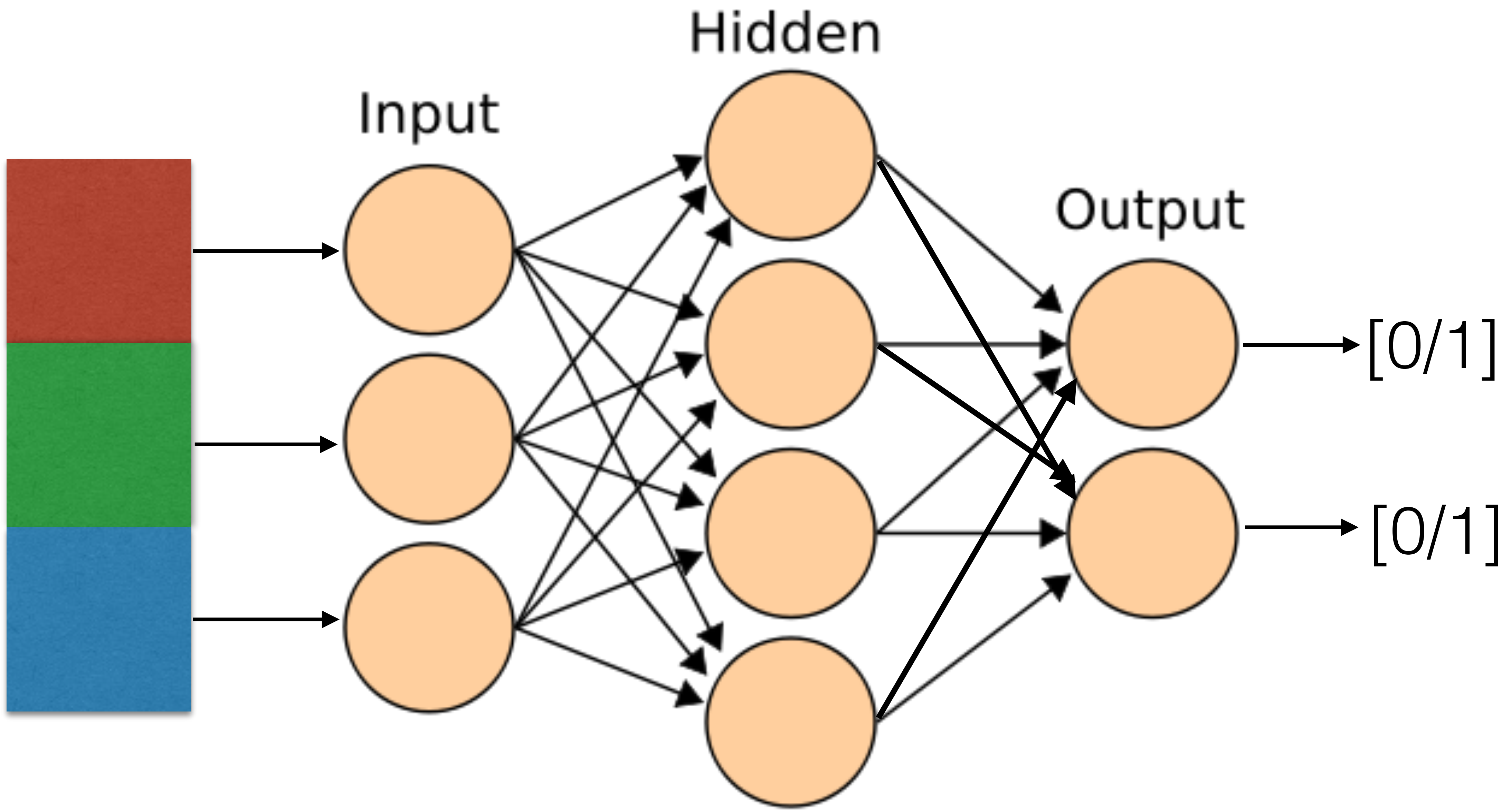
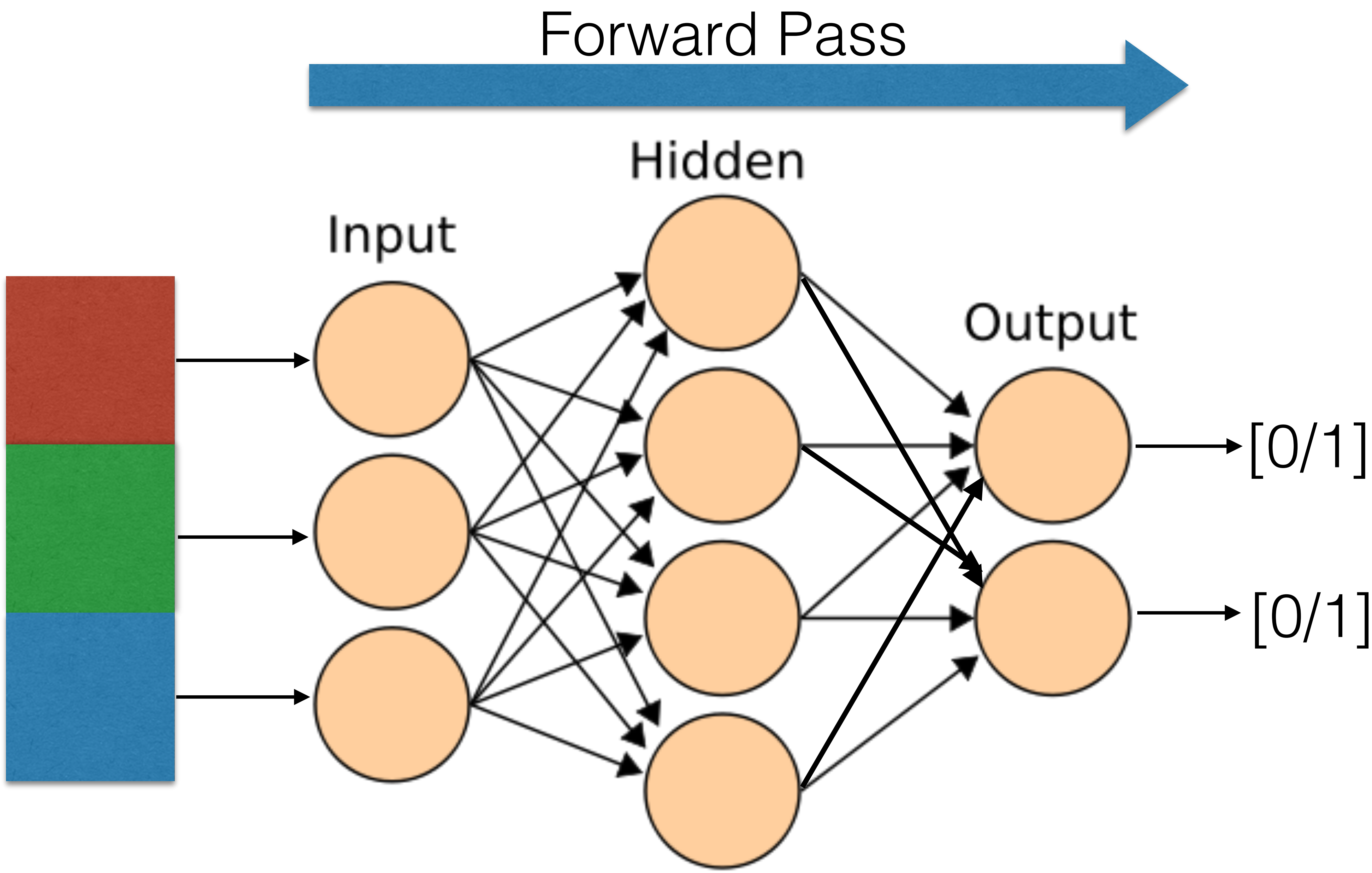
Feed Forward Artificial Neural Networks Trained Via Backpropagation

