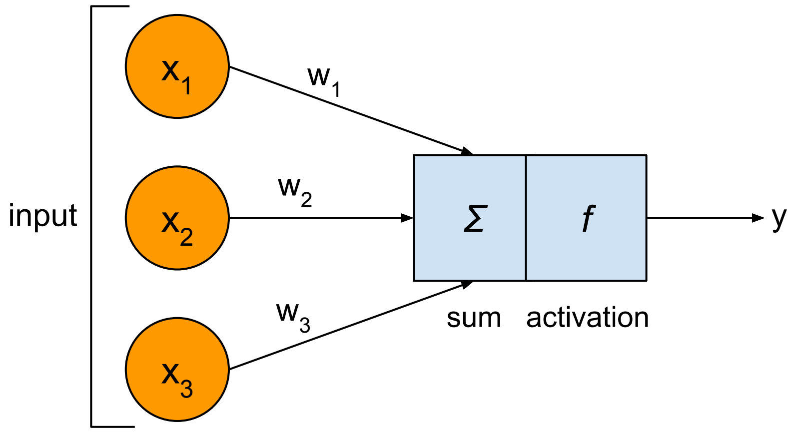
Computation in neural networks

Valerio Velardo

The artificial neuron



$$h = \sum_i x_i w_i = x_1 w_1 + x_2 w_2 + x_3 w_3$$

$$y = f(h) = f(x_1 w_1 + x_2 w_2 + x_3 w_3)$$

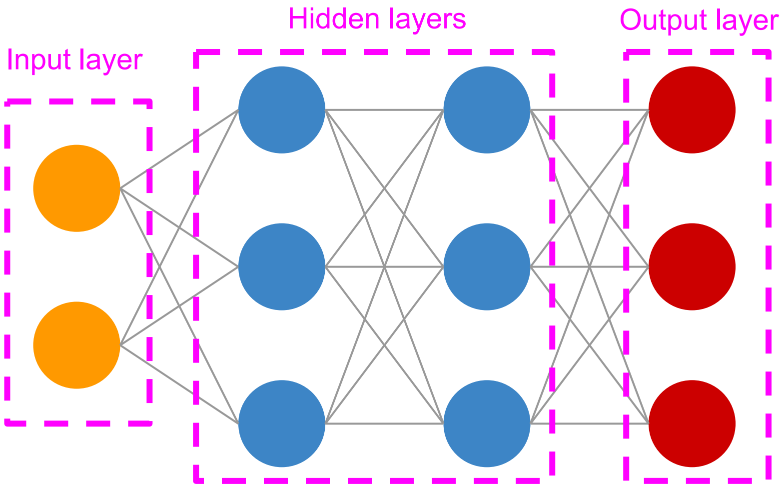
$$y = \frac{1}{1 + e^{-(x_1 w_1 + x_2 w_2 + x_3 w_3)}}$$

Why is a neural network needed?

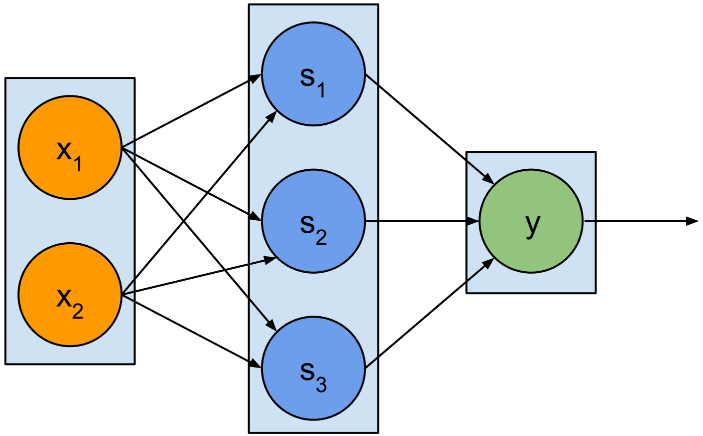
- A single neuron works for linear problems
- Real-world problems are complex
- ANNs can reproduce highly non-linear functions

The components of an artificial neural network (ANN)

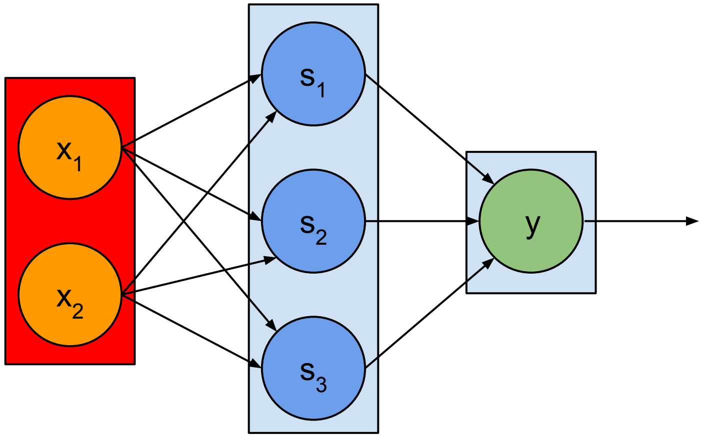
- Neurons
- Input, hidden, output layers
- Weighted connections
- Activation function



