Note: In here I am adding True or False, it can be Right or Wrong both first Check If option is right or wrong under that question summary. And Rest question which I am giving multiple answer question or single choice question both are only right answers, and there is lot's of changing questions are there so please check all the question first and then only select the correct option. Also In Multiple Question I have selected all the option so it will show you all he correct option so based on that select the right answers.

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

- [x] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.

Correct

Yes. The vector $w_j^{[i]}$ is the column vector of parameters of the i-th layer and j-th neuron of that layer.

- [x] $W^{[1]}$ is a matrix with rows equal to the parameter vectors of the first layer.

! This should not be selected

No. The parameter vectors are column vectors.

- [x] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.

! This should not be selected

No. The convention is that the superscript number in brackets indicates the number of layers.

- [x] $W^{[1]}$ is a matrix with rows equal to the transpose of the parameter vectors of the first layer.

Correct

Yes. We construct $W^{[1]}$ stacking the parameter vectors $w_j^{[1]}$ of all the neurons of the first layer.

- [x] W_1 is a matrix with rows equal to the parameter vectors of the first layer.

! This should not be selected

No. The notation convention is that the superscript number in brackets indicates the number of layers.

- [x] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.

! This should not be selected

No. The vectors $w_k^{[j]}$ are column vectors.

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False?

1 / 1 point

- (•) False
- ( ) True

Expand

Correct

Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

4. When building a binary classifier for recognizing cats (y=1) vs raccoons (y=0). Is better to use the sigmoid function as activation function for the hidden layers. True/False

1 / 1 point

- (•) False
- ( ) True

Expand

Correct

Yes. Using tanh almost always works better than the sigmoid function for hidden layers.

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)
- ( ) (5,)
- (•) (4,)
- ( ) (4, 1)

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.

7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

0 / 1 point

- (•) False
- ( ) True

Expand

Incorrect
No. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

8. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relatively large values, using np.random.randn(..,..)*1000. What will happen?

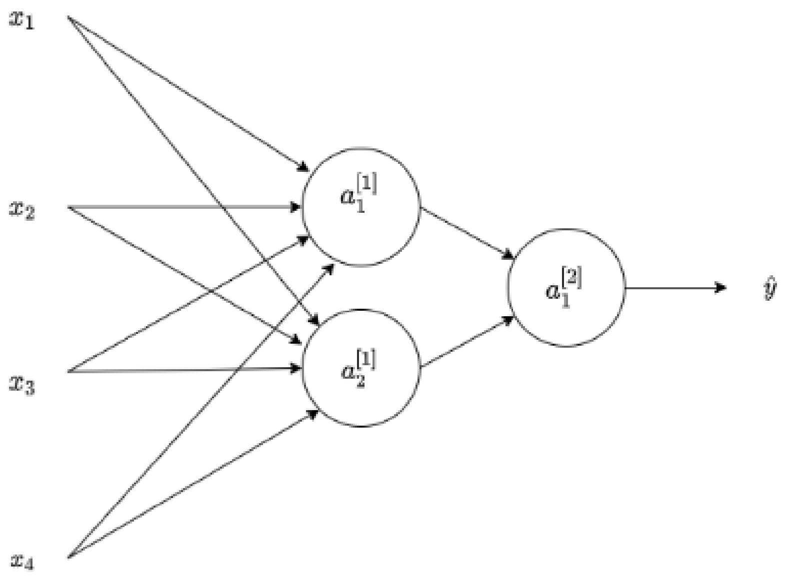
1 / 1 point

- (selected) 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.
- 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 a very small value to prevent divergence; this will slow down learning.
- 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.
- So long as you initialize the weights randomly gradient descent is not affected by whether the weights are large or small.

Expand

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:



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

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

! **This should not be selected**

No. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

✓ **Correct**

Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

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

✓ **Correct**

Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

✓ **Correct**

Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

0 / 1 point

- [x] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.

Correct

Yes. The vector $w_j^{[i]}$ is the column vector of parameters of the i-th layer and j-th neuron of that layer.

