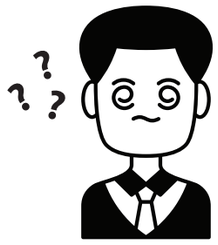
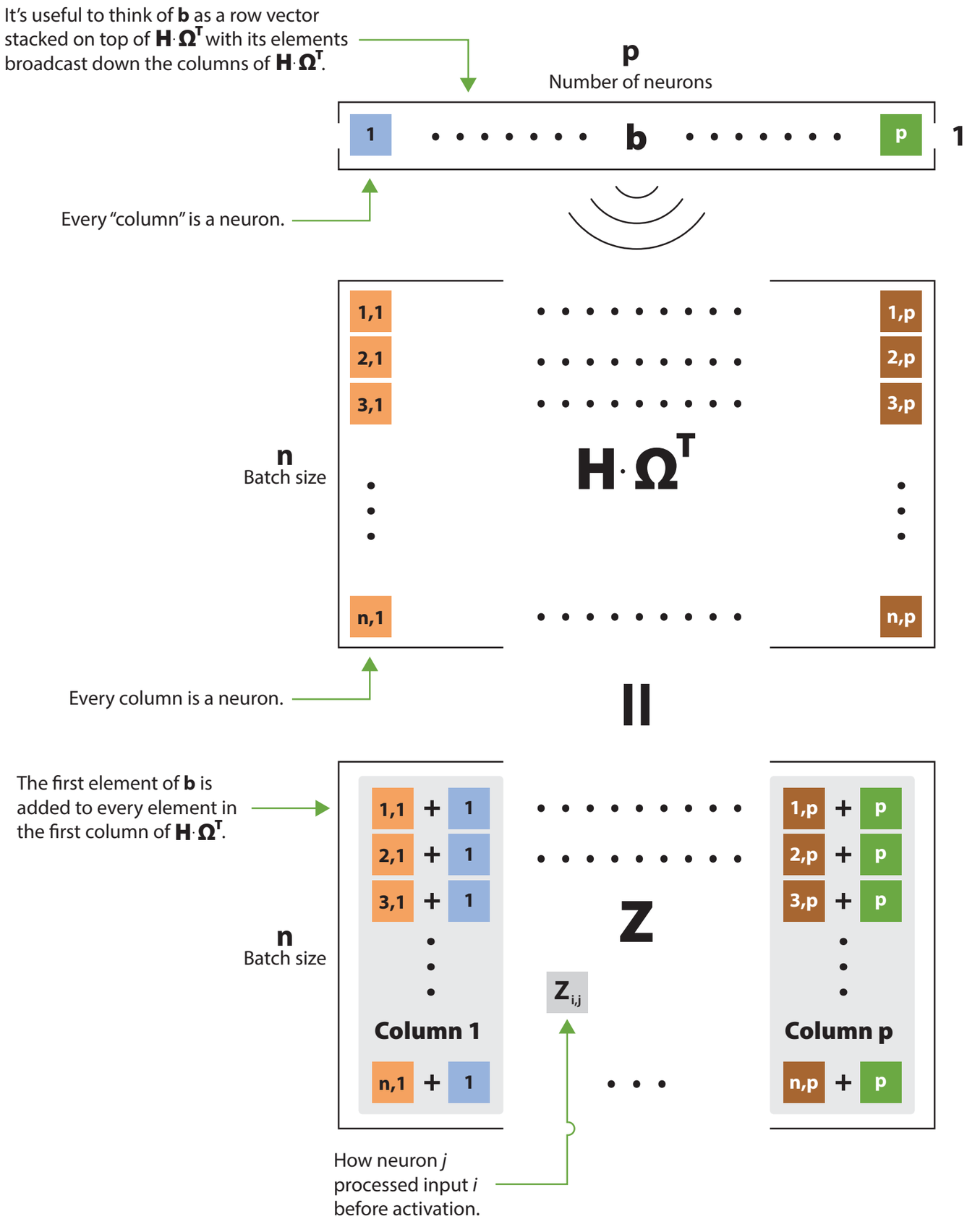


Neural Networks For People Who Get Confused Easily

by James McCammon (A fellow who gets confused easily)

Broadcasting

How "broadcasting" b works



Notes

1. What broadcasting is doing from a mathamtical perspective is taking the outer product of **b** and an implicitly defined, usually unshown vector of 1s, denoted as $\mathbf{1}_n$, where n is the same as our batch size. This outer product has the effect of creating a new matrix of the same size as $\mathbf{H}\cdot\mathbf{\Omega}^T$. We can then add the terms element wise.

The vector of 1s is constructed to be a column vector. We then transpose **b**. This gives us an (n x 1) vector of 1s times **b**, which is (1 x p). The result is a matrix that is (n x p), the same size as $\mathbf{H}\cdot\mathbf{\Omega}^T$. The outer product can be thought of using standard matrix multiplication, taking the vectors of 1s as an (n x 1) matrix and b as an (1 x p) matrix. Standard matrix mutliplication rules then apply. This new matrix contains the elements of **b** in each row, copied n times. See the example below for an example.

$$\mathbf{1}_4 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \otimes \mathbf{b}^T = \begin{bmatrix}1 & 2 & 3 & 4 & 5\end{bmatrix} = \begin{bmatrix}1\times1 & 1\times2 & 1\times3 & 1\times4 & 1\times5\\ 1\times1 & 1\times2 & 1\times3 & 1\times4 & 1\times5\\ 1\times1 & 1\times2 & 1\times3 & 1\times4 & 1\times5\\ 1\times1 & 1\times2 & 1\times3 & 1\times4 & 1\times5\end{bmatrix}$$