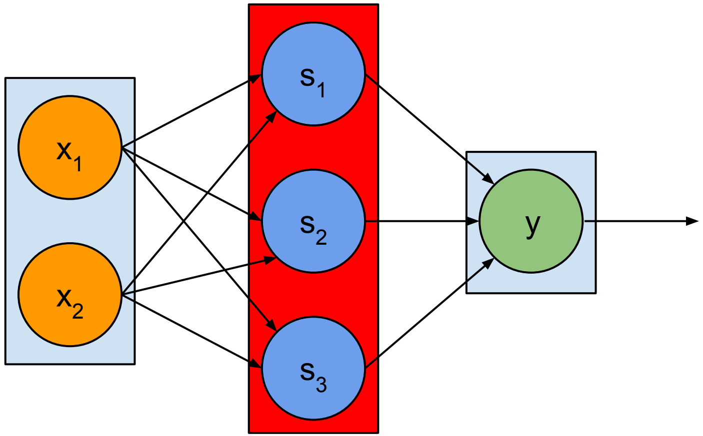
The multilayer perceptron (MLP)



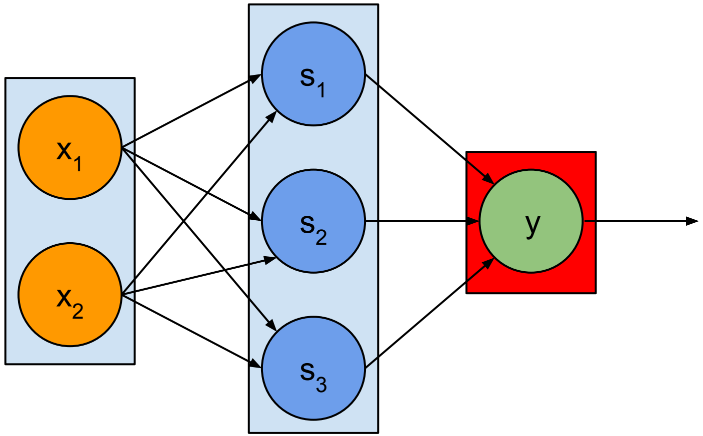
Computation in MLP



Computation in MLP



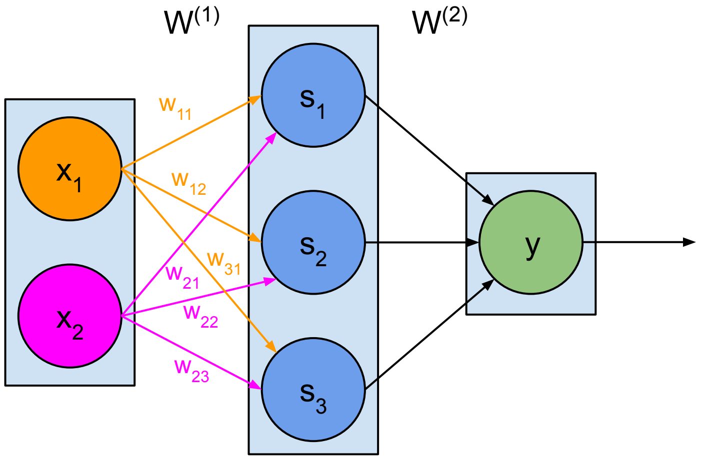
Computation in MLP



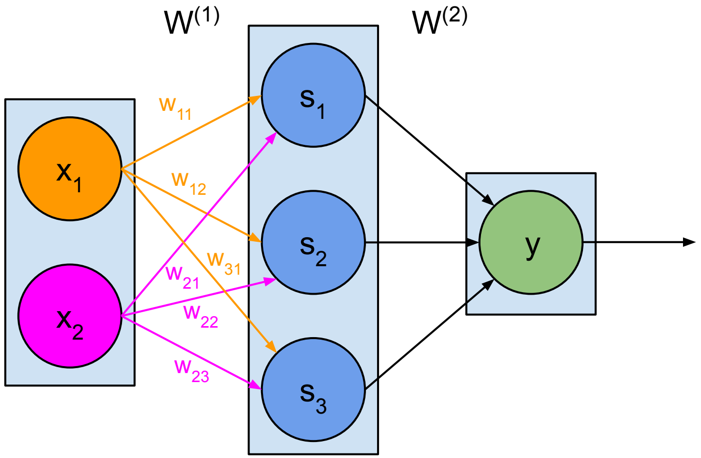
Computation in MLP

- Weights
- Net inputs (sum of weighted inputs)
- Activations (output of neurons to next layer)