- [x] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.

This should not be selected

No. The vector $w_j^{[i]}$ is the column vector of parameters of the jth neuron in the i-th layer.

- [x] $a_3^{[2]}$ denotes the activation vector of the second layer for the third example.

This should not be selected

No. In our convention, the subscript refers to the neuron number.

- [x] $a^{[3](2)}$ denotes the activation vector of the second layer for the third example.

This should not be selected

No. In our convention, the superscript in brackets () indicates the example number, while superscript in the brackets [] indicates the layer.

- [x] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.

This should not be selected

No. The vectors $w_k^{[j]}$ are column vectors.

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

Correct

Yes. In our convention $a^{[j]}$ denotes the activation function of the j-th layer.

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

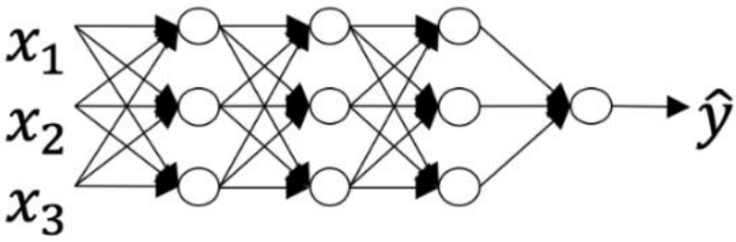
Expand

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 the following represents the activation output of the second neuron of the third layer applied to the fourth example?

1 / 1 point



- [x] $a_2^{[3](4)}$
- [ ] $a_2^{[4](3)}$
- [ ] $a_4^{[3](2)}$
- [ ] $a_3^{[4]2}$

Expand

Correct

Yes. The superscript in brackets indicates the layer number, the superscript in parenthesis represents the number of examples, and the subscript the number of the neuron.

4. When building a binary classifier for recognizing cats (y=1) vs raccoons (y=0). Is better to use the sigmoid function as activation function for the hidden layers. True/False

1 / 1 point

- [x] False
- [ ] True

Expand

Correct

Yes. Using tanh almost always works better than the sigmoid function for hidden layers.

5. Consider the following code:

1 / 1 point

A = np.random.randn(4,3)

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,)
- ( ) (1, 3)
- (•) (4, 1)

Expand

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

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 the 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.

Expand

Correct

7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

0 / 1 point

- (selected) False
- True

Expand

Incorrect

No. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

8. You have built a network using the tanh activation for all the hidden units. You initialize the weights to relatively large values, using np.random.randn(..,..)*1000. What will happen?

1 / 1 point

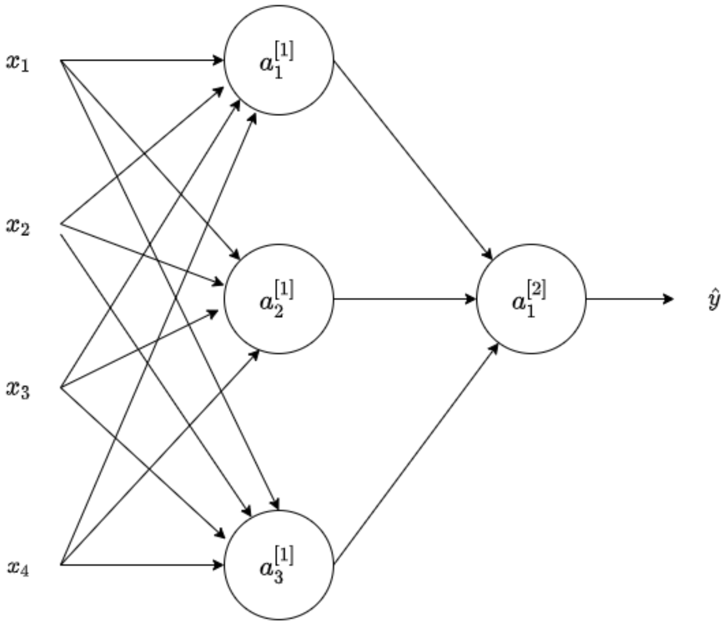
- (selected) This will cause the inputs of the tanh to also be very large, thus causing gradients to be close to zero. The optimization algorithm will thus become slow.
- This will cause the inputs of the tanh to also be very large, thus causing gradients to also become large. You therefore have to set α to a very small value to prevent divergence; this will slow down learning.
- 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.
- So long as you initialize the weights randomly gradient descent is not affected by whether the weights are large or small.

