{(1)}$ will have shape (2, 4)

☒ $b^{(1)}$ will have shape (4, 1)

☒ $W^{(1)}$ will have shape (4, 2)

☐ $b^{(1)}$ will have shape (2, 1)

☒ $W^{(2)}$ will have shape (1, 4)

☐ $b^{(2)}$ will have shape (4, 1)

☐ $W^{(2)}$ will have shape (4, 1)

☒ $b^{(2)}$ will have shape (1, 1)

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

Which of the following statements are True? (Check all that apply.)

- ☐ $W^{(1)}$ will have shape (2, 4)
- ☒ $b^{(1)}$ will have shape (4, 1)
- ☒ $W^{(2)}$ will have shape (4, 2)
- ☐ $b^{(1)}$ will have shape (2, 1)
- ☒ $W^{(2)}$ will have shape (1, 4)
- ☐ $b^{(2)}$ will have shape (4, 1)
- ☐ $W^{(2)}$ will have shape (4, 1)
- ☒ $b^{(2)}$ will have shape (1, 1)

10. In the same network as the previous question, what are the dimensions of $Z^{(1)}$ and $A^{(1)}$?

1 point

- ☐ $Z^{(1)}$ and $A^{(1)}$ are (4,2)
- ☒ $Z^{(1)}$ and $A^{(1)}$ are (4,m)
- ☐ $Z^{(1)}$ and $A^{(1)}$ are (1,4)
- ☐ $Z^{(1)}$ and $A^{(1)}$ are (4,1)

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

1. Which of the following are true? (Check all that apply.)

1 / 1 point

- ☒ X is a matrix in which each column is one training example.
- ☒ $a_4^{(2)}$ is the activation output by the 4th neuron of the 2nd layer.
- ☒ $a^{(2)}$ denotes the activation vector of the 2nd layer.
- ☐ $a_4^{(2)}$ is the activation output of the 2nd layer for the 4th training example.
- ☐ $a^{(2)(12)}$ denotes activation vector of the 12th layer on the 2nd training example.
- ☐ X is a matrix in which each row is one training example.
- ☒ $a^{(2)(12)}$ denotes the activation vector of the 2nd layer for the 12th training example.

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?

1 / 1 point

- ☒ True
- ☐ False

✓ Correct

Yes. As seen in lecture the output of the tanh is between -1 and 1, it thus centers the data which makes the learning simpler for the next layer.

3. Which of these is a correct vectorized implementation of forward propagation for layer L , where $1 \leq L \leq L$?

1 / 1 point

- ☐ $Z^{[L]} = W^{[L]}A^{[L]} + b^{[L]}$
 $A^{[L+1]} = g^{[L+1]}(Z^{[L]})$
- ☐ $Z^{[L]} = W^{[L-1]}A^{[L]} + b^{[L-1]}$
 $A^{[L]} = g^{[L]}(Z^{[L]})$
- ☒ $Z^{[L]} = W^{[L]}A^{[L-1]} + b^{[L]}$
 $A^{[L]} = g^{[L]}(Z^{[L]})$
- ☐ $Z^{[L]} = W^{[L]}A^{[L]} + b^{[L]}$
 $A^{[L+1]} = g^{[L]}(Z^{[L]})$

✓ Correct

4. You are building a binary classifier for recognising cucumbers ($y=1$) vs. watermelons ($y=0$). Which one of these activation functions would you recommend using for the output layer?

1 / 1 point

- ☐ ReLU
- ☐ Leaky ReLU
- ☒ sigmoid
- ☐ tanh

✓ Correct

Yes. Sigmoid outputs a value between 0 and 1 which makes it a very good choice for binary classification. You can classify as 0 if the output is less than 0.5 and classify as 1 if the output is more than 0.5. It can be done with tanh as well but it is less convenient as the output is between -1 and 1.

5. Consider the following code:

1 / 1 point

```
1 A = np.random.randn(4,3)
2 B = np.vstack([A, A]).T
```

What will be B.shape? (If you're not sure, feel free to run this in python to find out).

- ☒ (4, 1)
- ☐ (1, 3)
- ☐ (4,)
- ☐ (, 3)

✓ Correct

Yes, we use (keepdims = True) to make sure that A.shape is (4,1) and not (4,). It makes our code more rigorous.

