Congratulations! You passed!

Grade received 90% **Latest Submission Grade** 90% **To pass** 80% or higher

Go to next item

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

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

Expand

Correct
Great, you got all the right answers.

2. The tanh activation is not always better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data, making learning complex for the next layer. True/False? 0 / 1 point

- ( ) False
- (•) True

Expand

Incorrect
No. 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

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

Expand

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

- tanh
- Leaky ReLU
- (selected) sigmoid
- ReLU

Expand

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

```
#+begin_src python

x = np.random.rand(4, 5)

y = np.sum(x, axis=1)

#+end_src
```

What will be y.shape?

- (1, 5)
- (4, 1)
- (5,)
- (selected) (4,)

Expand

Correct
Yes. By using axis=1 the sum is computed over each row of the array, thus the resulting array is a column vector with 4 entries. Since the option keepdims was not used the array doesn't keep the second dimension.

6. Suppose you have built a neural network with one hidden layer and tanh as activation function for the hidden layers. Which of the following is a best option to initialize the weights? 1 / 1 point

- Initialize the weights to large random numbers.
- Initialize all weights to 0.
- Initialize the weights to small random numbers.
- Initialize all weights to a single number chosen randomly.

Expand

Correct
The use of random numbers helps to "break the symmetry" between all the neurons allowing them to compute different functions. When using small random numbers the values $z^{[k]}$ will be close to zero thus the activation values will have a larger gradient speeding up the training process.

7. A single output and single layer neural network that uses the sigmoid function as activation is equivalent to the logistic regression. True/False 1 / 1 point

- True
- False

Expand

Correct
Yes. The logistic regression model can be expressed by $\hat{y} = \sigma\left(W\,x + b\right)$. This is the same as $a^{[1]} = \sigma(W^{[1]}\,X + b)$.

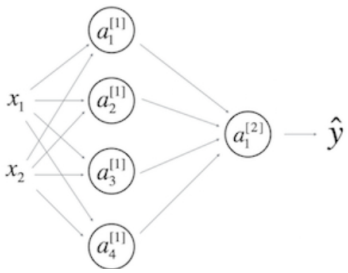
8. Which of the following is true about the ReLU activation functions? 1 / 1 point

- They are only used in the case of regression problems, such as predicting house prices.
- They are increasingly being replaced by the tanh in most cases.
- They are the go to option when you don't know what activation function to choose for hidden layers.
- They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.

Expand

Correct

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

✓ Correct

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

✓ Correct

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

✓ Correct

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

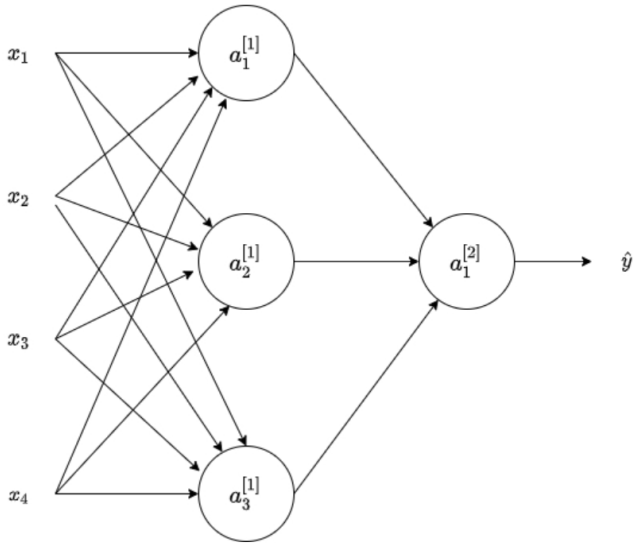
✓ Correct

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

Expand

✓ Correct
Great, you got all the right answers.

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



What are the dimensions of $Z^{[1]}$ and $A^{[1]}$?

- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, 1)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (3, 1)
- (•) $Z^{[1]}$ and $A^{[1]}$ are (3, m)
- ( ) $Z^{[1]}$ and $A^{[1]}$ are (4, m)

Expand

Correct

Yes. The $Z^{[1]}$ and $A^{[1]}$ are calculated over a batch of training examples. The number of columns in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of examples in the batch, m. And the number of rows in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of neurons in the first layer.