Expand

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:



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

- [x] $b^{[2]}$ will have shape (3, 1)

! **This should not be selected**

No. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $W^{[1]}$ will have shape (4, 3).

! **This should not be selected**

No. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

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

✓ **Correct**

- [x] $b^{[1]}$ will have shape (3, 1).

✓ **Correct**

Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $W^{[1]}$ will have shape (3, 4).

✓ **Correct**

Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [x] $b^{[1]}$ will have shape (1, 3)

! **This should not be selected**

No. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False?

1 / 1 point

- True
- False

Expand

Correct
Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

3. Which of the following is a correct vectorized implementation of forward propagation for layer 2?

1 / 1 point

- $Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$
  $A^{[2]} = g(Z^{[2]})$
- $Z^{[1]} = W^{[1]} X + b^{[1]}$
  $A^{[1]} = g^{[1]}(Z^{[1]})$
- $Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$
  $A^{[2]} = g^{[2]}(Z^{[2]})$
- $Z^{[2]} = W^{[2]} X + b^{[2]}$
  $A^{[2]} = g^{[2]}(Z^{[2]})$

Expand

Correct
Yes. The elements of layer two are represented using a superscript in brackets.

4. When building a binary classifier for recognizing cats (y=1) vs raccoons (y=0). Is better to use the sigmoid function as activation function for the hidden layers. True/False

1 / 1 point

- True
- False

Expand

Correct
Yes. Using tanh almost always works better than the sigmoid function for hidden layers.

5. Consider the following code:

1 / 1 poi

A = np.random.randn(4,3)

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

Expand

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

6. Suppose you have built a neural network with one hidden layer and tanh as activation function for the hidden layer. You decide to initialize the weights to small random numbers and the biases to zero. The first hidden layer's neurons will perform different computations from each other even in the first iteration. True/False?

1 / 1 point

- ( ) False No. Since the weights are most likely different, each neuron will do a different computation.
- (•) True Yes. Since the weights are most likely different, each neuron will do a different computation.

Expand

Correct

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

- ○ False
- ◉ True

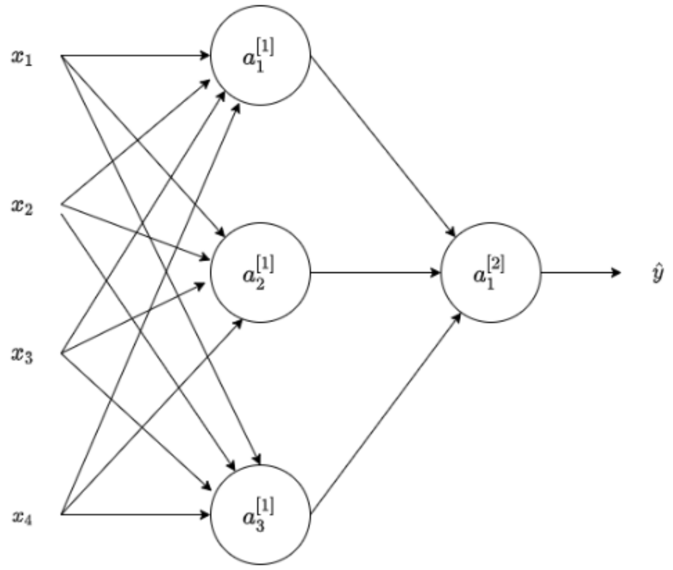
Expand

Correct

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

9. Consider the following 1 hidden layer neural network:

1 / 1 point



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

- [x] $b^{[1]}$ will have shape (3, 1).

