Congratulations! You passed!

Grade received 100% To pass 80% or higher

Go to next item

Shallow Neural Networks

Latest Submission Grade 100%

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

- [ ] $a^{[2](13)}$ denotes activation vector of the 13th layer on the 2nd training example.
- [ ] X is a matrix in which each row is one training example.

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

Correct

- [ ] $a_4^{[2]}$ is the activation output of the 2nd layer for the 4th training example
- [x] X is a matrix in which each column is one training example.

Correct

- [x] $a^{[2]}$ denotes the activation vector of the 2nd layer.

Correct

- [x] $a^{[2](13)}$ denotes the activation vector of the 2nd layer for the 13th training example.

Correct

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

- (•) 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

- ( ) • $Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$
  • $A^{[l+1]} = g^{[l+1]}(Z^{[l]})$
- ( ) • $Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$
  • $A^{[l+1]} = g^{[l]}(Z^{[l]})$

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

- (•) • $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$
  • $A^{[l]} = g^{[l]}(Z^{[l]})$

Correct

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 / 1 point

- ( ) ReLU
- ( ) Leaky ReLU
- (•) sigmoid
- ( ) tanh

Correct

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

5. Consider the following code: 1 / 1 point

```
1  A = np.random.randn(4,3)
2  B = np.sum(A, axis = 1, keepdims = True)
```

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

- ( ) (, 3)
- ( ) (4,)
- (•) (4, 1)
- ( ) (1, 3)

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

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.

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

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.

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

( ) It doesn't matter. So long as you initialize the weights randomly gradient descent is not affected by whether

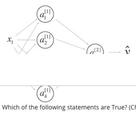
Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

optimization algorithm will thus become slow.

- ( ) 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.
- ( ) 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.

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



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

- [ ] $W^{[1]}$ will have shape (2, 4)
- [x] $b^{[1]}$ will have shape (4, 1)

Correct

Back
Shallow Neural Networks
Graded Quiz • 30 min
Due Jan 25, 2:59 AM EST

- [ ] $b^{[1]}$ will have shape (2, 1)
- [x] $W^{[2]}$ will have shape (1, 4)

Correct

- [ ] $b^{[2]}$ will have shape (4, 1)
- [ ] $W^{[2]}$ will have shape (4, 1)
- [x] $b^{[2]}$ 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 (4,1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4,2)
- (•) $Z^{[1]}$ and $A^{[1]}$ are (4,m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (1,4)