HOML Ch.10 Exercise #6

Suppose you have an MLP composed of one input layer with 10 passthrough neurons, followed by one hidden layer with 50 artificial neurons, and finally one output layer with 3 artificial neurons. All artificial neurons use the ReLU activation function. (*Answers in italics*.)

Before we answer the questions, here's a quick summary of what the author stated about the equation relating to the output of fully connected layers $h_{\mathbf{W}, \mathbf{b}}(\mathbf{X}) = \phi(\mathbf{XW} + \mathbf{b})$:

> "**X** represents the matrix of input features. It has one row per instance and one column per feature.
>
> • The weight matrix **W** contains all the connection weights except for the ones from the bias neuron. It has one row per input neuron and one column per artificial neuron in the layer.
>
> • The bias vector **b** contains all the connection weights between the bias neuron and the artificial neurons. It has one bias term per artificial neuron.
>
> • The function ϕ is called the activation function: when the artificial neurons are TLUs, it is a step function."

Using this information, let's answer the following questions:

• What is the shape of the input matrix **X**?
*Since **X** has 10 passthrough neurons, its shape is m by 10 where m is the batch size.*

• What are the shapes of the hidden layer's weight vector $\mathbf{W}_h$ and its bias vector $\mathbf{b}_h$?
__X__ has 10 passthrough neurons and the hidden layer that follows, $\mathbf{W}_h$ has 50 artificial neurons, so the shape of W is 10 by 50. The bias vector $\mathbf{b}_h$ has a shape of 50 because there have to be as many bias terms as there are artificial neurons.

• What are the shapes of the output layer's weight vector $\mathbf{W}_o$ and its bias vector $\mathbf{b}_o$?
The output layer $\boldsymbol{W}_o$ is receiving 50 neurons and outputting 3, so its shape is 50 by 3. And the bias vector $\boldsymbol{b}_o$ has 3 terms, corresponding to the number of output neurons.

• What is the shape of the network's output matrix **Y**?
Its shape should be m by 3, where m is the batch size and 3 is the number of outputs.

• Write the equation that computes the network's output matrix **Y** as a function of **X**, $\mathbf{W}_h$, $\mathbf{b}_h$, $\mathbf{W}_o$, and $\mathbf{b}_o$.
We have an input layer and two ReLU activation layers, each with a bias vector. So, following the structure of the output of fully connected layers, $\boldsymbol{Y} = ReLU(ReLU(\boldsymbol{X}\boldsymbol{W}_h + \boldsymbol{b}_h)\boldsymbol{W}_o + \boldsymbol{b}_o)$
**** Additional notes from the author: ReLU activations turn every negative number in the matrix to zero, and the bias vector is added to every row in the matrix (broadcasting).*