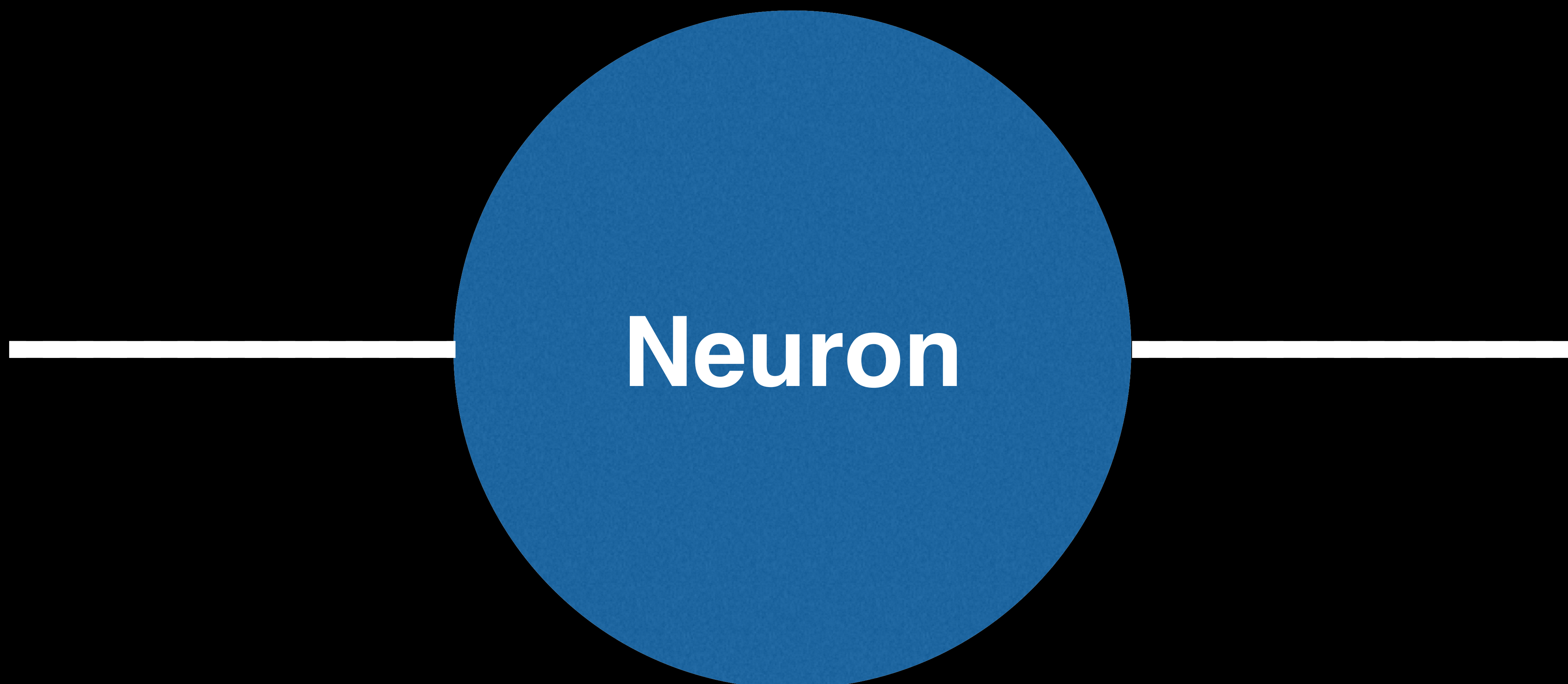
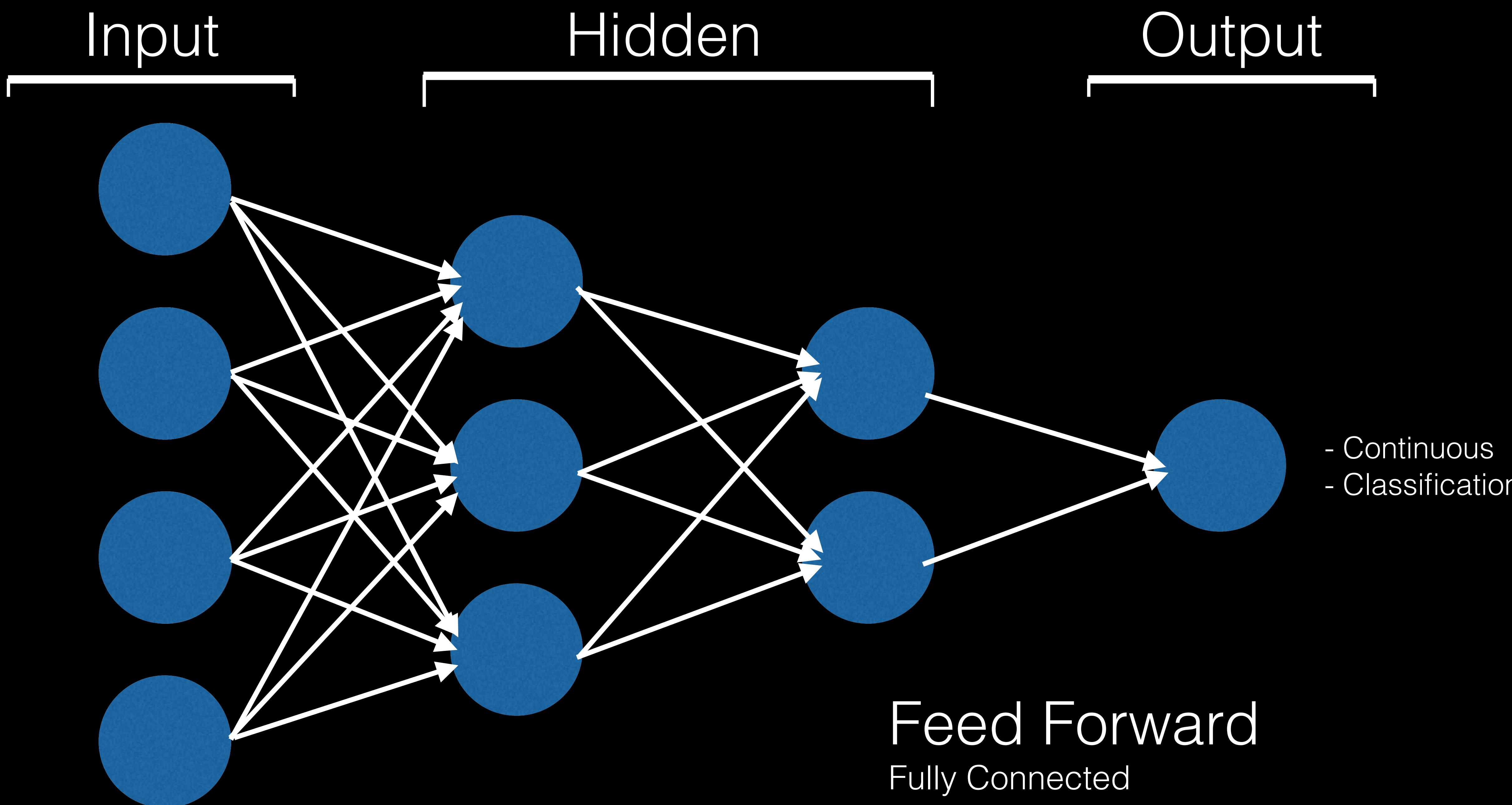
Intro to Neural Networks