✓ Correct
Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $W^{[1]}$ will have shape (3, 4).

✓ Correct
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [ ] $W^{[1]}$ will have shape (4, 3).
- [ ] $b^{[2]}$ will have shape (3, 1)
- [x] $b^{[2]}$ will have shape (1,1)

✓ Correct

2. In which of the following cases is the linear (identity) activation function most likely used?

1 / 1 point

- (selected) When working with regression problems.
- The linear activation function is never used.
- As activation function in the hidden layers.
- For binary classification problems.

Expand

Correct
Yes. In problems such as predicting the price of a house it makes sense to use the linear activation function as output.

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

1 / 1 point

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

Expand

Correct

4. The use of the ReLU activation function is becoming more rare because the ReLU function has no derivative for $c = 0$. True/False?

1 / 1 point

- (selected) False
- True

Expand

Correct
Yes. Although the ReLU function has no derivative at $c = 0$ this rarely causes any problems in practice. Moreover it has become the default activation function in many cases, as explained in the lectures.

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

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. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

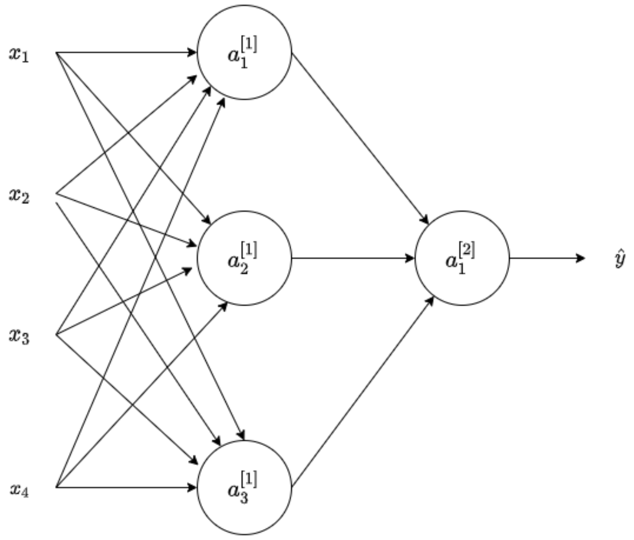
1 / 1 point

- ◉ True
- ○ False

Expand

Correct
Yes. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

9. Consider the following 1 hidden layer neural network:



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

- [x] $b^{[1]}$ will have shape (1, 3)

! **This should not be selected**

No. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $W^{[1]}$ will have shape (3, 4).

✓ **Correct**

Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [x] $b^{[2]}$ will have shape (3, 1)

! **This should not be selected**

No. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

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

✓ **Correct**

Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $W^{[1]}$ will have shape (4, 3).

! **This should not be selected**

No. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

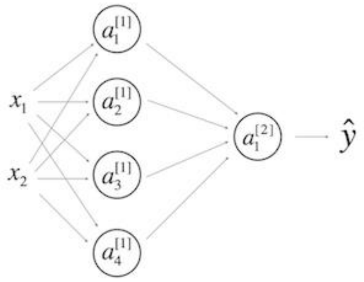
- [x] $b^{[1]}$ will have shape (3, 1).

✓ **Correct**

Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

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

1 / 1 point



- ○ $Z^{[1]}$ and $A^{[1]}$ are (1,4)
- ○ $Z^{[1]}$ and $A^{[1]}$ are (4,2)
- ◉ $Z^{[1]}$ and $A^{[1]}$ are (4,m)
- ○ $Z^{[1]}$ and $A^{[1]}$ are (4,1)

Expand

Correct

Final Attempt:

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False?

1 / 1 point

- ○ True
- ◉ False

Expand

Correct
Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

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

1 / 1 point

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

Expand

Correct

4. The use of the ReLU activation function is becoming more rare because the ReLU function has no derivative for $c = 0$. True/False?

1 / 1 point

- ○ True
- ◉ False

Expand

Correct
Yes. Although the ReLU function has no derivative at $c = 0$ this rarely causes any problems in practice. Moreover it has become the default activation function in many cases, as explained in the lectures.