Joseph Suarez'15

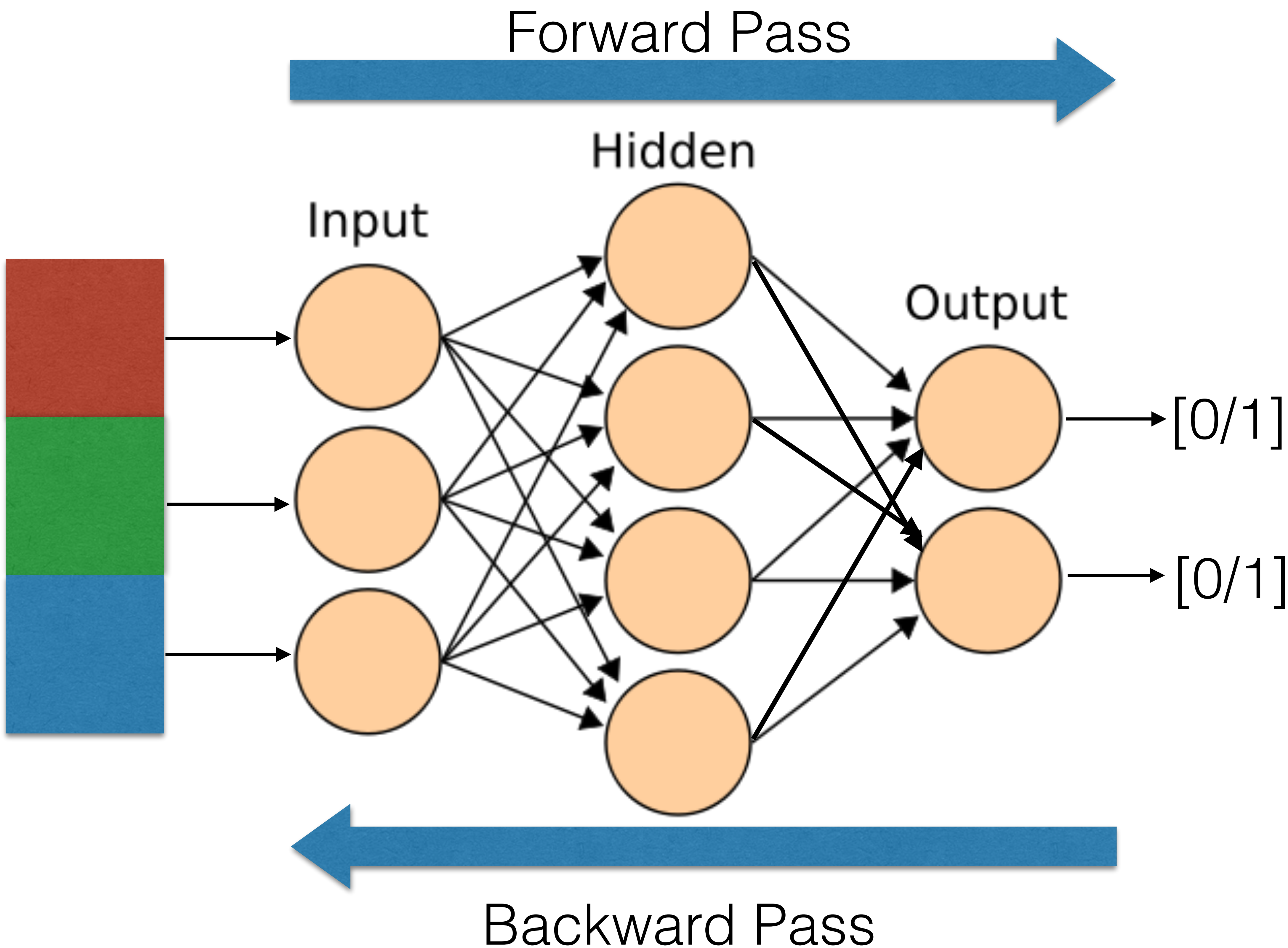
Architecture



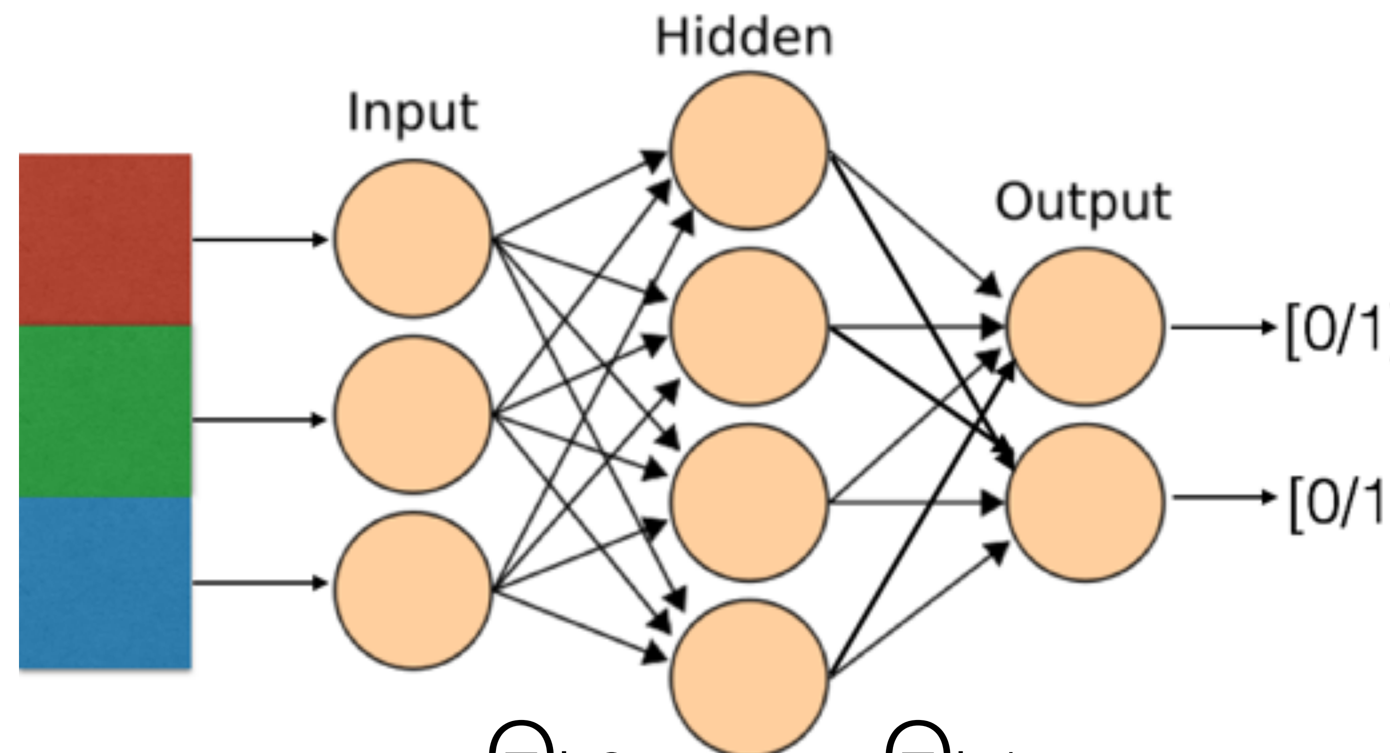
Architecture



Architecture



Forward Pass

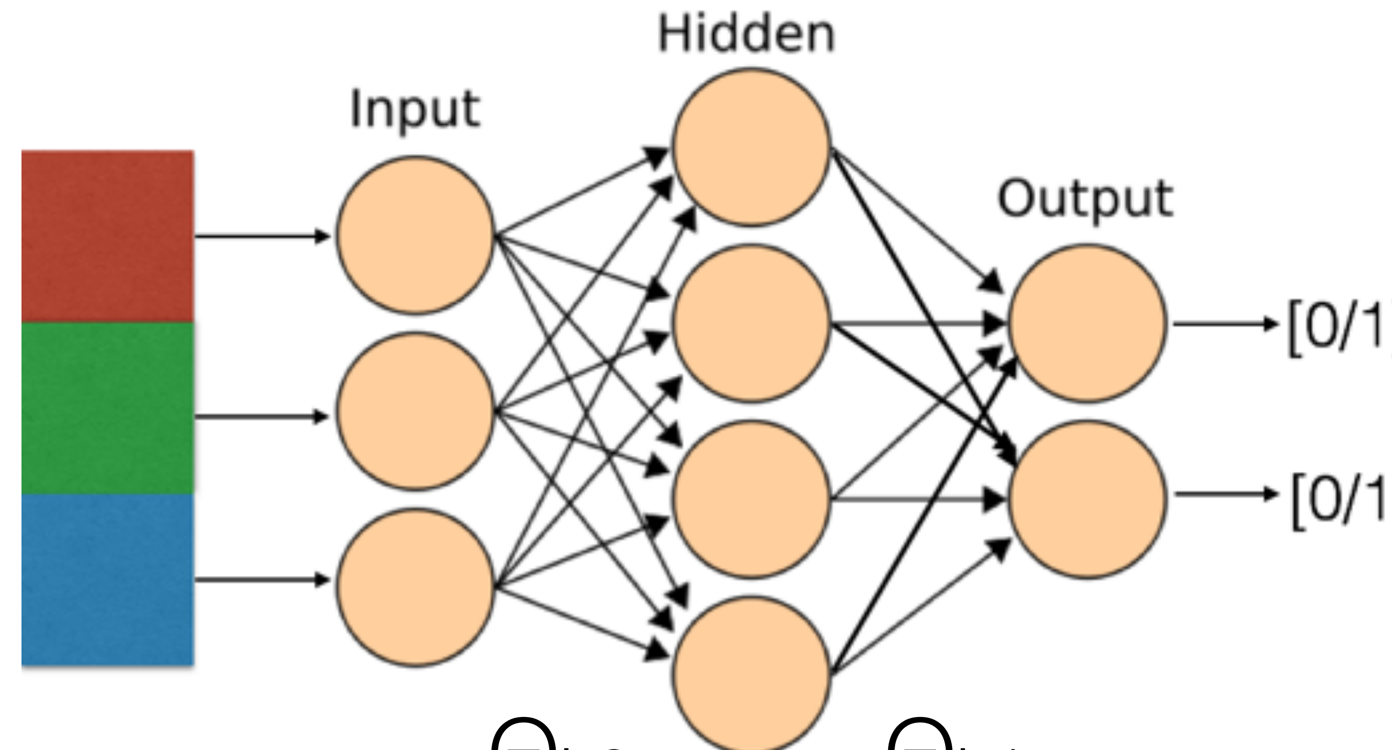


Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

Forward Pass

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix}$$



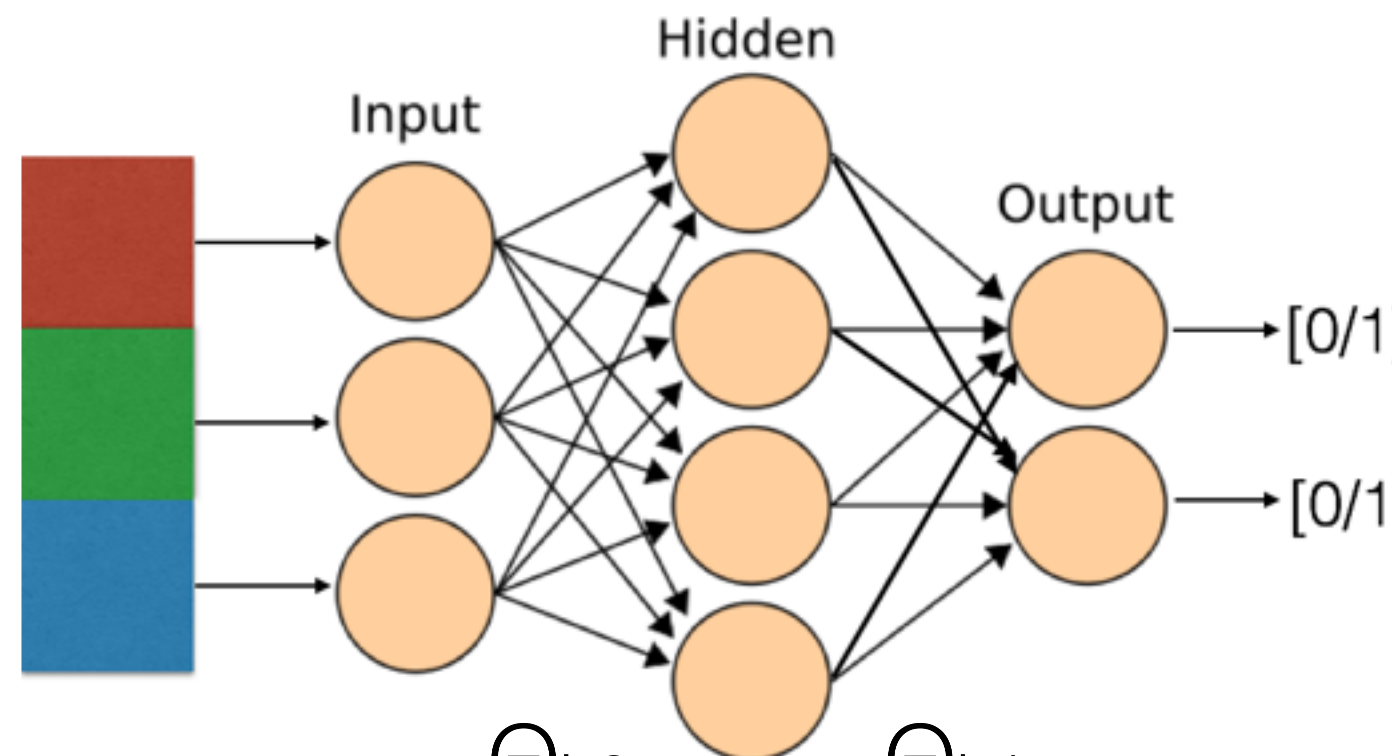
Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

Forward Pass