Forward Propagation

David Yerrington


Neuron

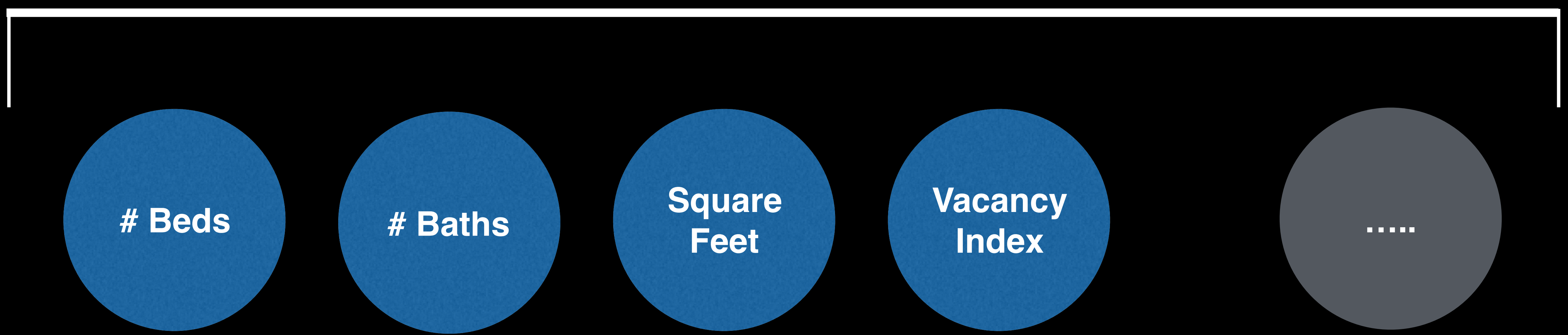


Feed Forward

Fully Connected

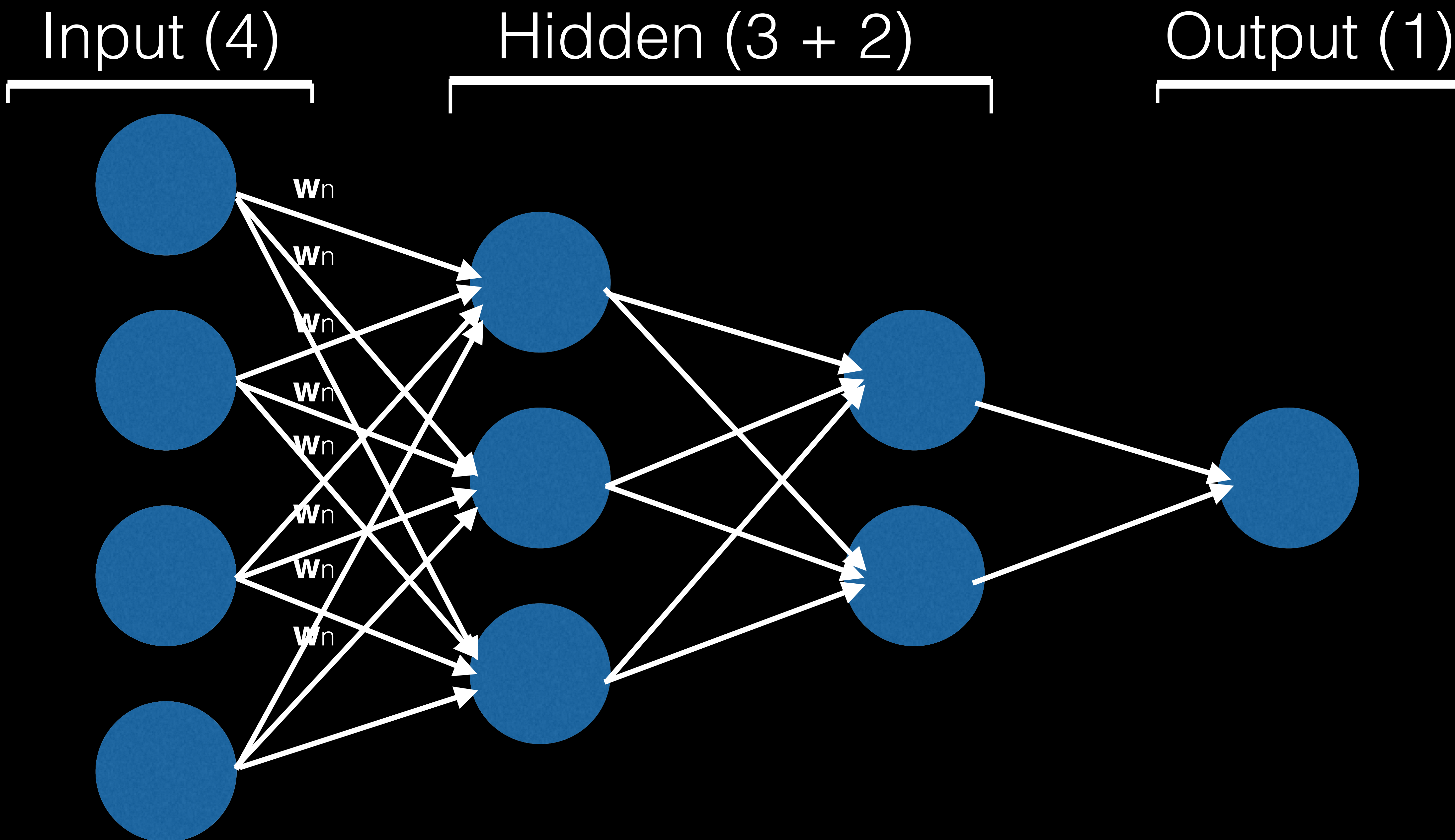
= *Neuron (aka: node)*

Input Layer



X_1	X_2	X_3	X_4	X_n
3	3	1500	1.2	
