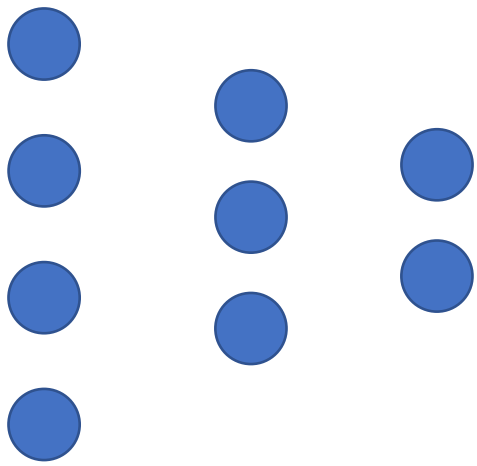
Feedforward



A 3 layers Neural network

Layer structure

The layers are L0,L1,L2
L0 is input layer(4 neural)
L1 is hidden layer(3 neural)
L2 is output layer(2 neural)

Weights and bias

The weights and biases for L0~L1 is matrix $W_{3\times 4}$ and vector $b_{3\times 1}$
The weights and biases for L1~L2 is matrix $W_{2\times 3}$ and vector $b_{2\times 1}$
If define the neural number of a specific layer l is n^l
Then the weights and biases from l to $l+1$ is:

$$W_{n^{l+1}\times n^l}$$
$$b_{n^{l+1}\times 1}$$

Backpropagation

<table>
<tr><td colspan="4">L
(Last layer)</td><td>C
(Cost function)</td></tr>
<tr><td>z^L</td><td colspan="2">δ^L</td><td>a^L</td><td>c</td></tr>
<tr><td colspan="2">$\mathrm{W}^L a^{L-1} + b^L$</td><td colspan="2">$\sigma(z^L)$</td><td>$f_c(a^L)$</td></tr>
</table>

Backpropagation gradient from cost function to last layer:

$$C = f_c(a^L)$$
$$\delta^L = (\nabla_{a^L} C) \odot \sigma^{'}(z^L)$$

<table>
<tr><td colspan="4">$l-1$
(previous layer)</td><td colspan="4">l
(some layer)</td></tr>
<tr><td>z^{l-1}</td><td colspan="2">δ^{l-1}</td><td>a^{l-1}</td><td>z^l</td><td colspan="2">δ^l</td><td>a^l</td></tr>
<tr><td colspan="2">$\mathrm{W}^{l-1} a^{l-2} + b^{l-1}$</td><td colspan="2">$\sigma(z^{l-1})$</td><td colspan="2">$\mathrm{W}^l a^{l-1} + b^l$</td><td colspan="2">$\sigma(z^l)$</td></tr>
</table>

Backpropagation gradient from some layer to previous layer:

$$\delta^{l-1} = ((W^l)^T \delta^l) \odot \sigma^{'}(z^{l-1})$$

Backpropagation to parameter:

$$\frac{\partial C}{\partial b^l} = \delta^l$$
$$\frac{\partial C}{\partial W^l} = a^{l-1}(\delta^l)^T$$