Weights

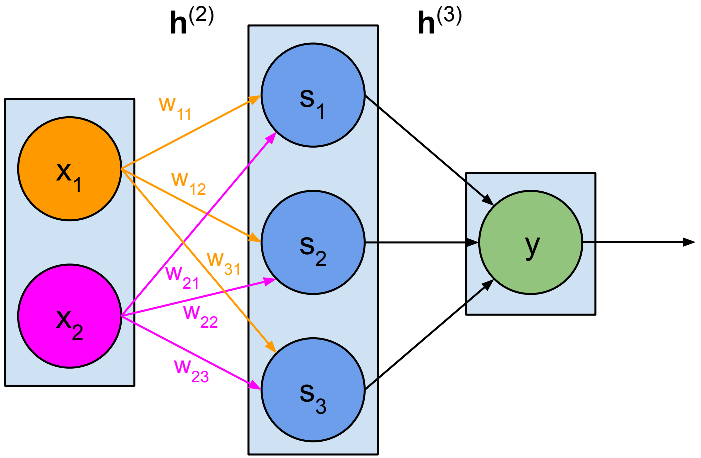


Weights

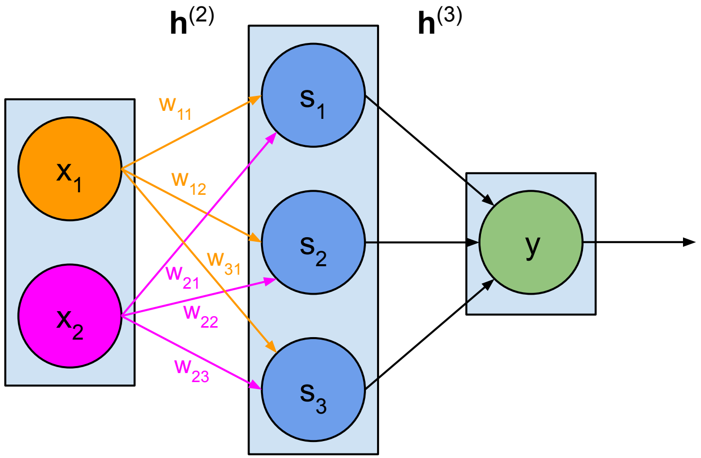


$$W^{(1)} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

Net input

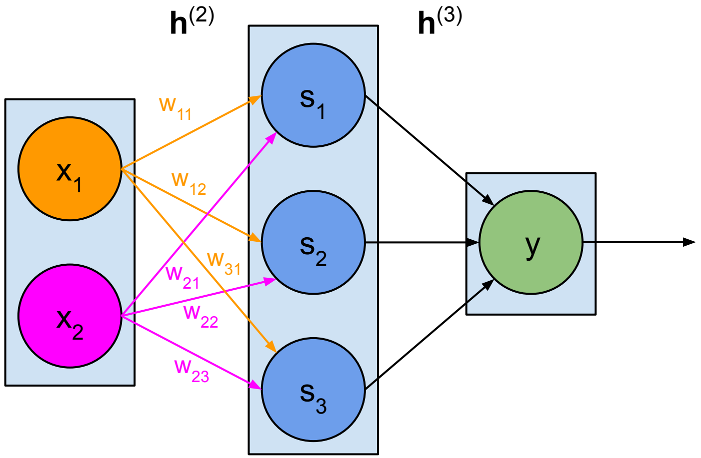


Net input



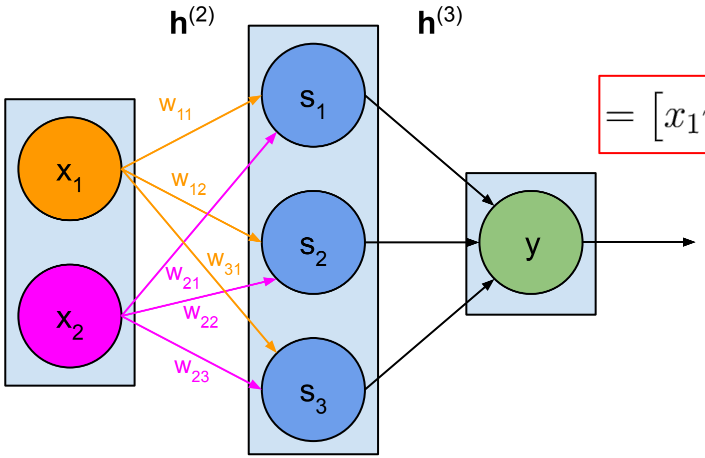
$$\boxed{\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}}$$

Net input



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)} = \boxed{\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}}$$