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix}$$

$$\mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$



Θ_{i-2} Θ_{i-1}

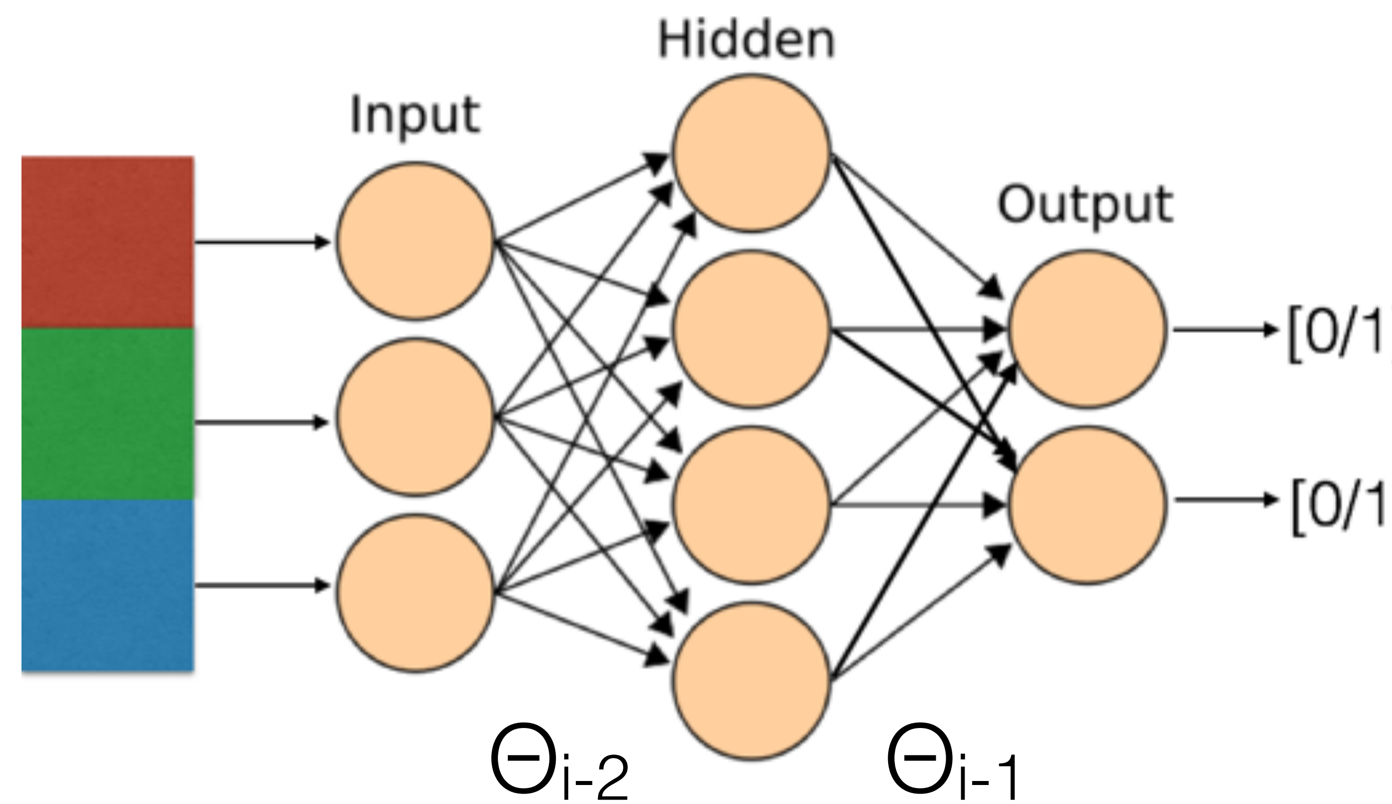
l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