1	1	700	0.55	
2	1	850	1.12	
4	2	1825	1.43	
0	1	500	0.31	

Each observation is sent as the input of the first layer in our network.



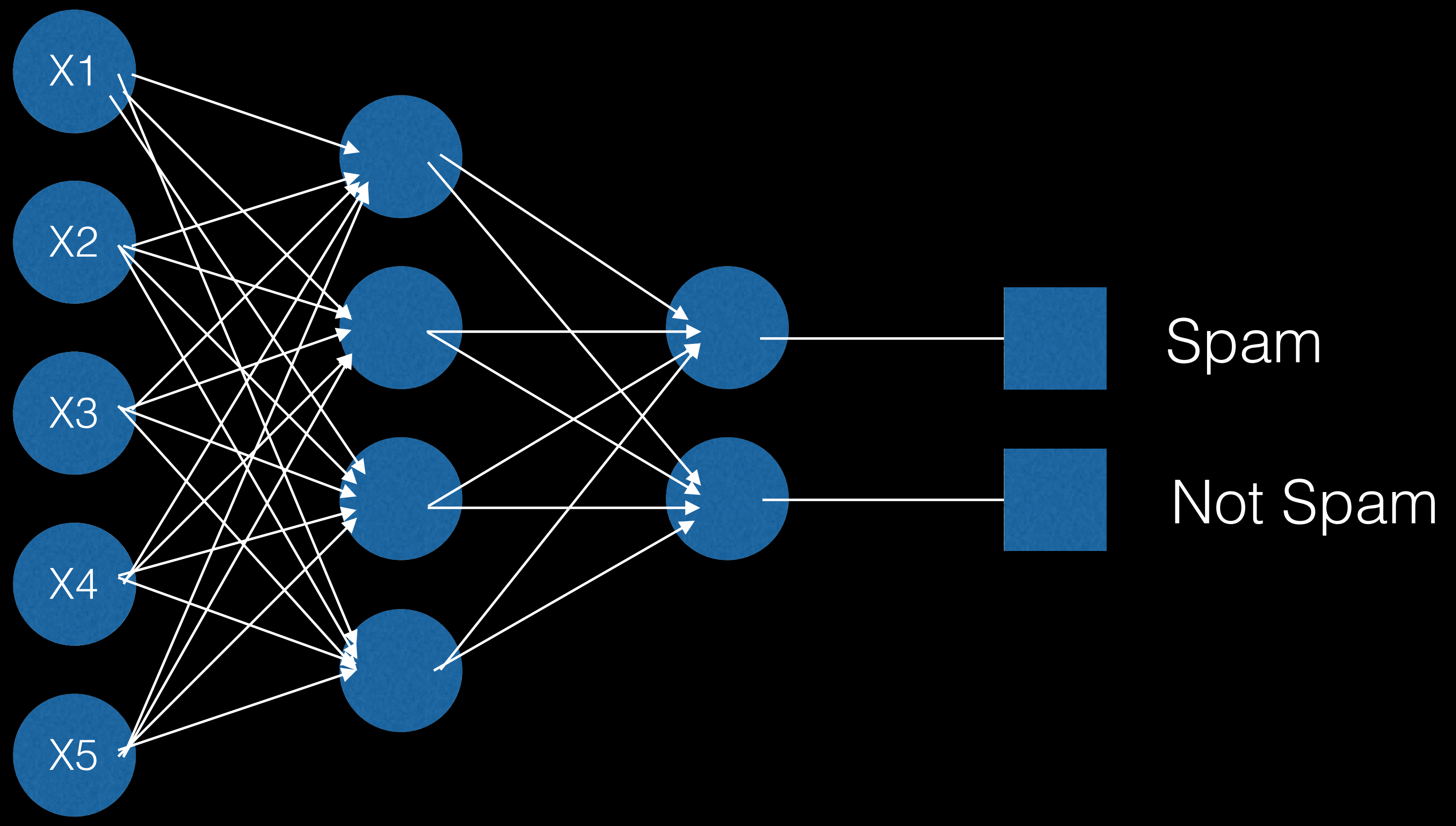
Each neuron is connected to other neurons via weights, $\mathbf{w}_n$.

