CS244 Exercise 6

Task 1: Neural Network Architectures

A perceptron is similar to a neural network that has no hidden layers and a single output unit with which of the following activation functions?

$$g(z) = \frac{1}{1+e^{-z}}$$

$$g(z) = \max(0, z)$$

$$g(z) = \begin{cases} 1 \text{ if } z>0 \\ -1 \text{ if } z \leq 0 \end{cases}$$

$$g(z) = \tanh(z)$$

Suppose we have data with four features about movies that we would like to classify into one of three classes: romantic comedy, horror, documentary. Which of the following two neural network architectures is more powerful (where "more powerful" means it can do everything the other can do plus more)?

- a single neural network with 4 inputs, 15 units in a hidden layer, and 3 units in the output layer
- three separate neural networks, each with 4 inputs, 5 units in a hidden layer, and 1 unit in the output layer

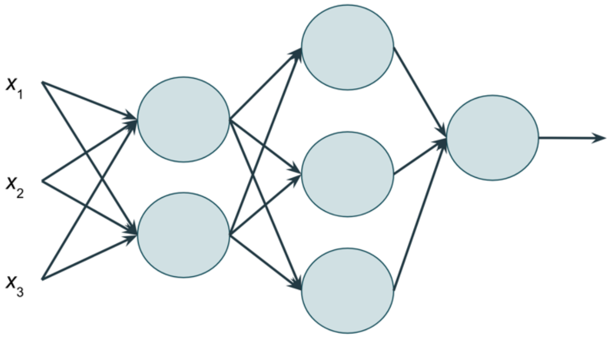
Which of the following is true about a neural network with 50 units in the input layer, no hidden layers, and 10 units in the output layer?

- it is similar to 10 different logistic regression classifiers, each with 50 inputs
- it is similar to 10 different logistic regression classifiers, each with 5 inputs
- it can learn non-linear relationships of the features
- it is equivalent to a decision tree with a height of 50 since it has no hidden layers
- all of the above

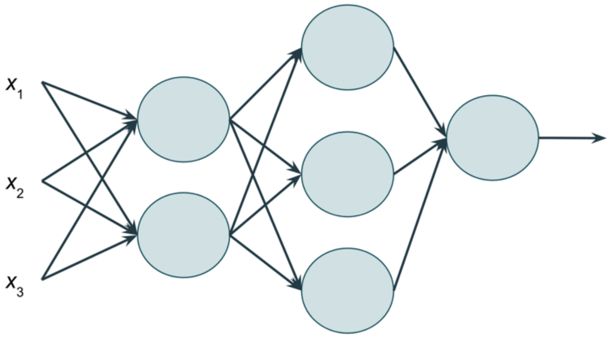
Suppose you have a neural network where, for one of the hidden layers, all units in the layer have identical values for their weight parameters as the other units in the layer. Which of the following is true?

- when new training data are provided to the neural network and training progresses, the units in the layer will continue to all have identical values for their weight parameters
- removing all units except one from the layer would result in an equivalent classifier
- all units in the layer will have the same input and output as the other units in the layer
- since the neural network has one or more hidden layers, it is possible that it may be able to model any continuous function, e.g., log(x), 1/x, sin(x), e^x, $\sqrt{x}$
- all of the above

Consider the neural network below. *Ignoring* any weight parameters associated with bias terms, how many values for weight parameters would need to be learned for this neural network?



Consider the neural network below. *Including* weight parameters associated with bias terms, how many values for weight parameters would need to be learned for this neural network?

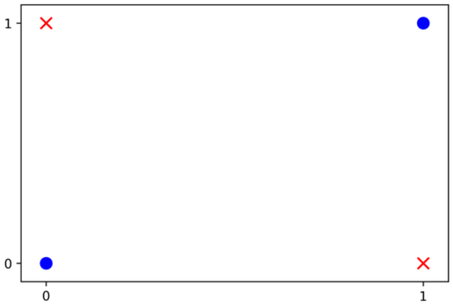


Task 2: Neural Networks for Boolean Logic Gate Functions

Suppose we have binary inputs that only take on values of 0 or 1. Below are truth tables and plots for the Boolean logic gate functions **AND** and **XOR**.