Inbox (5,615) x (1) WhatsApp x Coursera Respons... x Shallow Neural N... x deep-learning-co... x Neural Networks x 29200515-7fd15... x 29200515-7fd15... x

← → ↻ ⌂ coursera.org/learn/neural-networks-deep-learning/exam/7llnl/shallow-neural-networks/attempt?redirectToCover=true

Apps YouTube Gmail In-Browser Tutorial... GitHub - yanshengj... ai Telugu Movies 202... Google Images Krish Naik - YouTube Sci-Hub: removing... research papers Optical Character R...

← Shallow Neural Networks Graded Quiz • 30 min Due May 25, 11:59 AM +05

6. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true? 1 / 1 point

- ☒ Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
- ☐ Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have "broken symmetry".
- ☐ Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in lecture.
- ☐ The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

✓ Correct

7. Logistic regression's weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to "break symmetry". True/False? 1 / 1 point

- ☐ True
- ☒ False

✓ Correct

Yes, Logistic Regression doesn't have a hidden layer, if you initialize the weights to zeros, the first example x fed in the logistic regression will output zero but the derivatives of the Logistic Regression depend on the input x (because there's no hidden layer) which is not zero. So at the second iteration, the weights values follow x 's distribution and are different from each other if x is not a constant vector.

Inbox (5,615) x (1) WhatsApp x Coursera Respons... x Shallow Neural N... x deep-learning-co... x Neural Networks x 29200515-7fd15... x 29200515-7fd15... x

← → ↻ ⌂ coursera.org/learn/neural-networks-deep-learning/exam/7llnl/shallow-neural-networks/attempt?redirectToCover=true

Apps YouTube Gmail In-Browser Tutorial... GitHub - yanshengj... ai Telugu Movies 202... Google Images Krish Naik - YouTube Sci-Hub: removing... research papers Optical Character R...

← Shallow Neural Networks Graded Quiz • 30 min Due May 25, 11:59 AM +05

✓ Correct

Yes, Logistic Regression doesn't have a hidden layer. If you initialize the weights to zeros, the first example x fed in the logistic regression will output zero but the derivatives of the Logistic Regression depend on the input x (because there's no hidden layer) which is not zero. So at the second iteration, the weights values follow x 's distribution and are different from each other if x is not a constant vector.

8. 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? 1 / 1 point

- ☐ 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.
- ☐ 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.
- ☐ 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.
- ☒ 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.

✓ Correct

Yes, tanh becomes flat for large values, this leads its gradient to be close to zero. This slows down the optimization algorithm.

9. Consider the following 1 hidden layer neural network: 1 / 1 point

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

Which of the following statements are True? (Check all that apply).

- ☐ $W^{(1)}$ will have shape (2, 4)
- ☒ $b^{(1)}$ will have shape (4, 1)

✓ Correct
- ☒ $W^{(1)}$ will have shape (4, 2)

✓ Correct
- ☐ $b^{(1)}$ will have shape (2, 1)
- ☒ $W^{(2)}$ will have shape (1, 4)

✓ Correct
- ☐ $b^{(2)}$ will have shape (4, 1)
- ☐ $W^{(2)}$ will have shape (4, 1)
- ☒ $b^{(2)}$ will have shape (1, 1)

✓ Correct

Shallow Neural Networks
Graded Quiz • 30 min

Due May 25, 11:59 AM +05

✓ Correct

- ☐ $b^{(1)}$ will have shape (2, 1)
- ☒ $W^{(2)}$ will have shape (1, 4)

✓ Correct
- ☐ $b^{(2)}$ will have shape (4, 1)
- ☐ $W^{(2)}$ will have shape (4, 1)
- ☒ $b^{(2)}$ will have shape (1, 1)

✓ Correct

10. In the same network as the previous question, what are the dimensions of $Z^{(1)}$ and $A^{(1)}$?

1 / 1 point

- ☐ $Z^{(1)}$ and $A^{(1)}$ are (4,2)
- ☒ $Z^{(1)}$ and $A^{(1)}$ are (4,m)
- ☐ $Z^{(1)}$ and $A^{(1)}$ are (1,4)
- ☐ $Z^{(1)}$ and $A^{(1)}$ are (4,1)

✓ Correct