Forward Pass

$$\boldsymbol{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix}$$

$$\mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

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



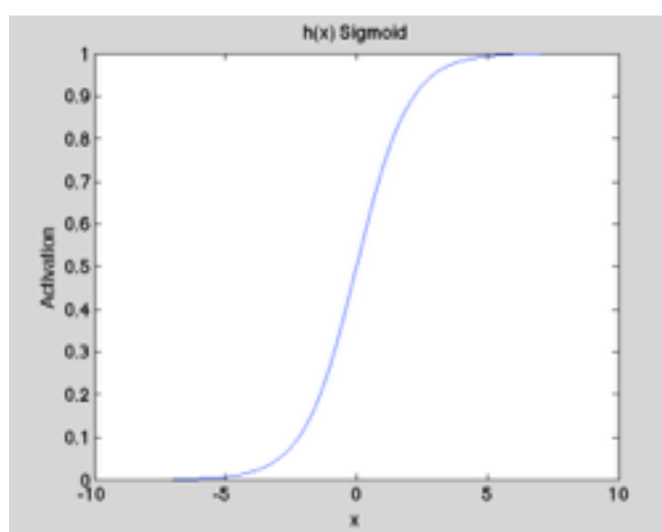
l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

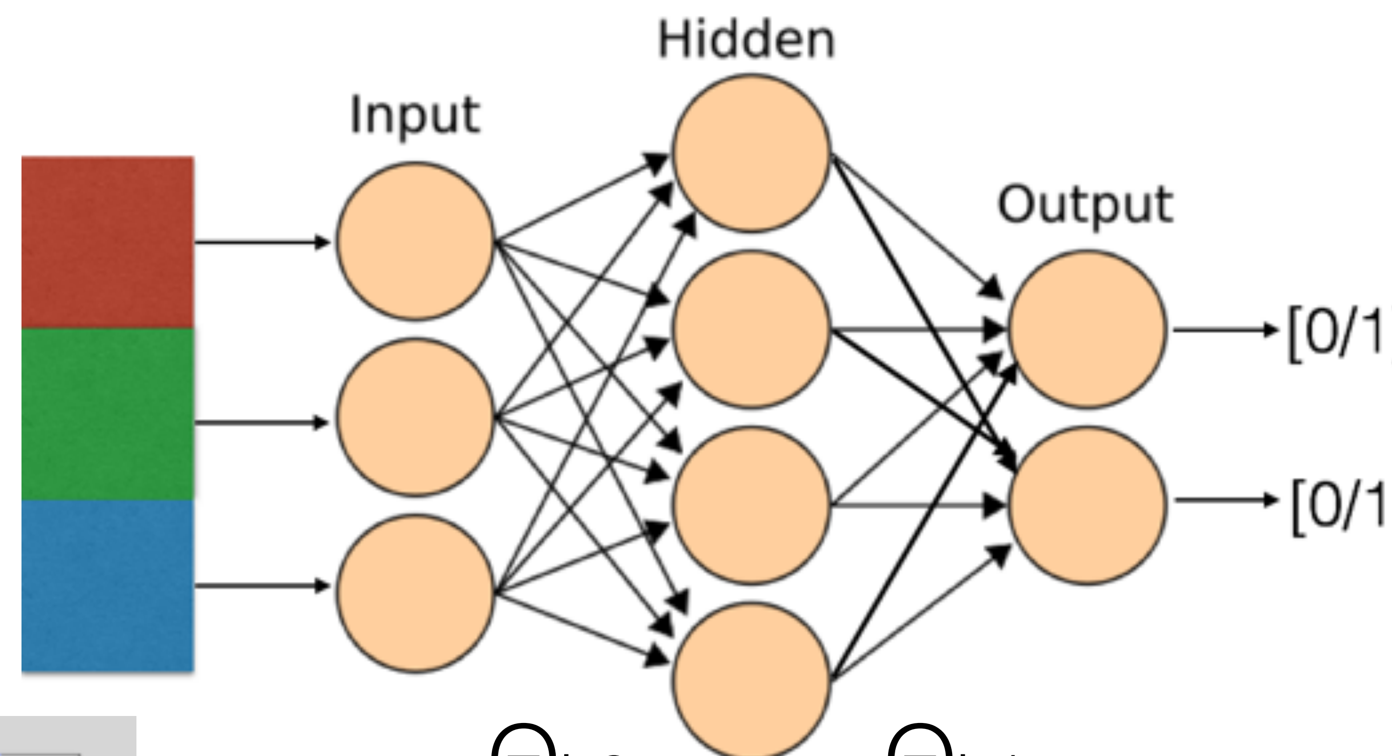
Forward Pass