5. Consider the following code:

1 / 1 point

A = np.random.randn(4,3)

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).

- (x) (4, 1)
- ( ) (1, 3)
- ( ) (4,)
- ( ) (3,)

Expand

Correct

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

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 compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in the 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.
- (x) 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".

Expand

Correct

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

- ○ False
- ◉ True

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 are true about the tanh function?

1 / 1 point

- [ ] The derivative at $c = 0$ is not well defined.
- [x] The tanh is mathematically a shifted version of the sigmoid function.

  Correct

  Yes. You can see the shape of both is very similar but tanh passes through the origin.
- [ ] The slope is zero for negative values.
- [x] For large values the slope is close to zero.

  Correct

  Yes. We can see in the graph of the $y = tanh(c)$ how as the values of c increase the curve becomes flatter.
- [ ] For large values the slope is larger.

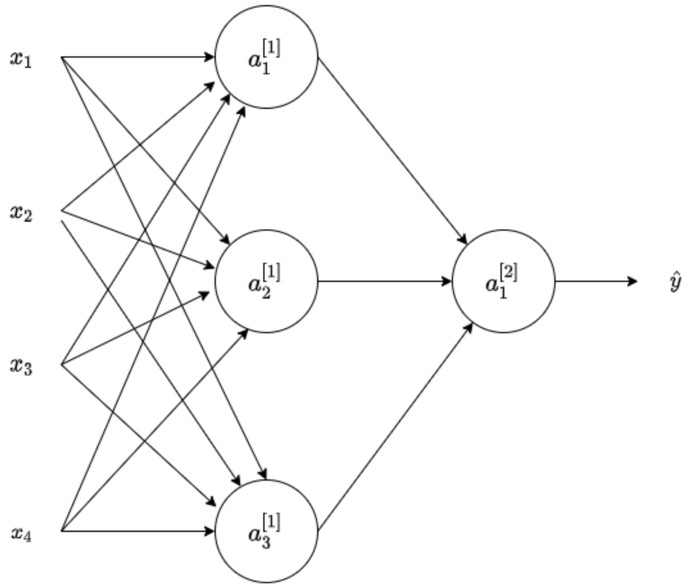
Expand

Correct

Great, you got all the right answers.

9. Consider the following 1 hidden layer neural network:

1 / 1 point



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

- [x] $b^{[1]}$ will have shape (3, 1).

✓ **Correct**
Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [ ] $W^{[1]}$ will have shape (4, 3).
- [ ] $b^{[2]}$ will have shape (3, 1)
- [x] $W^{[1]}$ will have shape (3, 4).

✓ **Correct**
Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the

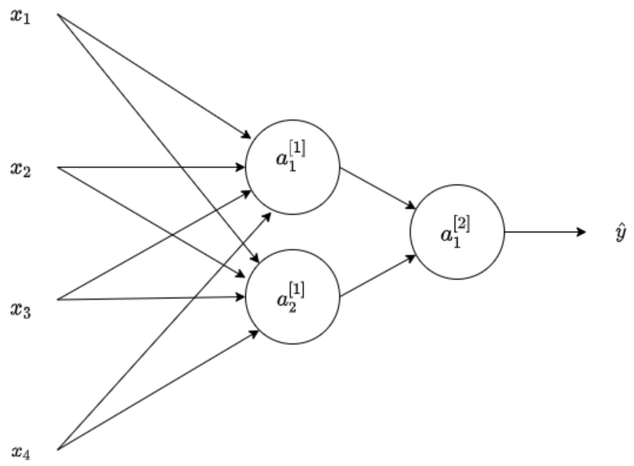
- [x] $b^{[2]}$ will have shape (1,1)

✓ **Correct**
Yes. $b^{[k]}$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [ ] $b^{[1]}$ will have shape (1, 3)

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 (2, m)
- $Z^{[1]}$ and $A^{[1]}$ are (2, 1)
- $Z^{[1]}$ 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 $Z^{[1]}$