n = feature index

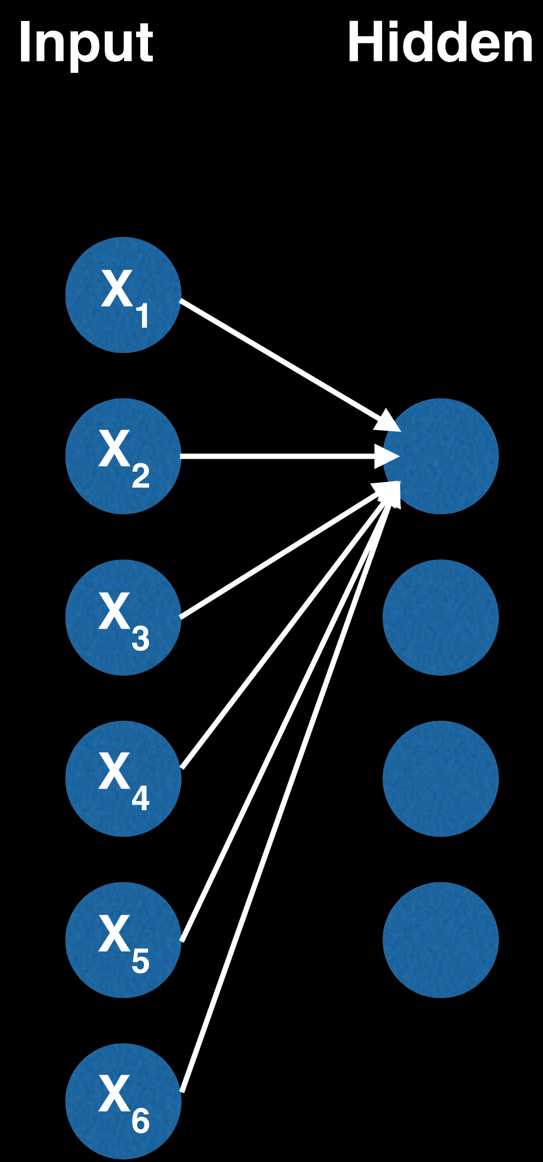
Activity (3 mins)

With a partner, identify:

- Input layer
- Hidden layer(s)
- Output layer
- # of nodes
- # of weights



Each neuron is connected to other neurons via weights.

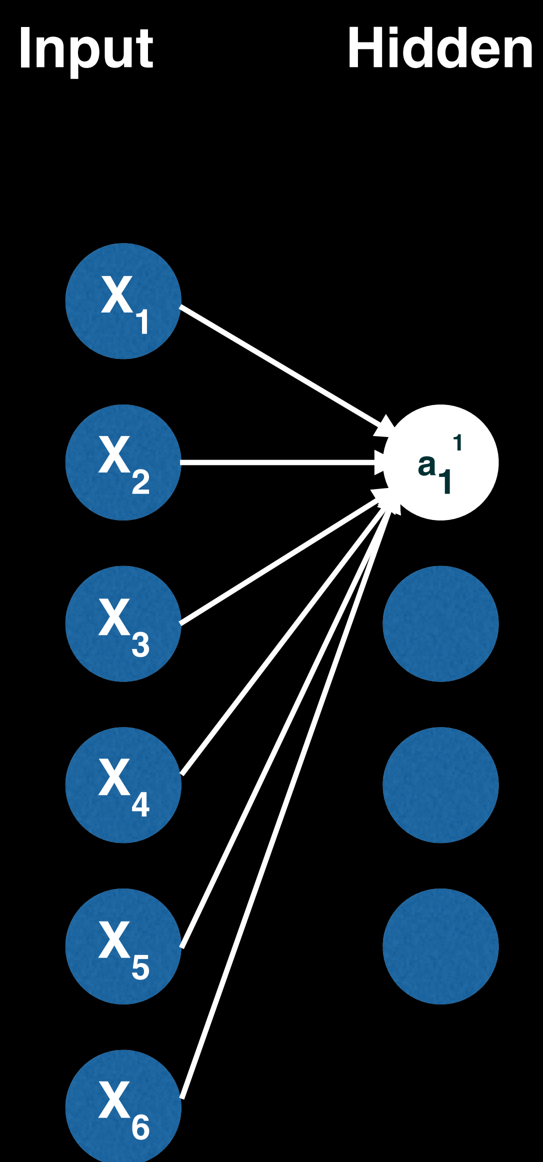


$$\mathbf{a}^0 = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \\ a_4^0 \\ a_5^0 \\ a_6^0 \end{bmatrix} \quad \mathbf{w}^0 = \begin{bmatrix} w_{0,1} & w_{0,2} & w_{0,3} & \cdots & w_{0,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Data, that is sent from one layer to neurons in the next. We will represent with the vector **a.**

Every output from one layer to the next, **a**, has a corresponding weight **w**.

