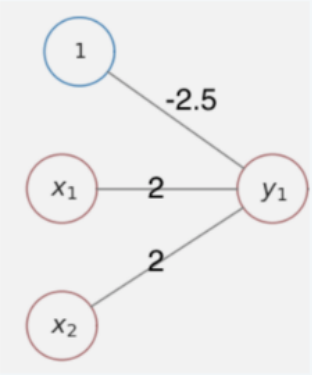
Week 1 Lesson 1 Quiz

1. Which of the following logical operations does the following perceptron represent? Note that x_1 and x_2 are binary.

3 / 3 pc



- ○ x_1 NAND x_2
- ◉ x_1 AND x_2
- ○ x_1 XOR x_2
- ○ x_1 OR x_2

Correct

Your answer is correct. The correct answer is x_1 AND x_2.

Consider Logical AND. Logical AND outputs 1 only when both inputs x_1 and x_2 are 1. For every other case, AND should output 0. The weights are the same for

2. True or False: The following training set can be classified exactly by a single perceptron.

x_1	0	1	0	1
x_2	0	0	1	1
y	0	0	1	1

- (●) True
- ( ) False

Correct
Your answer is correct. A single perceptron is a linear classifier. A linear classifier can properly separate these data points.

3. Which of the following can be guaranteed to behave as Activation functions and are not difficult to train? Select one or more:

5 / 6

- [ ] Step function
- [x] ReLu

Correct
ReLu is guaranteed to behave as an Activation function and is not difficult to train.

- [ ] sine
- [x] Sigmoid

Correct
Sigmoid is guaranteed to behave as an Activation function and is not difficult to train.

- [ ] cosine
- [x] Tanh

Correct
Tanh is guaranteed to behave as an Activation function and is not difficult to train.

You didn't select all the correct answers

4. Consider your model is being trained using the Perceptron algorithm. Let W be the current Weight and x be a misclassified instance. Which of the following statements are valid? 3 / 4 poi

- [ ] If x is a positive instance classified as negative, then $W = W - x$
- [x] If x is a positive instance classified as negative, then $W = W + x$

Correct
Your answer is correct. If x is a positive instance misclassified as negative, then adjust the weight with $W = W + x$.

- [ ] If x is a negative instance classified as positive, then $W = W - x$
- [ ] If x is a negative instance classified as positive, then $W = W + x$

You didn't select all the correct answers

5. What are the possible **hyperparameters** that can be tuned for a Multi-Layered Perceptron (MLP)? Select one or more: 4 / 4 poi

- [x] Activation functions

Correct
Activation functions are a hyperparameter that can be tuned for MLP.

- [x] Number of nodes in a layer

Correct
Number of nodes in a layer are a hyperparameter that can be tuned for MLP.

- [ ] Weights
- [x] Number of hidden layers

Correct
Number of hidden layers are a hyperparameter that can be tuned for MLP.

Quiz

Week 1 Lesson 2 Quiz

1. Which of the following statements is NOT true? 7 / 7 poin

- ( ) Back propagation is more computationally complex than forward propagation.
- ( ) Vanishing gradients can occur and lead to slow convergence.
- ( ) Forward propagation is necessary for computing the activations at each layer.
- (•) Back propagation can always find the global optimum regardless of weights initialization.

Correct
Your answer is correct. Regardless of weight initialization, backpropagation cannot always find the global optimum.

What are the output values in the hidden layer output vector *h*? Give each answer to three decimal places.

What is the first element in the hidden layer output vector (h1) ?

0.500

Correct

Your answer is correct.

Good notation examples for written answers:

$W_{1,2}^{X}$ for the weight between X_1 and h_2

W_1^h for the weight between h_1 and a

$$W^h = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

What are the output values in the hidden layer output vector *h*? Give each answer to three decimal places.

What is the second element in the hidden layer output vector (h2)?

Highlight

0.268

Correct

Your answer is correct.

Good notation examples for written answers:

$W_{1,2}^{X}$ for the weight between X_1 and h_2

W_1^h for the weight between h_1 and a

Good notation examples for coding answers:

`Wx` for the W^X matrix

`Wh` for the W^h matrix

The weight matrix for h-a is

$$W^h = \begin{pmatrix} -1 \\ 1/2 \\ 1/2 \end{pmatrix}$$

What are the output values in the hidden layer output vector h? Give each answer to three decimal places.