x_1	x_2	**AND**
0	0	0
0	1	0
1	0	0
1	1	1

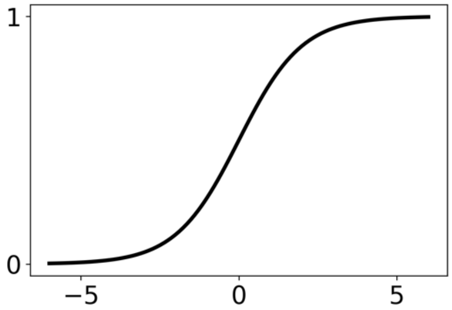
x_1	x_2	**XOR**
0	0	0
0	1	1
1	0	1
1	1	0



Notice that **AND** appears linearly separable (you could draw a line through the figure separating the positive and negative examples) whereas **XOR** does not. Thus, a simple neural network to model the **AND** function might not have a hidden layer whereas a simple neural network to model the **XOR** function might have a hidden layer. Which of the following Boolean logic gate functions are linearly separable?

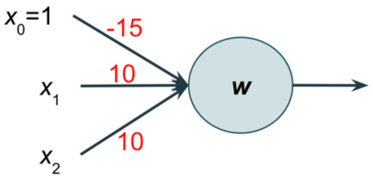
- NAND
- OR
- NOR
- XNOR

Recall the sigmoid function, as shown below, that may be used as the activation function in logistic regression and in neural networks.



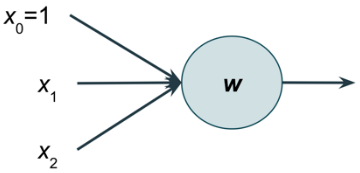
As a point of reference, given an input value less than -5, the sigmoid function returns a value of approximately 0 (less than 0.01). Similarly, given an input value greater than 5, the sigmoid function returns a value of approximately 1 (greater than 0.99).

Consider the neural network below that has three input units: x_0 is the bias term and is always 1, x_1 can take on a value of 0 or 1, and x_2 can take on a value of 0 or 1. The single unit in the output layer uses a sigmoid activation function.



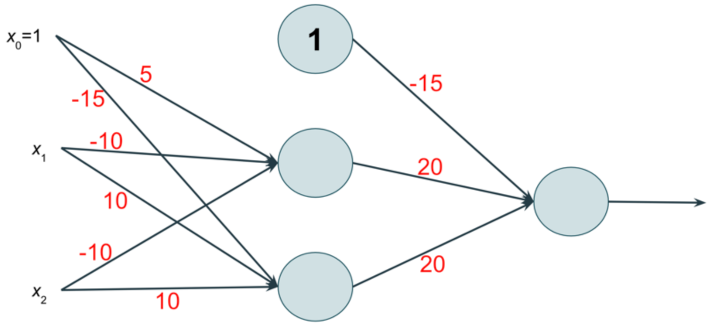
For the single unit in the output layer, if its three weight parameters are set to $\boldsymbol{w}$ = (-15, 10, 10) then the neural network models the **AND** function. Why? If either x_1 or x_2 is 0 then the input to the sigmoid activation function will be -5 or -15 and the output will be approximately 0. If both x_1 and x_2 are 1 then the input to the sigmoid activation function will be 5 and the output will be approximately 1.

Consider the neural network below that has three input units: x_0 is the bias term and is always 1, x_1 can take on a value of 0 or 1, and x_2 can take on a value of 0 or 1. The single unit in the output layer uses a sigmoid activation function.



What are values for the three weight parameters such that the neural network would model the **OR** function?

Consider the neural network below that has three input units: x_0 is the bias term and is always 1, x_1 can take on a value of 0 or 1, and x_2 can take on a value of 0 or 1. There is one hidden layer with three units, one of which is a bias unit that always outputs 1 and the other two of which use sigmoid activation functions. There is one unit in the output layer that uses a sigmoid activation function.



Which of the following Boolean logic gate functions does this neural network model?

- NAND
- NOR
- XOR
- XNOR

Download the Jupyter Notebook for Exercise 6 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

Task 3: Forward Propagation

What five values are output by the neural network, i.e., what value is output for each of the *n*=5 data points?

Using *tanh* as the activation function, what five values are output by the neural network?

Task 4: Image Classification

How many images are there in the dataset? How many features does each image have?

What is the accuracy of your neural network on the *training* data? What is the accuracy of your neural network on the *testing* data?

Task 5: Recycling

What is the accuracy of your neural network on the *training* data? What is the accuracy of your neural network on the *testing* data?