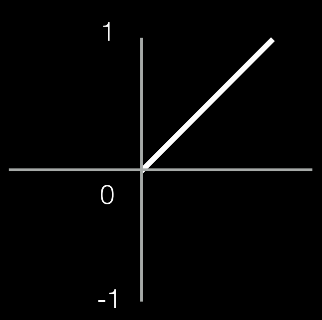
This is very similar to linear regression betas (theres also a bias term that is like the intercept!).



Activation

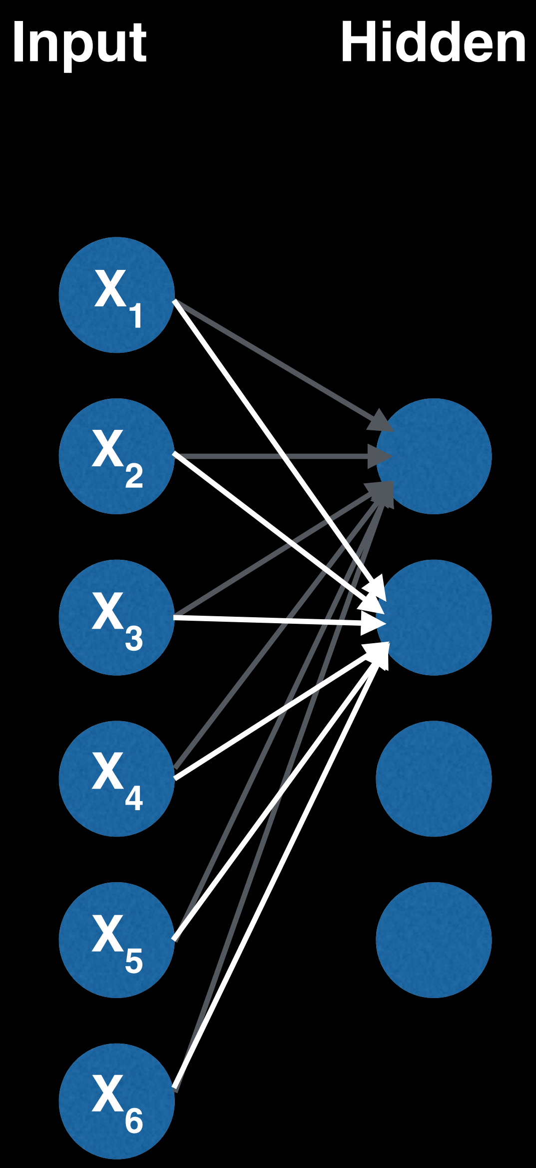
$$\sigma\left(w_{0,1}\, a_1^0 + w_{0,2}\, a_2^0 + w_{0,3}\, a_3^0 + w_{0,4}\, a_4^0 + w_{0,5}\, a_5^0 + w_{0,6}\, a_6^0 + B_0 \right)$$