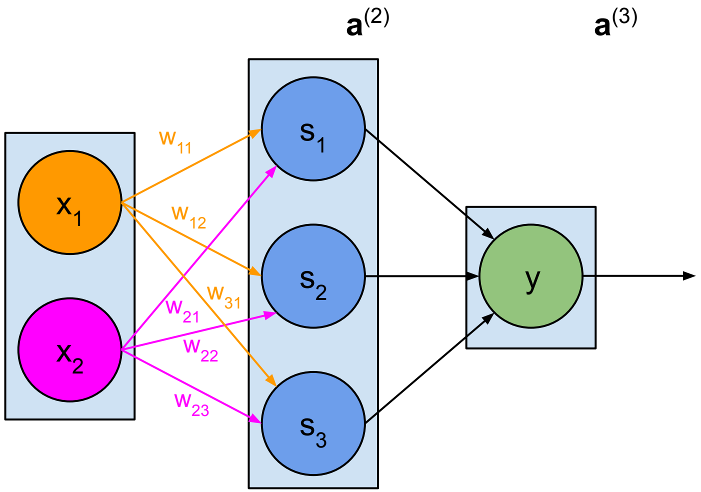
Net input



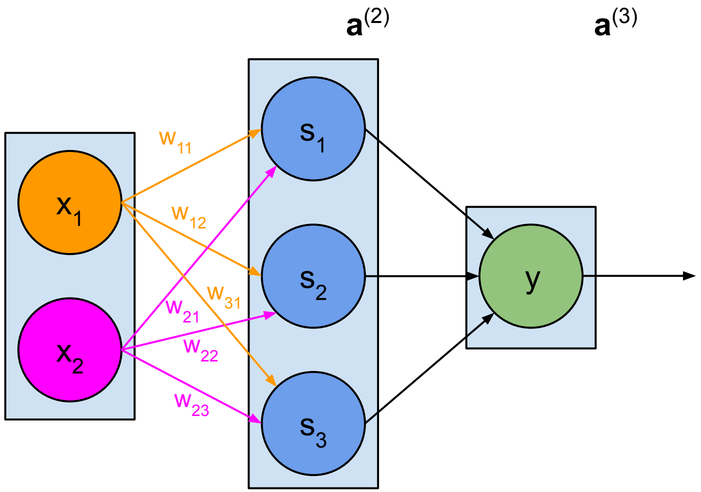
$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

$$\boxed{= \begin{bmatrix} x_1 w_{11} + x_2 w_{21} & x_1 w_{12} + x_2 w_{22} & x_1 w_{13} + x_2 w_{23} \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}}$$

Activation

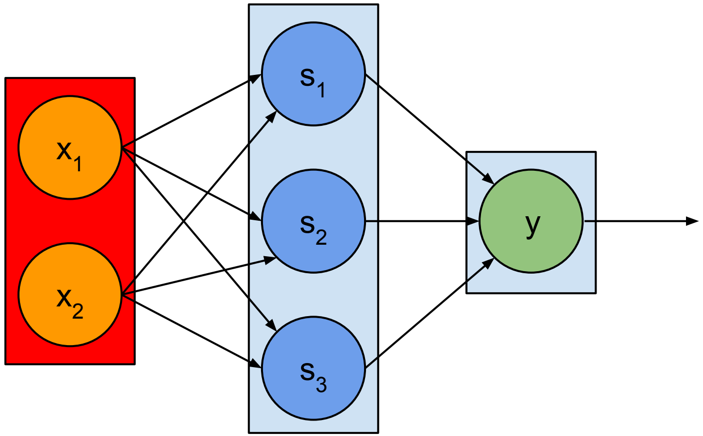


Activation



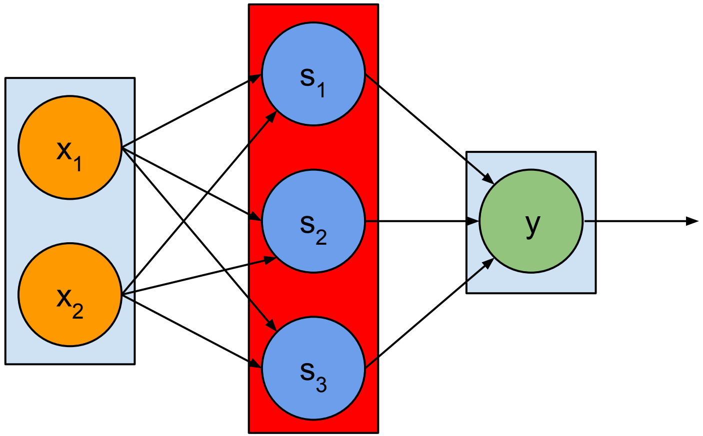
$$\mathbf{a}^{(2)} = f(\mathbf{h}^{(2)})$$

Computation in MLP (1st layer)



$$\mathbf{x}$$

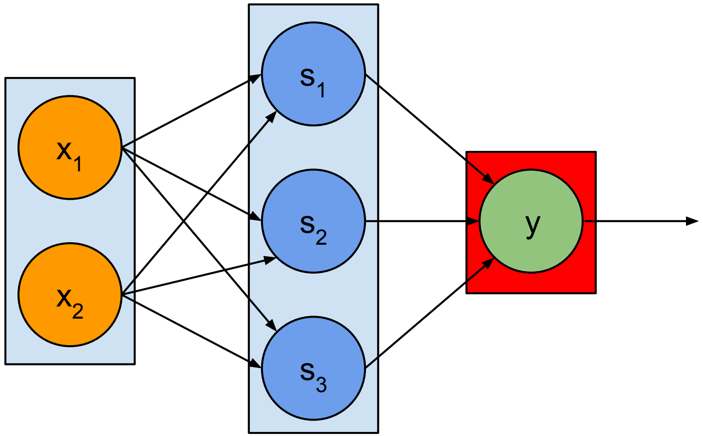
Computation in MLP (2nd layer)



