Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 70% or higher

Go to next item

1. Which of the following gates can't be represented by a perceptron? **1 / 1 point**

- ( ) NAND
- ( ) NOT
- ( ) None of the above can be represented.
- (•) All of the above can be represented.

Correct
As shown in the lecture, a perceptron can represent a NAND gate. Furthermore, if we feed the same value to both inputs of a NAND gate, we arrive at a NOT gate (as seen in the lecture). Therefore, both NAND and NOT gates can be represented by a Perceptron.

2. Is it possible to design a perceptron network to solve the logical expression below, given that the inputs x_1 and x_2 are binary? **1 / 1 point**

$$z = (\bar{x}_1 + x_2) \times (x_1 + x_2) + x_1$$

- (•) Yes
- ( ) No
- ( ) Cannot be decided based on the given information

Correct
A perceptron network can solve any logical expression, since any logical

gate can be represented as a perceptron network.

3. What is the difference between a perceptron and a sigmoid neuron? **1 / 1 point**

- (●) A perceptron outputs a binary decision whereas a sigmoid neuron's output is continuous.
- ( ) A perceptron is easier to train than a sigmoid neuron.
- ( ) A perceptron can take an unlimited number of inputs, whereas a sigmoid neuron cannot.
- ( ) None of the above

Correct
A perceptron uses a step function as its activation function and hence its output binary. A neuron uses a sigmoid function as its activation function and hence its output is continuous between 0 and 1.

Due to the continuous activation function of a Sigmoid Neuron, it is easier to train than a perceptron, as it is easier to determine how to change the input to get the desired output. Furthermore, both a perceptron and a sigmoid neuron can take an unlimited number of inputs.

4. Why is a sigmoid function a better activation function than a step function? **1 / 1 point**

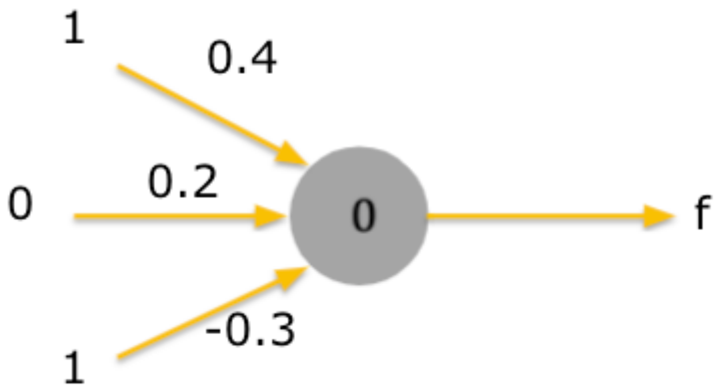
- ( ) A sigmoid function is not a better activation function than a step function
- (●) Using a sigmoid function, small changes in the weight cause small changes in the output, which facilitates training
- ( ) A sigmoid function can cause sudden changes in the output
- ( ) None of the above

Correct
Since a sigmoid function is a continuous monotonic function, small changes in its input cause small changes in the output. Due to the monotonic property, if we increase the input, it will cause an increase in the output and the converse is also true. This allows us to guide the training process. For example, if our initial output for a perceptron was

leaning towards 0 but we need it to be 1, we can make an informed decision about increasing the input to that perceptron.

5. The input, weights, and activation function of a perceptron are shown below. Compute its output.

2 / 2 points



$$f = \begin{cases} 0 & \text{if} \sum_j w_j x_j + b \leq 0 \\ 1 & \text{if} \sum_j w_j x_j + b > 0 \end{cases}$$

- ○ 0.1
- ◉ 1
- ○ 0
- ○ 0.3

Correct
Substituting the values for the weights and biases in the equation above, we get $\sum_j w_j x_j + b > 0$. Hence, the output is 1.

6. How many perceptron units are typically used in the output when classifying data into N classes?

1 / 1 point

- ○ 1

- [x] N
- [ ] $log\, N$
- [ ] $N/2$

Correct
Since each perceptron gives a binary output of 0 or 1, we need N perceptron units in the output layer to predict if a layer belongs to a class or not.

7. Which of the following is true about learning rates? **1 / 1 point**

- [ ] Higher the learning rate, the better
- [ ] Lower the learning rate, the better
- [ ] When a network is small, lower learning rate is better
- [x] None of the above

Correct
The optimal learning rate is dependent on the initialization and the cost function curves which are often not known. Thus, there is no fixed rule for learning rate. Optimal learning rate needs to be determined by experimentation.

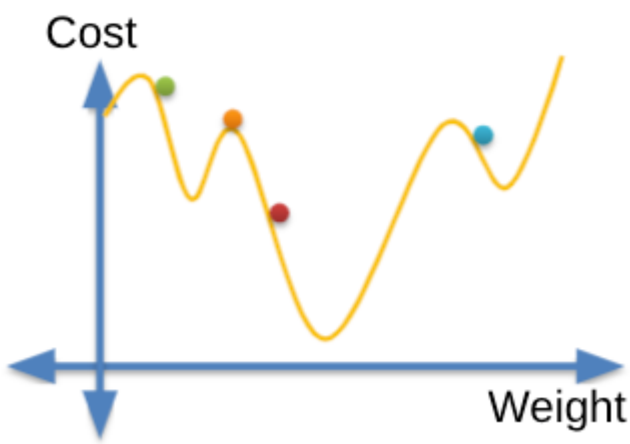
8. Which of the following is NOT equivalent to the derivative of a sigmoid function? **3 / 3 points**

- [ ] $\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$
- [ ] $\sigma'(z) = \frac{e^{z}}{e^{2z}+2e^{z}+1}$
- [x] $\sigma'(z) = \frac{1}{e^{-z}+2e^{-z}+1}$
- [ ] $\sigma'(z) = \sigma(z)(1-\sigma(z))$

Correct
Mathematically, all others are equivalent to $\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$.

9. Below is a plot of the cost function against the weight of a neural network. The colored dots represent different weight initializations for the network. Which of these is most likely to converge to the global minimum?

1 / 1 point



- ◯ Green
- ◯ Orange
- ◉ Red
- ◯ Blue

Correct
Except red, other initializations can converge to the local minimum.

10. Which of the following statements about Convolutional Neural Networks is false?

1 / 1 point

- ◯ In a Convolutional Neural Network, the convolution kernels are also learned.
- ◉ Convolutional Neural Network replaces all neural layers with convolutional layers.
- ◯ One convolutional layer can contain several distinct convolutions.
- ◯ All of the above

Correct
While a Convolutional Neural Network introduces a convolutional layer

(or layers), the classification necessitates the presence of a regular, fully connected neural network. The first answer choice is true, since the entire reason for using a Convolutional Neural Network is that the kernels, which tended to be hand-engineered, are now learned. The third answer choice is also true, since as shown in the lecture, one convolutional layer can contain many different kernels performing different convolutions.