$$\sigma =$$

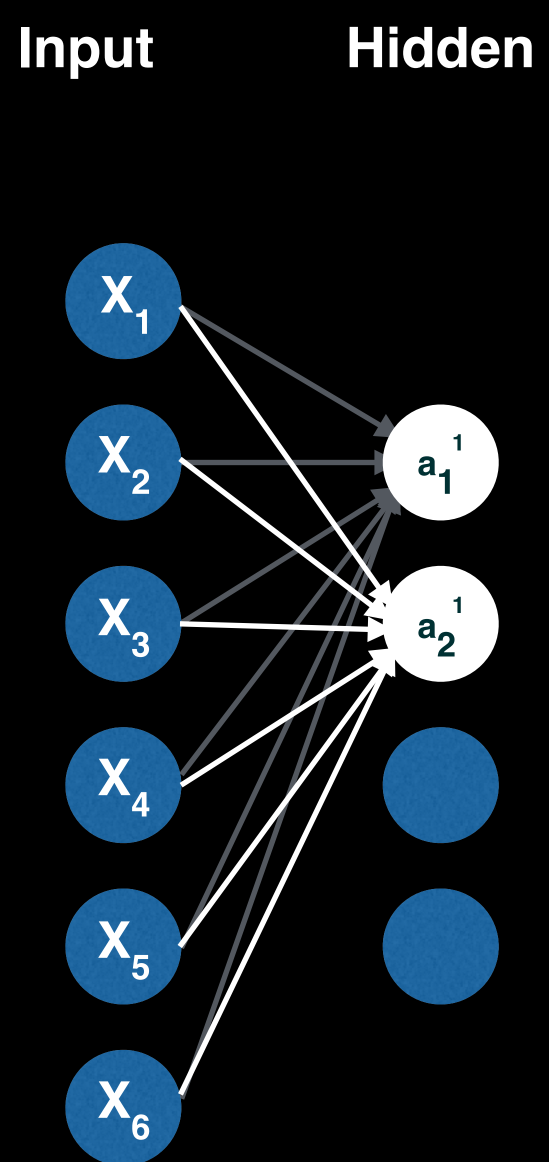


Linear Activation

$f(z) = \max(0, z)$



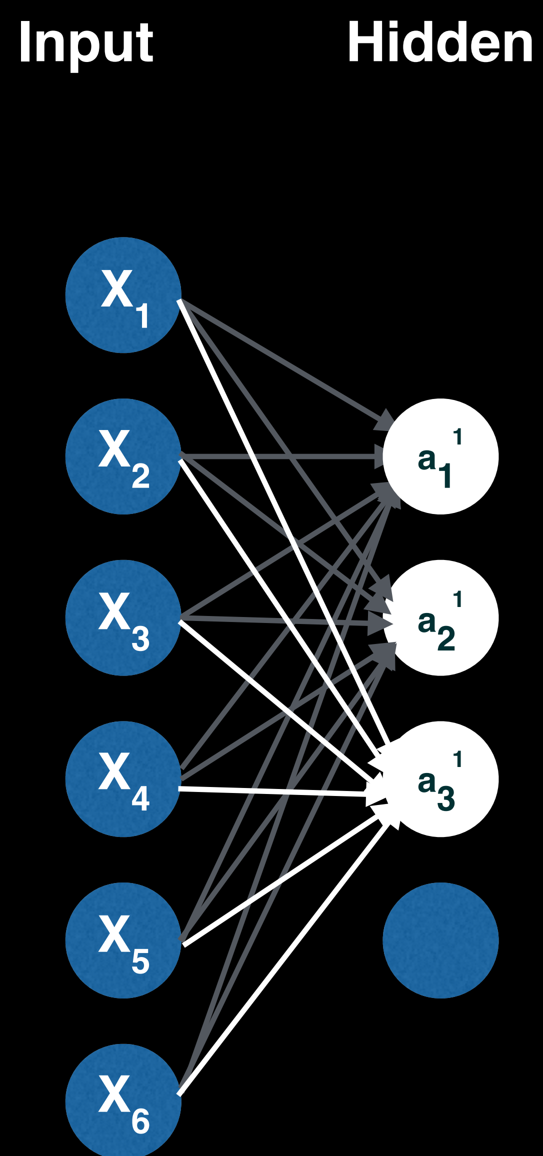
$$a^1 = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \\ a_4^0 \\ a_5^0 \\ a_6^0 \end{bmatrix} \quad W^0 = \begin{bmatrix} w_{0,1} & w_{0,2} & w_{0,3} & \cdots & w_{0,6} \\ w_{1,1} & w_{1,2} & w_{1,3} & \cdots & w_{1,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$



Activation

$$\sigma\left(w_{0,1}\, a_1^0 + w_{0,2}\, a_2^0 + w_{0,3}\, a_3^0 + w_{0,4}\, a_4^0 + w_{0,5}\, a_5^0 + w_{0,6}\, a_6^0 + B_0 \right)$$

$$\sigma\left(w_{1,1}\, a_1^0 + w_{1,2}\, a_2^0 + w_{1,3}\, a_3^0 + w_{1,4}\, a_4^0 + w_{1,5}\, a_5^0 + w_{1,6}\, a_6^0 + B_1 \right)$$

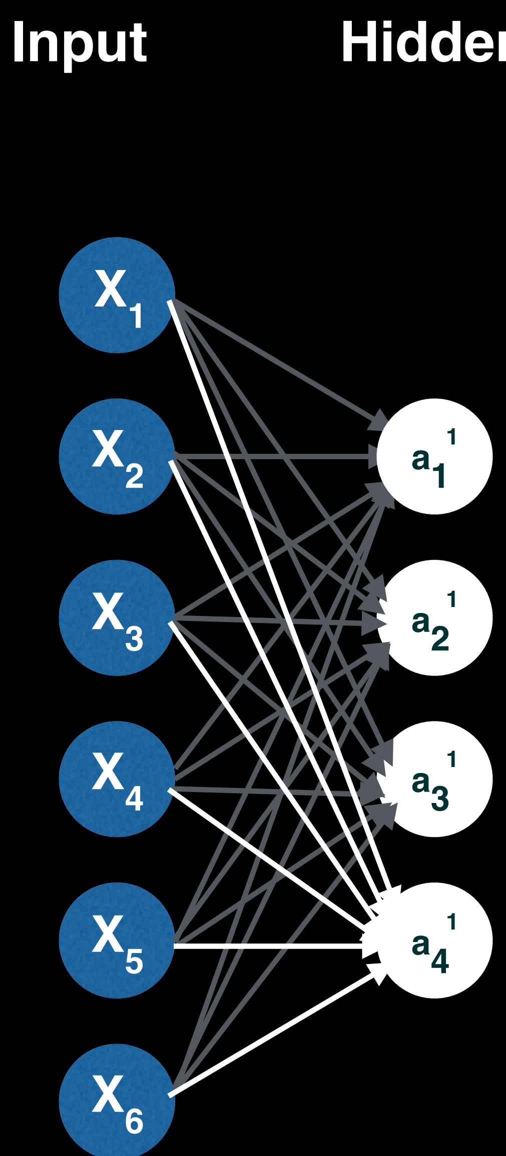


Activation

$$\sigma\left(w_{0,1}\, a_1^0 + w_{0,2}\, a_2^0 + w_{0,3}\, a_3^0 + w_{0,4}\, a_4^0 + w_{0,5}\, a_5^0 + w_{0,6}\, a_6^0 + B_0 \right)$$