What is the third element in the hidden layer output vector (h3)?

0.731

Correct
Your answer is correct.

Good notation examples for written answers:

$W_{1,2}^{X}$ for the weight between X_1 and h_2

W_1^h for the weight between h_1 and a

Good notation examples for coding answers:

`Wx` for the W^X matrix

`Wh` for the W^h matrix

What is the output value at a? Give your answer to three decimal places.

0.500

Correct
Your answer is correct.

Good notation examples for written answers:

W^X for the weight between X_1 and h_2

Highlight

W_1^h for the weight between h_1 and a

Good notation examples for coding answers:

`Wx` for the W^X matrix

`Wh` for the W^h matrix

`x` for input vector

`h` for hidden layer's output vector

computation graph is not used at all, it will not score properly. Try to fit the red boxes above. This question does not need coding and the answer can be easily obtained analytically.

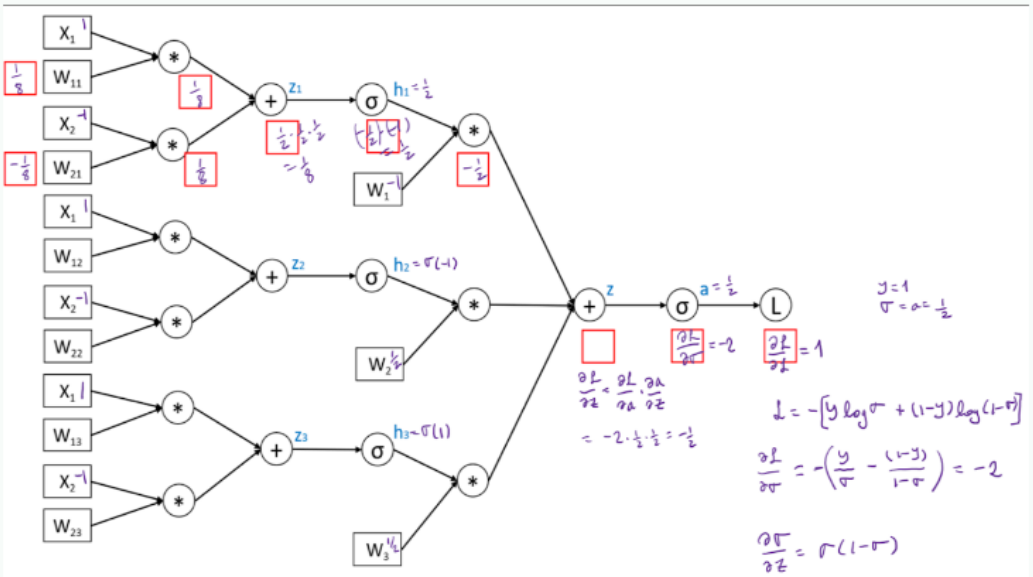
Highlight

Hint: You may use the property of $\frac{\partial \sigma(z)}{\partial z} = \sigma(1-\sigma)$

0.125

Correct

Your answer is correct.



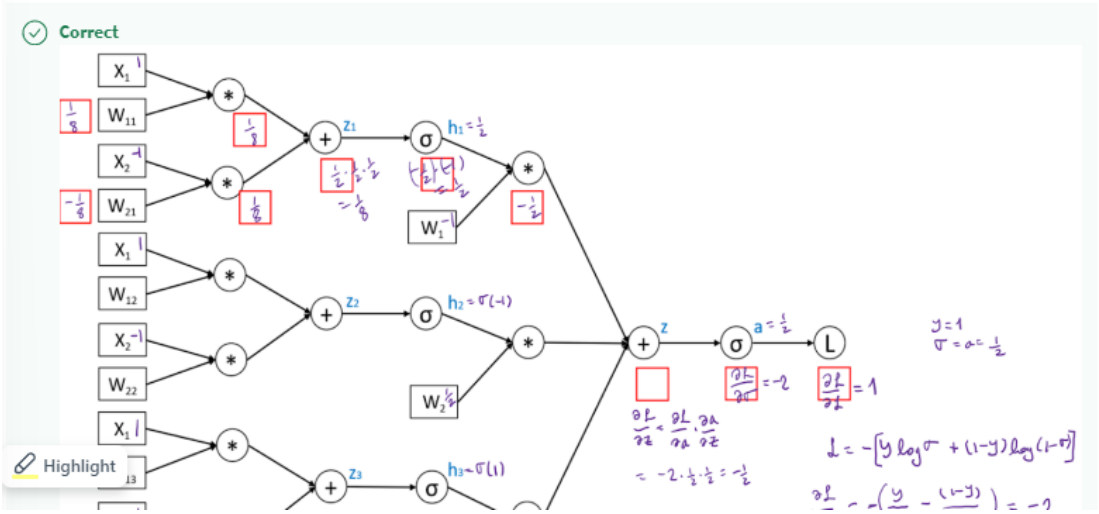
Hint: You may use the property of $\frac{\partial \sigma(z)}{\partial z} = \sigma(1-\sigma)$