$$\boldsymbol{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix}$$

$$\mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

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





Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

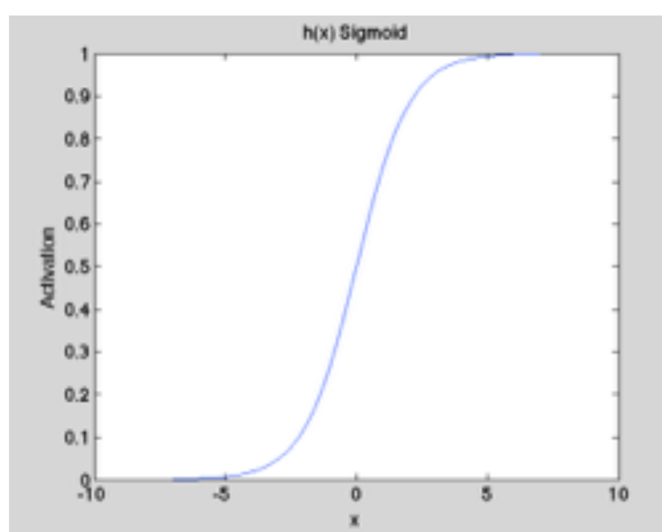
Forward Pass

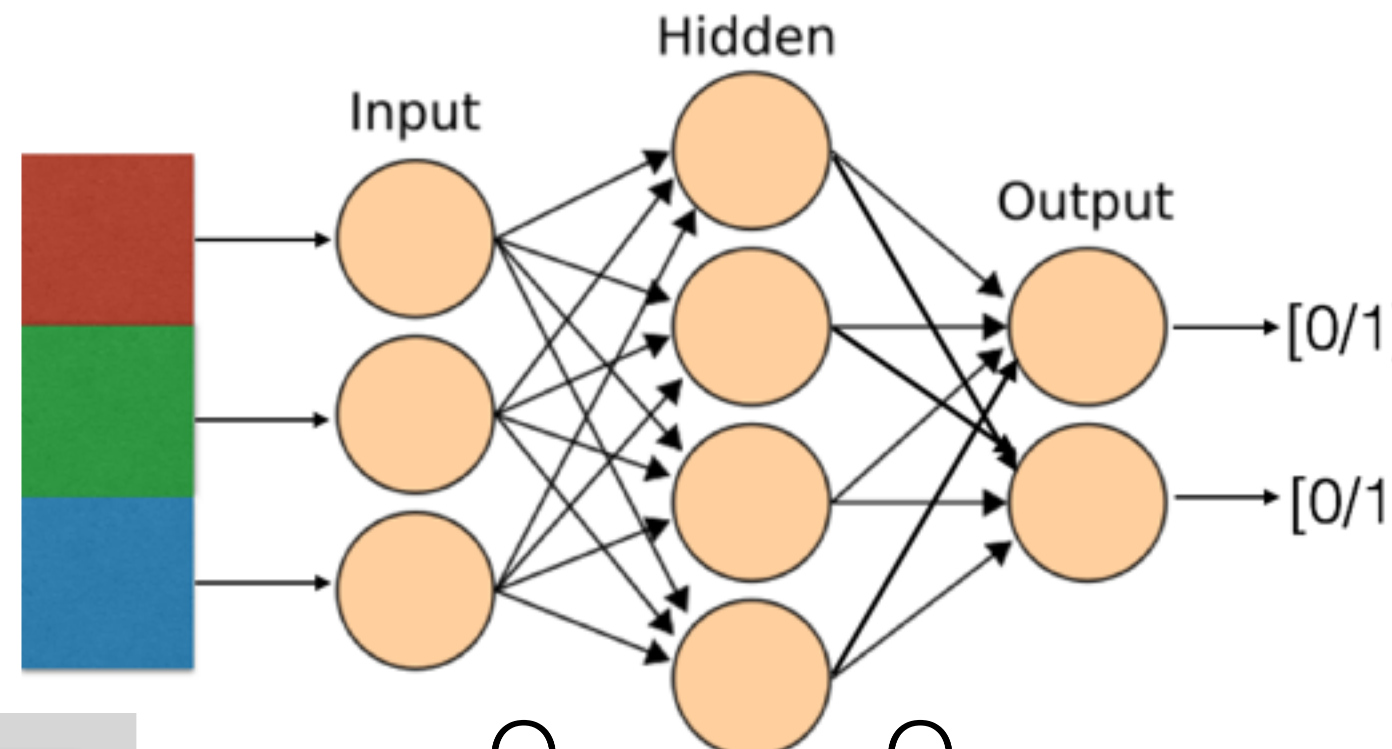
$$\boldsymbol{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix}$$