$$\sigma\left(w_{1,1}\, a_1^0 + w_{1,2}\, a_2^0 + w_{1,3}\, a_3^0 + w_{1,4}\, a_4^0 + w_{1,5}\, a_5^0 + w_{1,6}\, a_6^0 + B_1 \right)$$

$$\sigma\left(w_{2,1}\, a_1^0 + w_{2,2}\, a_2^0 + w_{2,3}\, a_3^0 + w_{2,4}\, a_4^0 + w_{2,5}\, a_5^0 + w_{2,6}\, a_6^0 + B_2 \right)$$



Activation

$$\sigma\left(w_{0,1}\, a_1^0 + w_{0,2}\, a_2^0 + w_{0,3}\, a_3^0 + w_{0,4}\, a_4^0 + w_{0,5}\, a_5^0 + w_{0,6}\, a_6^0 + B_0 \right)$$

$$\sigma\left(w_{1,1}\, a_1^0 + w_{1,2}\, a_2^0 + w_{1,3}\, a_3^0 + w_{1,4}\, a_4^0 + w_{1,5}\, a_5^0 + w_{1,6}\, a_6^0 + B_1 \right)$$

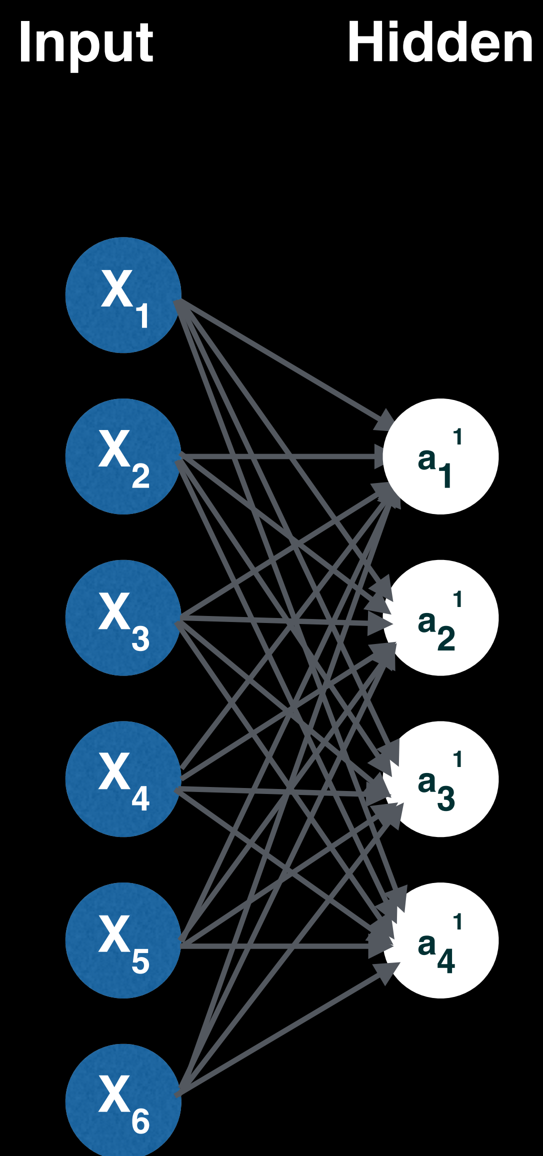
$$\sigma\left(w_{2,1}\, a_1^0 + w_{2,2}\, a_2^0 + w_{2,3}\, a_3^0 + w_{2,4}\, a_4^0 + w_{2,5}\, a_5^0 + w_{2,6}\, a_6^0 + B_2 \right)$$

$$\sigma\left(w_{3,1}\, a_1^0 + w_{3,2}\, a_2^0 + w_{3,3}\, a_3^0 + w_{3,4}\, a_4^0 + w_{3,5}\, a_5^0 + w_{3,6}\, a_6^0 + B_3 \right)$$