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$\mathbf{a}^{(2)} = f(\mathbf{h}^{(2)})$$

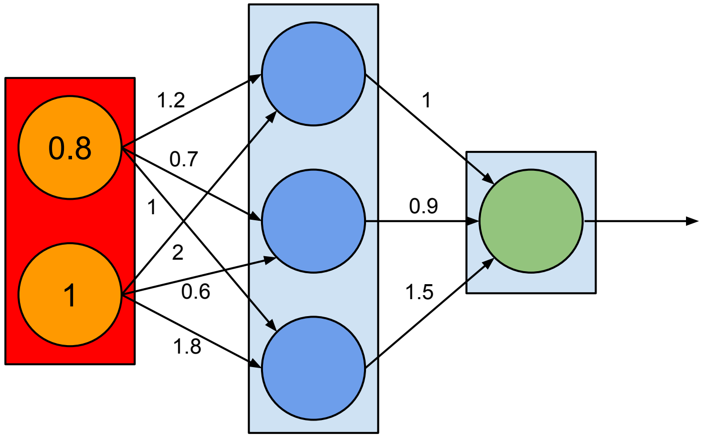
Computation in MLP (3d layer)



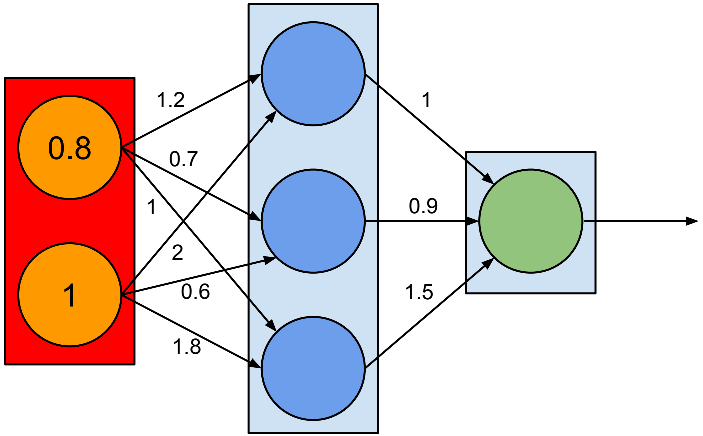
$$\mathbf{h}^{(3)} = \mathbf{a}^{(\mathbf{2})} W^{(2)}$$

$$y = f(\mathbf{h}^{(3)})$$

Sample computation

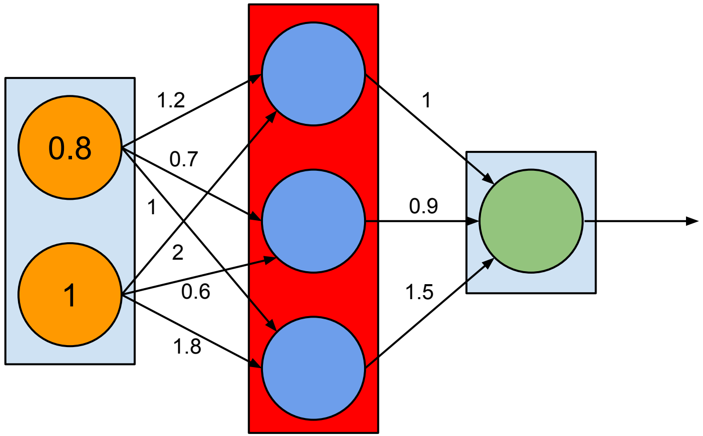


Sample computation



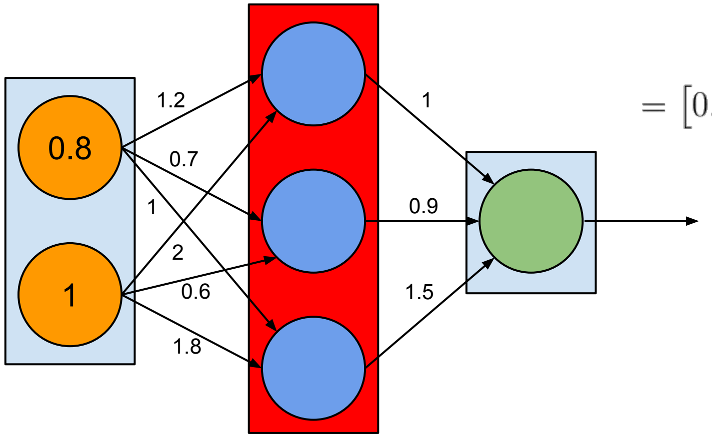
$$\mathbf{x} = \begin{bmatrix} 0.8 & 1 \end{bmatrix}$$