$$\mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

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

$$\boldsymbol{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})$$

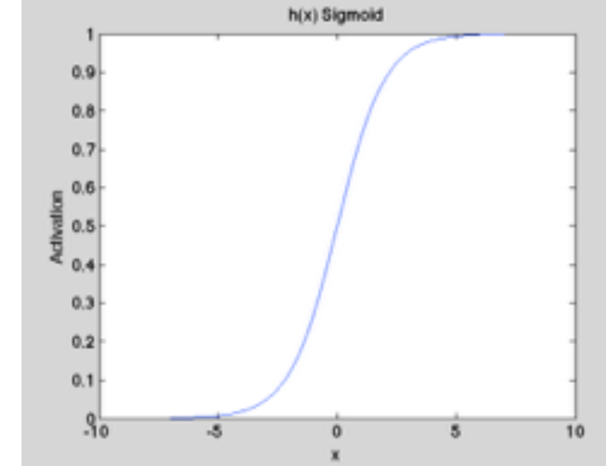




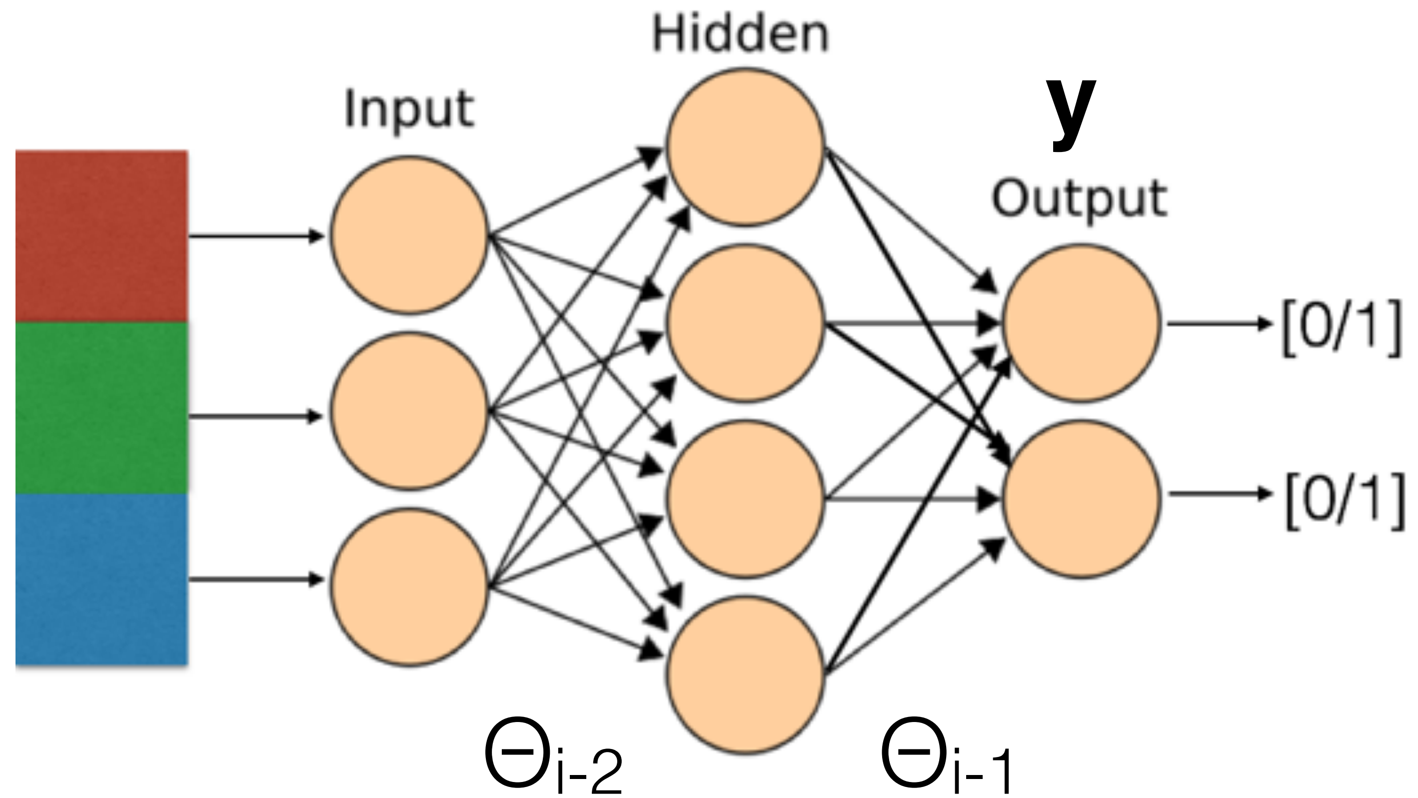
Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

Backward Pass



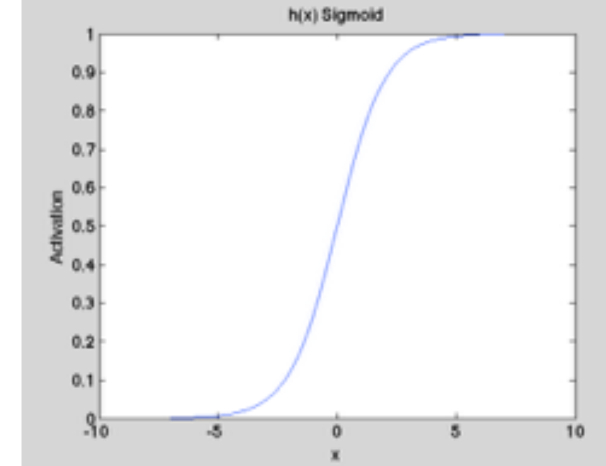
$$h(x) = \frac{1}{1+e^{-x}} \qquad \mathbf{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})$$



l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

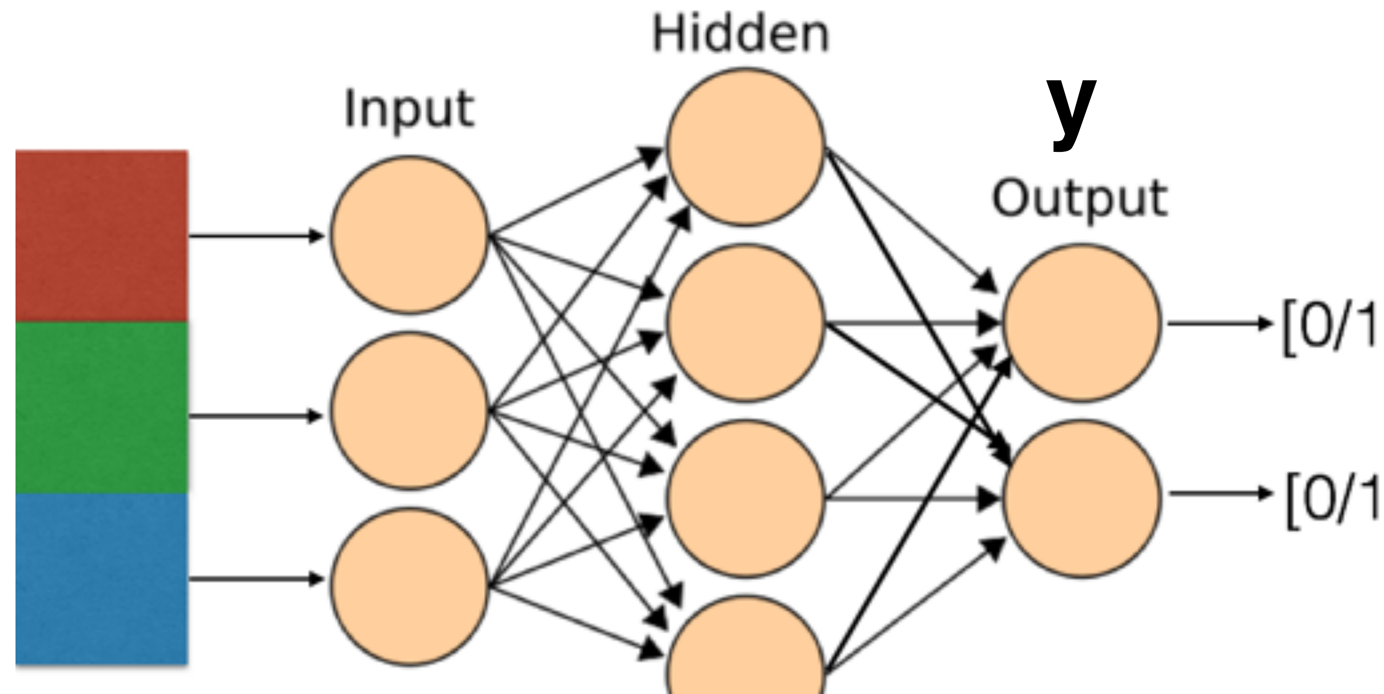
Backward Pass



$$h(x) = \frac{1}{1+e^{-x}} \qquad \mathbf{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y}$$

