CS344 Exercise 4

Task 1: Neural networks

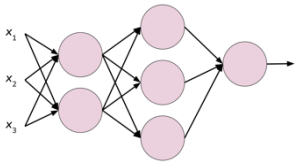
Neural networks can learn complex (non-linear) relationships between observed data and their labels

TRUE FALSE

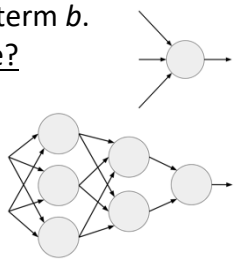
Units are just like neurons and artificial neural networks learn exactly the same way that the brain learns

TRUE FALSE

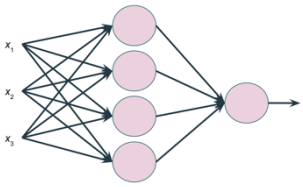
How many hidden layers does the neural network on the right have?



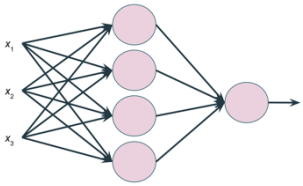
The unit on the right has 4 parameters: 3 for the weights ***w*** and 1 for the bias term *b*.
How many total parameters does the 6 node neural network on the right have?



Suppose the weights (not the biases) for the first (hidden) layer of the neural network on the right are stored in an array ***W***1. What are the dimensions of ***W***1?



If a batch of 32 data points are input to the neural network on the right, what are the dimensions of the array output by forward propagation?



Suppose a 2-layer neural network uses a sigmoid activation function for all units and has the following parameters:

$$\boldsymbol{W}1 = \begin{pmatrix} 0.5 & 1 & -4 \\ 3 & 6 & -1 \end{pmatrix} \qquad b1 = (2 \quad -3 \quad 1) \qquad \boldsymbol{W}2 = \begin{pmatrix} -6 \\ 1.1 \\ 5 \end{pmatrix} \qquad b2 = (6)$$

For a data point ***x*** = (5, 2), what value $\hat{y}$ (to 2 decimal places) would be predicted by forward propagation?

A. 0.01 B. 0.25 C. 0.5 D. 0.75 E. 0.99

In a binary classification problem, for a data point in class **1**, suppose a neural network predicts $\hat{y}$ = 0.6. What is the *loss* associated with the prediction for this point?

In a binary classification problem, for three data points with class labels y_1 = 1, y_2 = 0, and y_3 = 0, suppose a neural network makes predictions $\hat{y}_1$ = 0.8, $\hat{y}_2$ = 0.5, and $\hat{y}_3$ = 0.1.
What is the *cost* associated with the predictions for these three points?

While the cost function for logistic regression is concave, the cost function for neural networks normally contains multiple local minima.

TRUE FALSE

During backpropagation, for each layer of a neural network from the end to the beginning, gradients associated with the layer are calculated and the parameters for the layer are updated.

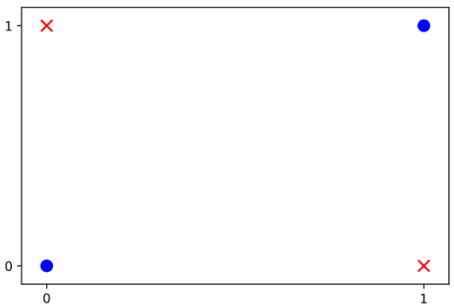
TRUE FALSE

Task 2: 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

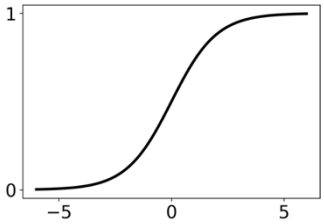


Notice that **AND** appears linearly separable (you could draw a line through the figure separating the positive and negative examples) whereas **XOR** does not. Thus, a logistic regression classifier could model the **AND** function well but not the **XOR** function. To model the **XOR** function, we could use a neural network with a hidden layer.

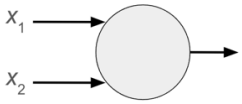
Which of the following Boolean logic gate functions are linearly separable (circle all that apply)?

NAND OR NOR XNOR

Recall the sigmoid function, as shown on the right. 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).



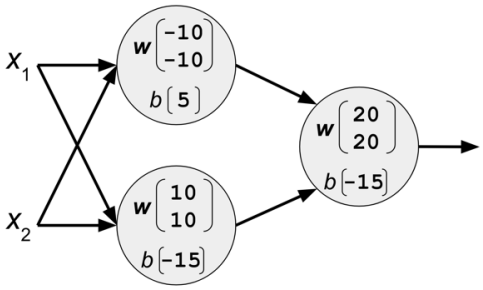
Consider the (simple) neural network below. The neural network has two binary inputs, x_1 and x_2, that can take on a value of either 0 or 1. The single unit in the output layer uses a sigmoid activation function.



If the unit's parameters are set to $\boldsymbol{w}$ = (10, 10) and b = (-15), 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.

Using this same neural network architecture (two binary inputs and a single unit in the output layer that uses a sigmoid activation function), what are parameter values for $\boldsymbol{w}$ and b such that the neural network would model the **OR** function?

Consider the neural network below. The neural network has two binary inputs, x_1 and x_2, that can take on a value of either 0 or 1. There are two units in the hidden layer and a single unit in the output layer. All units use the sigmoid activation function.



With the parameter values indicated in the image above, which one of the following Boolean logic gate functions does this neural network model?

NAND NOR XOR XNOR

Task 3: Practice with neural networks

Download the Jupyter Notebook for Exercise 4 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.

After training a NN with one hidden layer consisting of 5 units, what is the accuracy of the NN on the validation data?

After plotting the decision boundary, does the decision boundary appear approximately linear?

What is the shape of $\boldsymbol{W}1$? Why does $\boldsymbol{W}1$ have this shape, i.e., what dictates the number of rows of $\boldsymbol{W}1$ and the number of columns of $\boldsymbol{W}1$?

For how many iterations does gradient descent execute before it converges?

Looking at the plot of costs, does it appear to be increasing, decreasing, or neither?

Among the different architectures that you tried, what is the best validation accuracy that you observed?

What is your model's accuracy on the testing data?

On data from the file **`W.csv`**, what is the testing accuracy of the model whose architecture performed best on the validation data?

CS344 Exercise 4 Final Page

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME
Exercise	