Sample computation



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

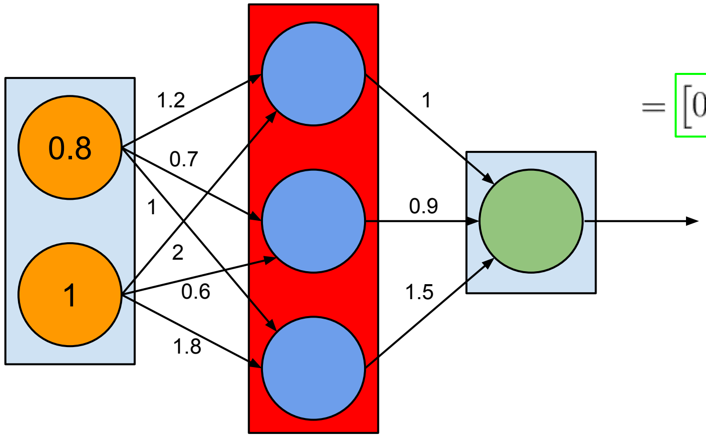
Sample computation



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$= \begin{bmatrix} 0.8 & 1 \end{bmatrix} \begin{bmatrix} 1.2 & 0.7 & 1 \\ 2 & 0.6 & 1.8 \end{bmatrix} = \begin{bmatrix} 0.8 \cdot 1.2 + 1 \cdot 2 & 0.8 \cdot 0.7 + 1 \cdot 0.6 & 0.8 \cdot 1 + 1 \cdot 1.8 \end{bmatrix}$$

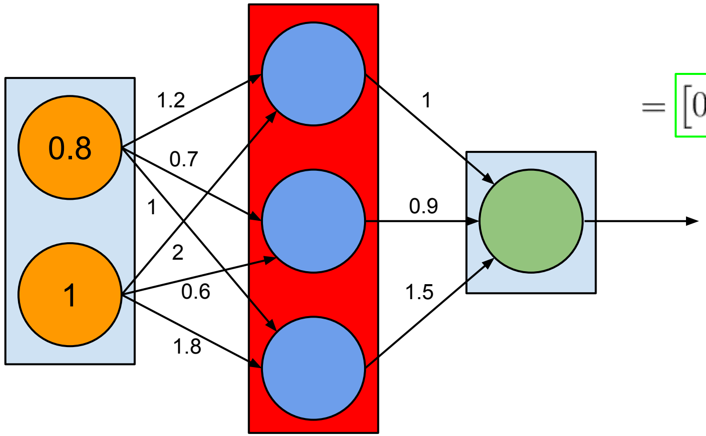
Sample computation



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$= \boxed{\begin{bmatrix} 0.8 & 1 \end{bmatrix}} \begin{bmatrix} \boxed{\begin{matrix} 1.2 \\ 2 \end{matrix}} & \begin{matrix} 0.7 \\ 0.6 \end{matrix} & \begin{matrix} 1 \\ 1.8 \end{matrix} \end{bmatrix} = \begin{bmatrix} \boxed{0.8 \cdot 1.2 + 1 \cdot 2} & 0.8 \cdot 0.7 + 1 \cdot 0.6 & 0.8 \cdot 1 + 1 \cdot 1.8 \end{bmatrix}$$

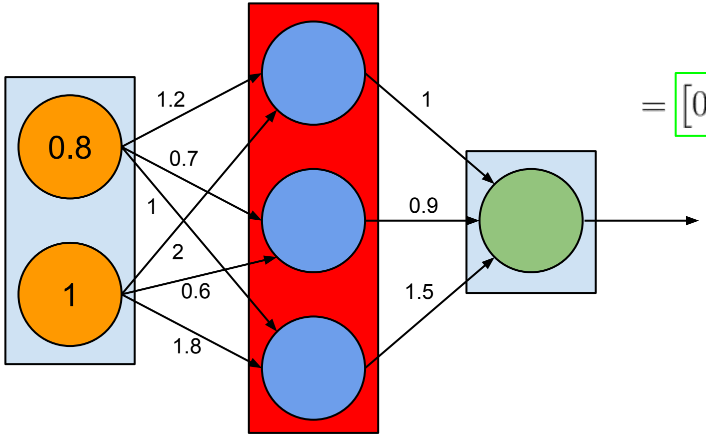
Sample computation



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$= \boxed{\begin{bmatrix} 0.8 & 1 \end{bmatrix}} \begin{bmatrix} 1.2 & \boxed{0.7} & 1 \\ 2 & \boxed{0.6} & 1.8 \end{bmatrix} = \begin{bmatrix} 0.8{\cdot}1.2 + 1{\cdot}2 & \boxed{0.8{\cdot}0.7 + 1{\cdot}0.6} & 0.8{\cdot}1 + 1{\cdot}1.8 \end{bmatrix}$$