$$\sigma\left(\sum_k w^i_{jk}\, a^{i-1}_k + b^i_j \right)$$

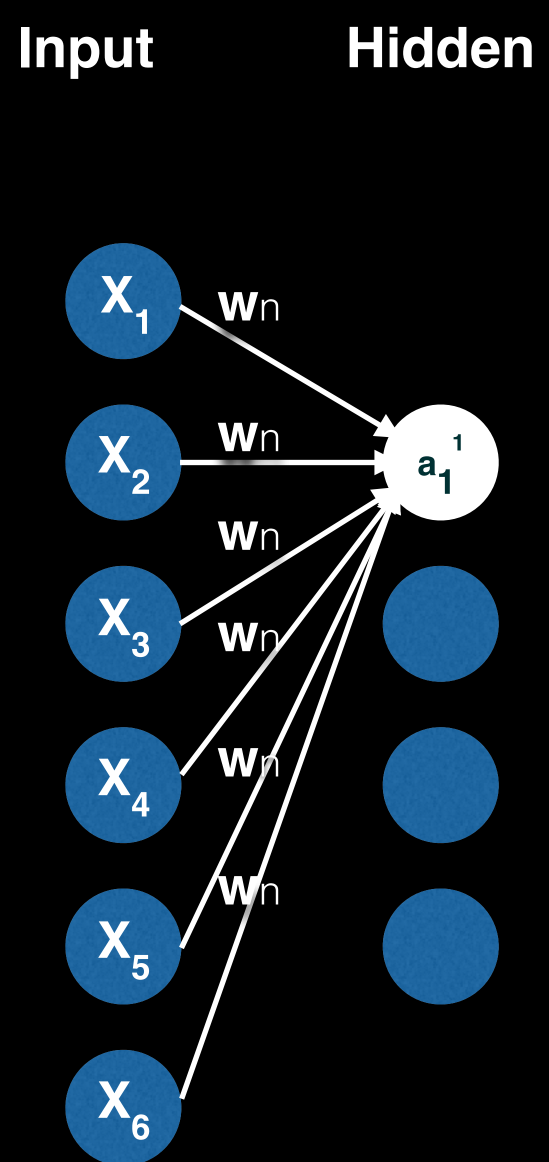
The output of the **k**th neuron in the previous layer

j = the weight of the neuron connecting to **k**th neuron $\mathbf{a}^{\mathbf{i-1}}$



$$a^{(1)} = \sigma(\qquad)$$

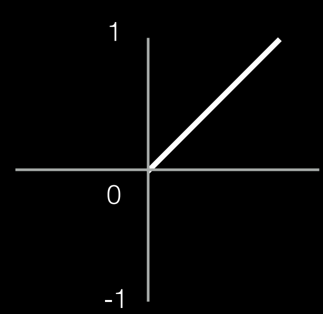
$$\sigma\left(\begin{bmatrix} w_{0,1} & w_{0,2} & w_{0,3} & \cdots & w_{0,6} \\ w_{1,1} & w_{1,2} & w_{1,3} & \cdots & w_{1,6} \\ \vdots & \vdots & \overset{W}{\vdots} & \vdots & \vdots \\ w_{n,1} & w_{n,2} & w_{n,3} & \cdots & w_{n,6} \end{bmatrix} \begin{bmatrix} a_1^0 \\ a_2^0 \\ \overset{a^{(0)}}{\vdots} \\ a_n^0 \end{bmatrix} + \begin{bmatrix} B_0 \\ B_1 \\ \overset{B}{\vdots} \\ b_n^0 \end{bmatrix} \right)$$



Activation Review

$$\sigma\left(w_{0,1}\, a_1^0 + w_{0,2}\, a_2^0 + w_{0,3}\, a_3^0 + w_{0,4}\, a_4^0 + w_{0,5}\, a_5^0 + w_{0,6}\, a_6^0 + B_0 \right)$$

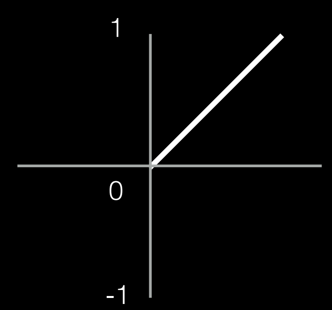
$\sigma =$



Linear Activation

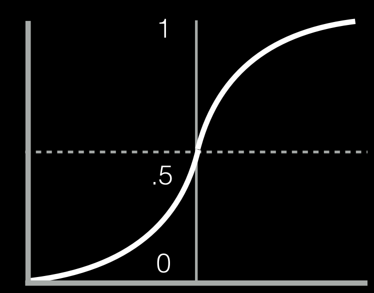
$f(z) = \max(0, z)$

Common Activation Functions



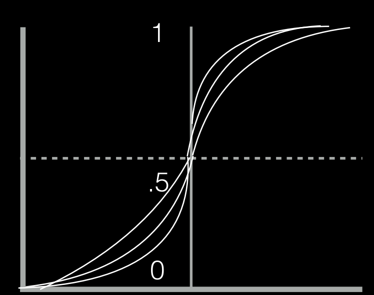
Linear Activation
RELU

f(z) = max(0, z)



Sigmoid

$$f(x) = \frac{1}{1 + e^{-x}}$$



Softmax

$$S(y_i) = \frac{e^{y_i}}{\sum e^{-y_i}}$$

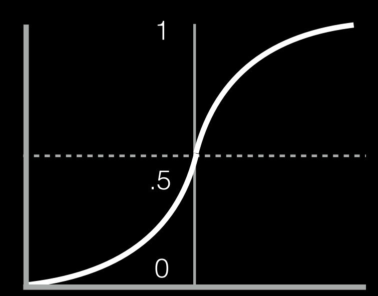
Common Output Functions

Regression

Linear Function

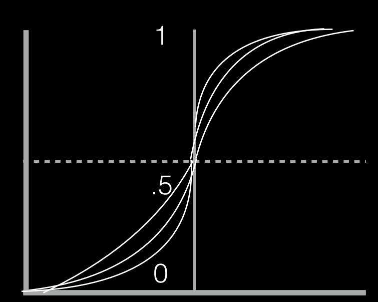
$$f(x) = x$$

Classification



Sigmoid

$$f(x) = \frac{1}{1 + e^{-x}}$$



Softmax

$$S(y_i) = \frac{e^{y_i}}{\sum e^{-y_i}}$$

Let’s code this example: