Matrix Multiplication is Useful .. Honest!

This lesson will revise all the basic concepts of matrices and how they are multiplied.

Previously we manually did the calculations for a 2-layer network with just 2 nodes in each layer. That was enough work, but imagine doing the same for a network with 5 layers and 100 nodes in each? Just writing out all the necessary calculations would be a huge task … all those combinations of combining signals, multiplied by the right weights, applying the sigmoid activation function, for each node, for each layer … argh!

So how can matrices help? Well, they help us in two ways. First, they allow us to compress writing all those calculations into a very simple short form. That's great for us humans because we don't like doing a lot of work because it's boring, and we're prone to errors anyway. The second benefit is that many computer programming languages understand working with matrices, and because the real work is repetitive, they can recognize that and do it very quickly and efficiently.

In short, matrices allow us to express the work we need to do concisely and easily, and computers can get the calculations done quickly and efficiently. Now we know why we're going to look at matrices, despite perhaps a painful experience with them at school, let's make a start and demystify them. A *matrix* is just a table, a rectangular grid, of numbers. That's it. There's nothing much more complex about a matrix than that. If you've used spreadsheets, you're already comfortable with working with numbers arranged in a grid. Some would call it a table. We can call it a matrix too. The following shows a table of numbers.

A	B	C
3	32	5
5	74	2

That's all a matrix is — a table or a grid of numbers — just like the following example of a matrix of size "2 by 3".

$$\begin{pmatrix} 23 & 43 & 22 \\ 43 & 12 & 54 \end{pmatrix}$$

It is a convention to use rows first then columns, so this isn't a "3 by 2" matrix, it is a "2 by 3" matrix. Also, some people use square brackets around matrices, and others use round brackets as we have. Actually, they don't have to be numbers; they could be quantities which we give a name to, but may not have assigned an actual numerical value to. So the following is a matrix, where each element is a *variable* which has a meaning and could have a numerical value, but we just haven't said what it is yet. Now, matrices become useful to us when we look at how they are multiplied. You may remember how to do this from school, but if not we'll look at it again. Here's an example of two simple matrices multiplied together.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} (1*5)+(2*7) & (1*6)+(2*8) \\ (3*5)+(4*7) & (3*6)+(4*8) \end{pmatrix}$$

$$= \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

You can see that we don't simply multiply the corresponding elements. The top left of the answer isn't $1 * 5$, and the bottom right isn't $4 * 8$. Instead, matrices are multiplied using different rules. You may be able to work them out by looking at the above example. If not, have a look at the following which highlights how the top left element of the answer is worked out.

$$\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}5 & 6\\ 7 & 8\end{pmatrix} = \begin{pmatrix}(1*5)+(2*7) & (1*6)+(2*8)\\ (3*5)+(4*7) & (3*6)+(4*8)\end{pmatrix}$$

$$= \begin{pmatrix}19 & 22\\ 43 & 50\end{pmatrix}$$

You can see the top left element is worked out by following the top row of the first matrix, and the left column of the second matrix. As you follow these rows and columns, you multiply the elements you encounter and keep a running total. So to work out the top left element of the answer we start to move along the top row of the first array where we find the number 1, and as we start to move down the left column of the second matrix we find the number 5. We multiply 1 and 5 and keep the answer, which is 5, with us. We continue moving along the row and down the column to find the numbers 2 and 7. Multiplying 2 and 7 gives us 14, which we keep too. We've reached the end of the rows and columns so we add up all the numbers we kept aside, which is $5 + 14$, to give us 19. That's the top left element of the result matrix. Following illustration shows the rule that we follow while multiplying two matrices when the number of rows in the first martix matches the the number of columns in the second one:

$$\begin{pmatrix}a & b & ..\\ c & d & ..\end{pmatrix}\begin{pmatrix}e & f\\ g & h\\ .. & ..\end{pmatrix} = \begin{pmatrix}(a*e)+(b*g)+... & (a*f)+(b*h)+...\\ (c*e)+(d*g)+... & (c*f)+(d*h)+...\end{pmatrix}$$

$$= \begin{pmatrix}ae+bg+... & af+bh+...\\ ce+dg+... & cf+dh+...\end{pmatrix}$$

This is just another way of explaining the approach we take to multiplying matrices. By using letters, which could be any number, we've made even

clearer the generic approach to multiplying matrices. It's a generic approach because it could be applied to matrices of different sizes too. When we said it works for matrices of different sizes, there is an important limit. You can't just multiply any two matrices, they need to be compatible. You may have seen this already as you followed the rows across the first matrix, and the columns down the second. If the number of elements in the rows don't match the number of elements in the columns then the method doesn't work. So you can't multiply a "2 by 2" matrix by a "5 by 5" matrix. Try it — you'll see why it doesn't work. To multiply matrices the number of columns in the first must be equal to the number of rows in the second.