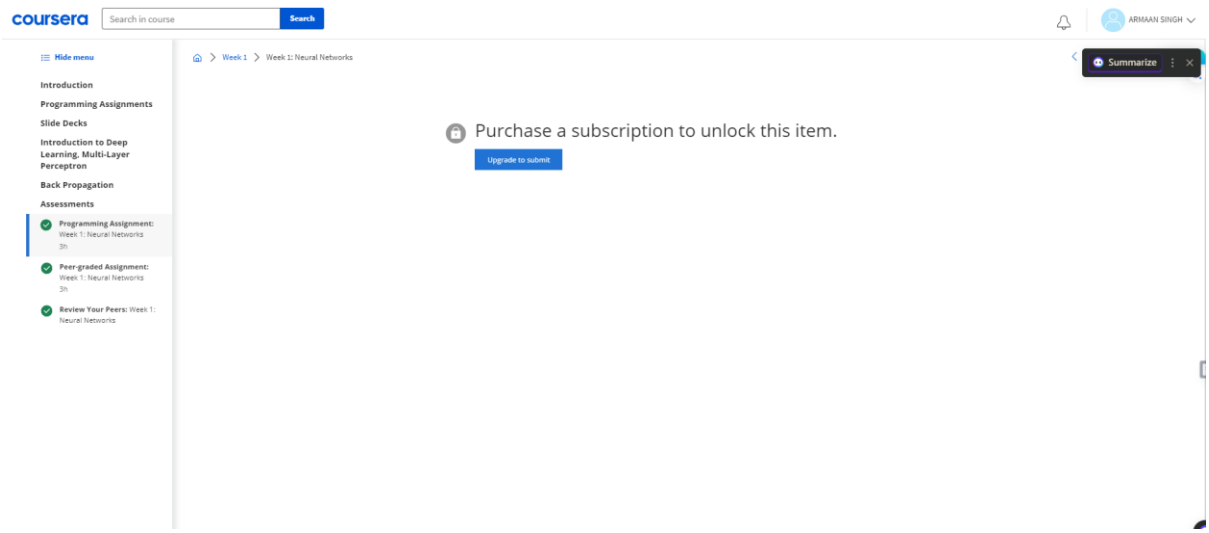
Calculate new weight using the old weight and learning learning as follows:

$$W21 \leftarrow W21 - \eta \frac{\partial L}{\partial W21}$$

0.558

Correct





Week 2 Quiz

1. Which of the following statements is **NOT** a reason for adding a dropout layer? **3 / 3 points**

- ○ Adding a dropout layer forces the network to have a redundant representation.
- ◉ Adding a dropout layer makes the network more expressive during training time.
- ○ Adding a dropout layer trains an ensemble of models.
- ○ Adding a dropout layer functions as regularization.

Correct

Your answer is correct. In the article *Surprising properties of dropout in deep networks,* Helmbold and Long (2018, p. 15) write, "dropout training can hijack part of the expressiveness of the wide layer to control the artificial variance due to dropout rather than fitting the underlying patterns in the data."

Helmbold, D. P., & Long, P. M. (2018). Surprising properties of dropout in deep networks. *Journal of Machine Learning Research*, *18*, 1–28. https://doi.org/https://www.jmlr.org/papers/volume18/16-549/16-549.pdf

```
Epoch 30/100
1500/1500 [==============================] - 5s 3ms/step - loss: nan - accuracy: 0.0993 - val_loss: nan - val_accuracy: 0.1030
Epoch 31/100
1500/1500 [==============================] - 5s 3ms/step - loss: nan - accuracy: 0.0993 - val_loss: nan - val_accuracy: 0.1030
Epoch 32/100
1500/1500 [==============================] - 4s 3ms/step - loss: nan - accuracy: 0.0993 - val_loss: nan - val_accuracy: 0.1030
Epoch 33/100
1500/1500 [==============================] - 4s 3ms/step - loss: nan - accuracy: 0.0993 - val_loss: nan - val_accuracy: 0.1030
Epoch 34/100
```

What are possible recommendations that can improve the result? (Select all that apply).

