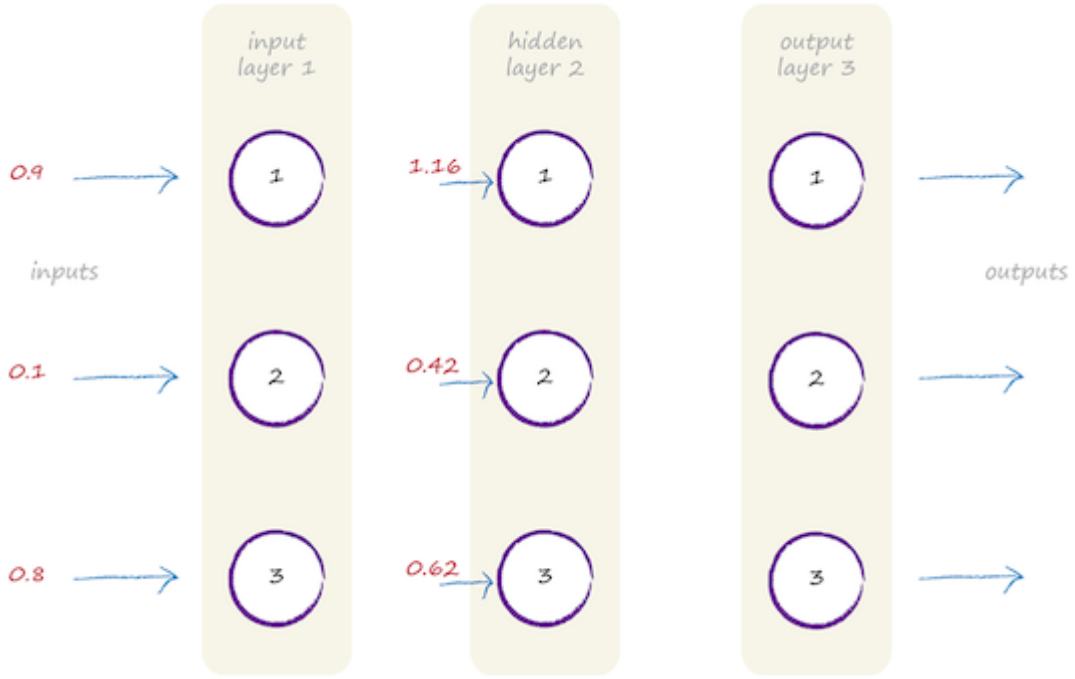
A Three Layer Example: Working on Hidden Layer

Moving forward, now we will calculate the weights for the hidden layer and calculate its output.

Let's visualize the combined moderated inputs which we calculated in the previous lesson, into the second hidden layer.



So far so good, but there's more to do. You'll remember those nodes apply a sigmoid activation function to make the response to the signal more like those found in nature. So let's do that:

$$O_{hidden} = sigmoid(X_{hidden})$$

The sigmoid function is applied to each element in X_{hidden} to produce the matrix which has the output of the middle hidden layer.

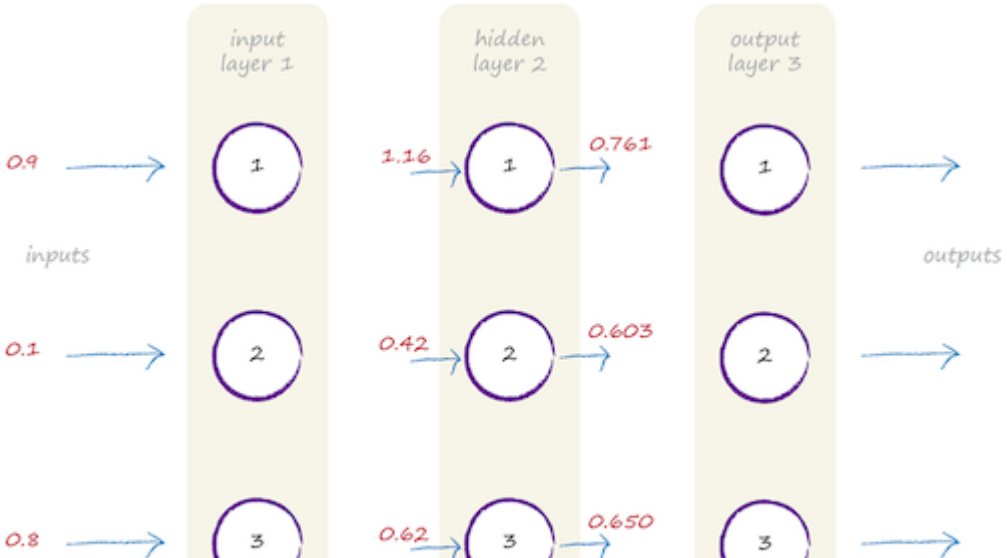
$$O_{hidden} = sigmoid \begin{pmatrix} 1.16 \\ 0.42 \\ 0.62 \end{pmatrix}$$

$$O_{hidden} = \begin{pmatrix} 0.761 \\ 0.603 \\ 0.650 \end{pmatrix}$$

Let's just check the first element to be sure. The sigmoid function is $y = \frac{1}{(1 + e^{-x})}$, so when x = 1.16, $e^{-1.16}$ is 0.3135. That means $y = \frac{1}{(1 + 0.3135)}$ = 0.761.

You can also see that all the values are between 0 and 1, because this sigmoid doesn't produce values outside that range. Look back at the graph of the logistic function to see this visually.

Phew! Let's pause again and see what we've done. We've worked out the signal as it passes through the middle layer. That is, the outputs from the middle layer. Which, just to be super clear, are the combined inputs into the middle layer which then have the activation function applied. Let's update the diagram with this new information.



If this were a two-layer neural network, we would stop now as these are the outputs from the second layer. We won't stop because we have another third layer.

How do we work out the signal through the third layer? It's the same approach as the second layer; there isn't any real difference. We still have incoming signals into the third layer, just as we did coming into the second layer. We still have links with weights to moderate those signals. And we still have an activation function to make the response behave like those we see in nature. So the thing to remember is, no matter how many layers we have, we can treat each layer like any other — with incoming signals which we combine, link weights to moderate those incoming signals, and an activation function to produce the output from that layer. We don't care whether we are working on the 3rd or 53rd or even the 103rd layer — the approach is the same.