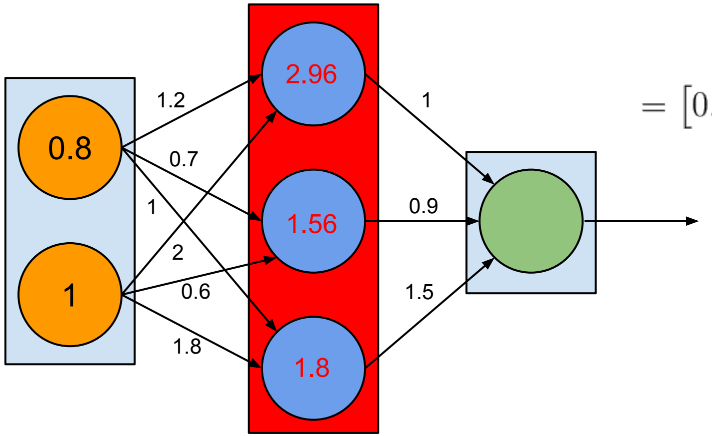
Sample computation



$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$= \boxed{\begin{bmatrix} 0.8 & 1 \end{bmatrix}} \begin{bmatrix} 1.2 & 0.7 & \boxed{\begin{matrix} 1 \\ 1.8 \end{matrix}} \\ 2 & 0.6 & \end{bmatrix} = \begin{bmatrix} 0.8{\cdot}1.2 + 1{\cdot}2 & 0.8{\cdot}0.7 + 1{\cdot}0.6 & \boxed{0.8{\cdot}1 + 1{\cdot}1.8} \end{bmatrix}$$

Sample computation

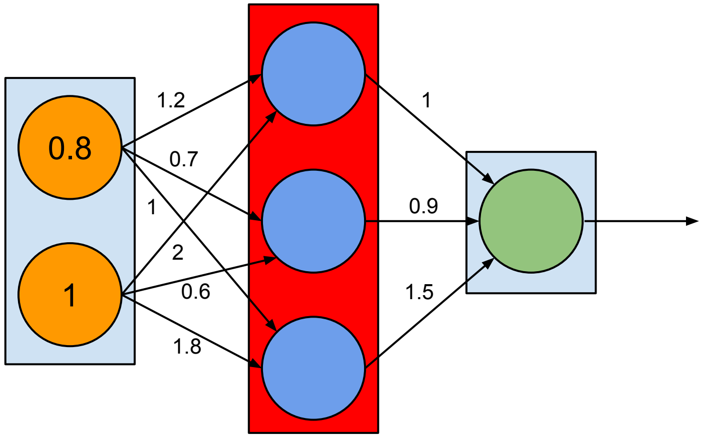


$$\mathbf{h}^{(2)} = \mathbf{x}W^{(1)}$$

$$= \begin{bmatrix} 0.8 & 1 \end{bmatrix} \begin{bmatrix} 1.2 & 0.7 & 1 \\ 2 & 0.6 & 1.8 \end{bmatrix} = \begin{bmatrix} 0.8 \cdot 1.2 + 1 \cdot 2 & 0.8 \cdot 0.7 + 1 \cdot 0.6 & 0.8 \cdot 1 + 1 \cdot 1.8 \end{bmatrix}$$