- ☑ Early stopping

Correct

Your answer is correct. Early stopping could improve the results.

- ☐ Increase learning rate
- ☐ Larger architecture
- ☐ More epochs
- ☑ Add regularization terms such as dropout, batch normalization, ridge, and lasso etc.

Correct

Your answer is correct. You could add regularization terms such as dropout, batch normalization, ridge and lasso, etc. to improve the result.

```
1500/1500 [==============================] - 4s 3ms/step - loss: nan - accuracy: 0.0993 - val_loss: nan - val_accuracy: 0.1030
Epoch 34/100
```

What is the most important fix you recommend to your friend?

- ○ You can add regularization terms such as dropout, batch normalization, ridge, and lasso etc.
- ◉ Reduce learning rate
- ○ Try a smaller architecture
- ○ Early stopping

Correct

Your answer is correct. The training log shows that it strongly overfits after certain epochs, and the loss even diverges due to the very high learning rate. The obvious observation is that the learning rate is very high, so I recommend reducing the learning rate first and seeing how it goes. The good values for the learning rate can be obtained by trying multiple learning rate values and monitoring the train/validation loss or accuracy, and choosing the one that leads the best validation accuracy.

Purchase a subscription to unlock this item.

Upgrade to submit

Quiz

Week 3 Quiz

Review Learning Objectives

1. Consider a **convolutional neural network model** that has three convolution layers. The first layer has 50 filters, the second layer has 100 filters, and the third layer has 200 filters. All convolution layers have stride=2, and the same padding. The input images are 300x400 pixels with RGB channels. Assume filter size of 3x3. **4 / 4**

How many trainable parameters does the CNN model have? Enter integer answer .

226700

Correct

Parameters in 1st layer; 50 filters with 3x3x3 size + 1 bias = 50x(3x3x3+1) = 1400

Parameters in 2nd layer; 100 filters with (3x3x50 + 1) = 45100

Parameters in 3rd layer; 200 filters with (3x3x100+1) = 180200

Total parameters = 1400+45100+180200 = 226700

2. Consider a **convolutional neural network model** that has three convolution layers. The first layer has 50 filters, the second layer has 100 filters, and the third layer has 200 filters. All convolution layers have stride=2, and the same padding. The input images are 300x400 pixels with RGB channels. **4 / 4 point**

What is the **feature map size** after the third convolution layer?

- ○ 75x100x100
- ○ 150x200x50
- ◉ 38x50x200

Correct

After applying the first convolutional layer, the first feature map size is 150x200x50. After the second layer, the feature map size is 75x100x100. After the third layer, the feature map size is 38x50x200.(the width 75 gets padded to 77, then divided by two is 38).

3. While training a CNN model, you receive the **OOM (out of memory) error message.** What can you do to resolve the issue? Choose all that apply. 4 / 4 p

- [x] Reduce the batch size

  Correct
  Reducing the mini-batch size will require less memory per batch.

- [ ] Reduce the number of epochs
- [x] Reduce the number of filters in conv layers

  Correct
  Reducing the number of filters can reduce the feature map depth.

- [ ] Reduce the stride of conv layers
- [x] Reduce the number of layers

  Correct
  A smaller number of layers will give fewer parameters and feature maps.

4. **True or False**: You can use bigger strides in a convolutional layer instead of a max-pooling layer for better accuracy. 2 / 2 po

- (•) True
- ( ) False

  Correct
  A conv layer with stride two can subsample equivalently with max pool with a 2x2 filter size. It costs more parameters, but the conv layer is learnable, so it generally tends to have better accuracy.

5. How many conv layers with 3x3 filters would you need to have the same receptive field as a conv layer with 11x11 filters? Assume stride=1 and no padding. ***Enter your answer for the number of layers as an integer***, E.g., 1 2 / 2 po

5

Correct
The correct answer is 5 layers. 1->3->5->7->9->11