$$\boxed{= \begin{bmatrix} 2.96 & 1.56 & 1.8 \end{bmatrix}}$$

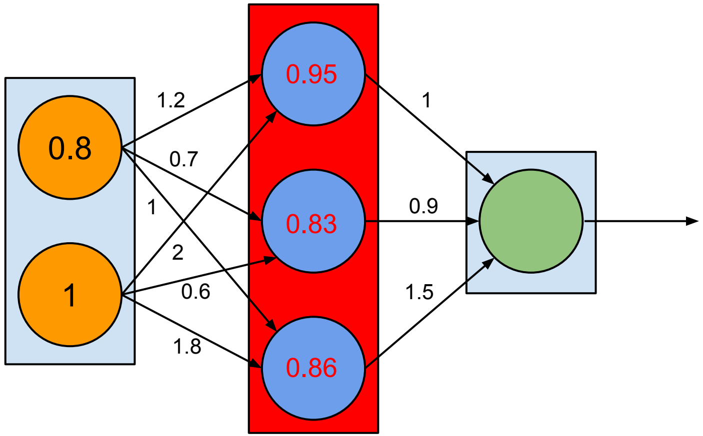
Sample computation



$$\mathbf{h}^{(2)} = \begin{bmatrix} 2.96 & 1.56 & 1.8 \end{bmatrix}$$

$$\mathbf{a}^{(2)} = \frac{1}{1 + e^{-\mathbf{h}^{(2)}}}$$

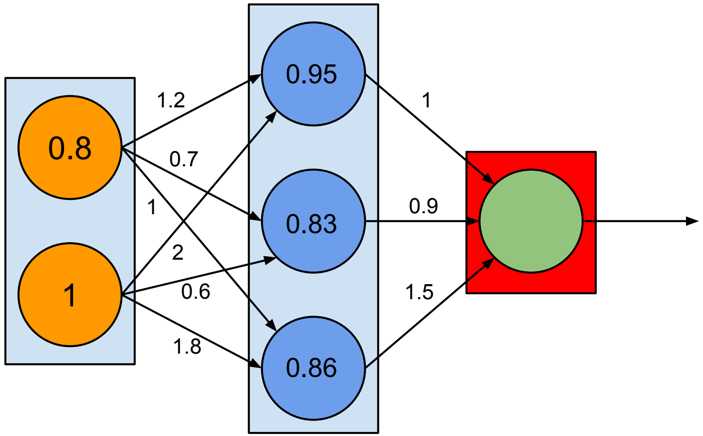
Sample computation



$$\mathbf{h}^{(2)} = \begin{bmatrix} 2.96 & 1.56 & 1.8 \end{bmatrix}$$

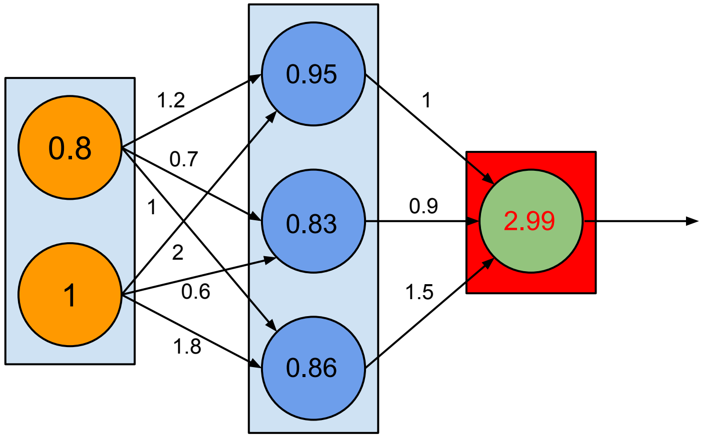
$$\mathbf{a}^{(2)} = \frac{1}{1+e^{-\mathbf{h}^{(2)}}} = \boxed{\begin{bmatrix} 0.95 & 0.83 & 0.86 \end{bmatrix}}$$

Sample computation



$$\mathbf{h}^{(3)} = \mathbf{a}^{(2)} W^{(2)}$$

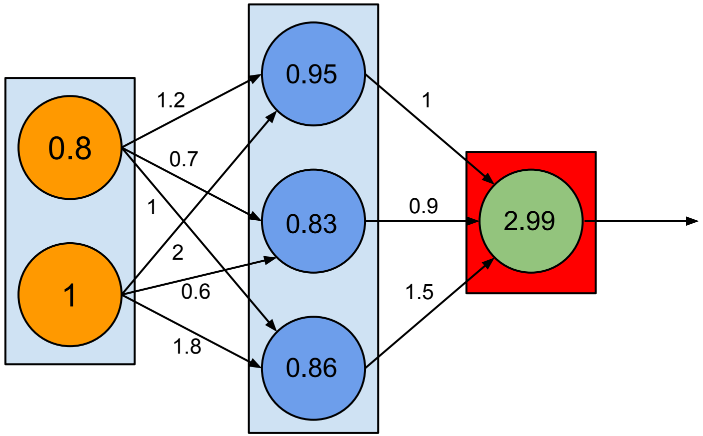
Sample computation



$$\mathbf{h}^{(3)} = \mathbf{a}^{(2)} W^{(2)}$$

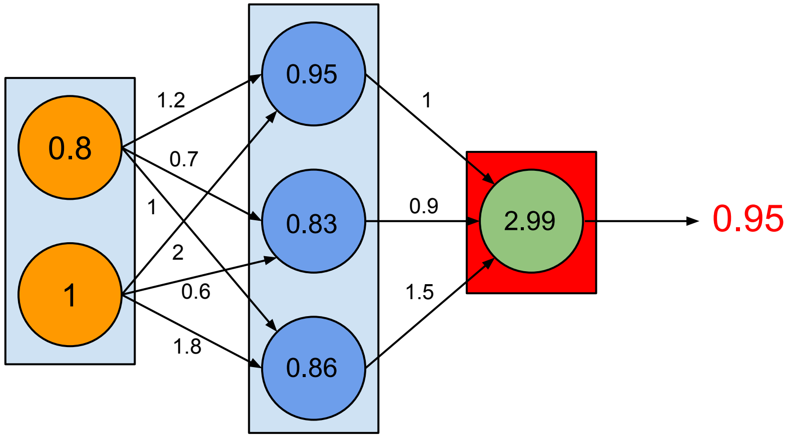
$$= \begin{bmatrix} 0.95 & 0.83 & 0.86 \end{bmatrix} \begin{bmatrix} 1 \\ 0.9 \\ 1.5 \end{bmatrix} = 2.99$$

Sample computation



$$y = \frac{1}{1 + e^{-\mathbf{h}^{(3)}}}$$

Sample computation



$$y = \frac{1}{1 + e^{-\mathbf{h}^{(3)}}}$$

$$\boxed{= \frac{1}{1 + e^{-2.99}} = 0.95}$$

Takeaway points

- ANNs work for complex problems
- Computation is distributed
- Signal moves from left to right
- Weights, net inputs and activations

What’s up next?