Last Layer

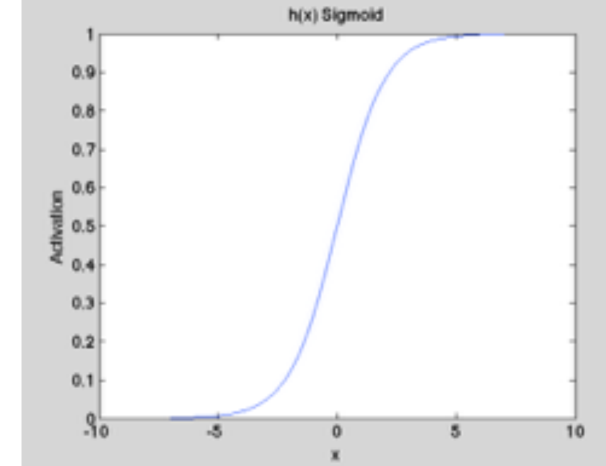


Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

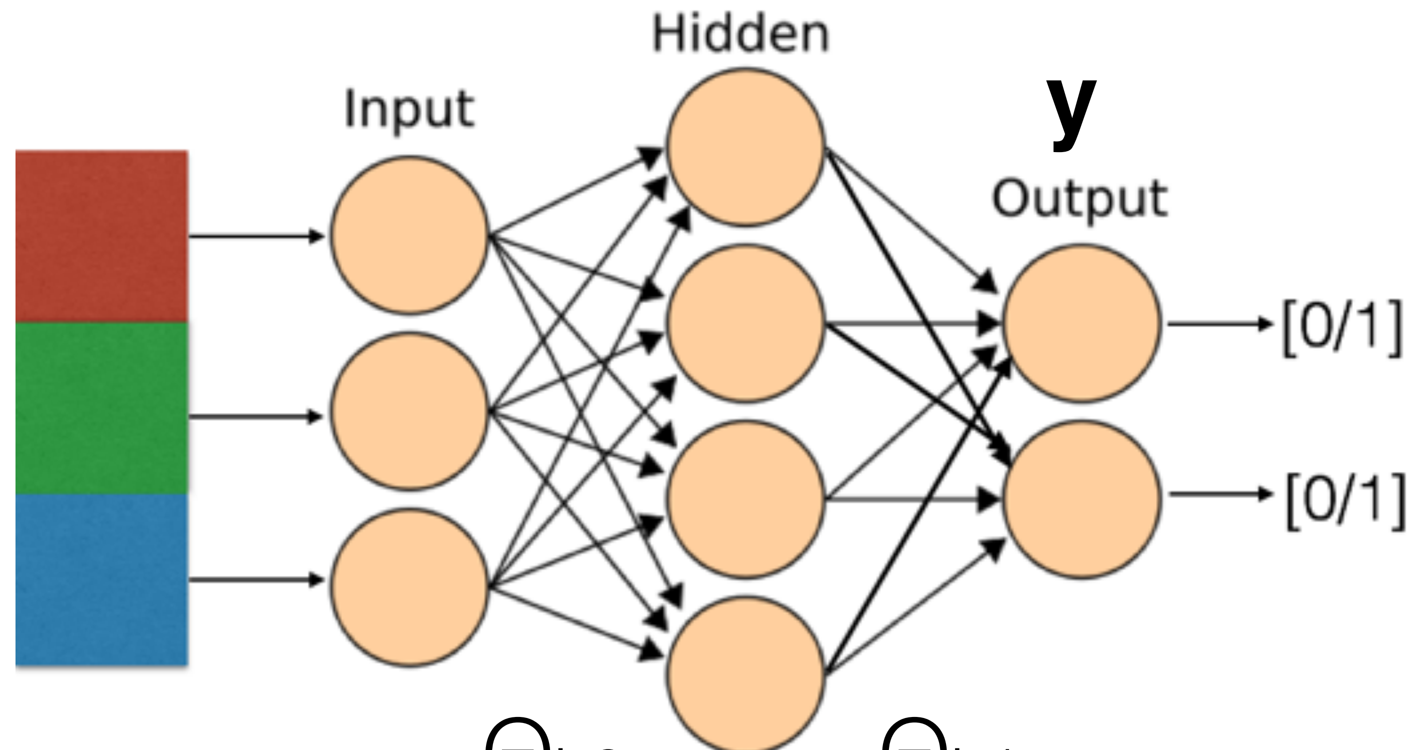
Backward Pass



$$h(x) = \frac{1}{1+e^{-x}} \qquad \mathbf{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y} = E$$

Last Layer



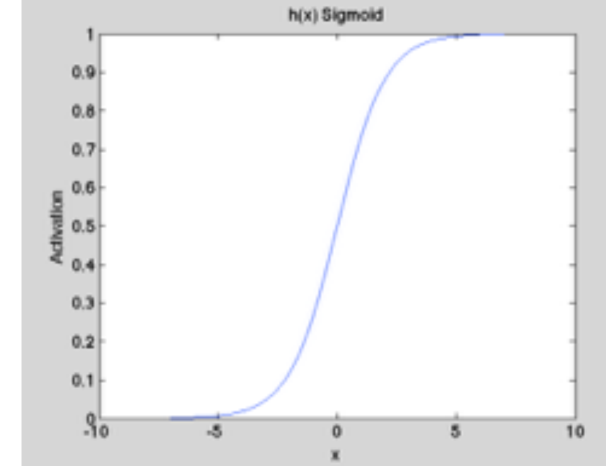
$\Theta_{i\text{-}2}$ $\Theta_{i\text{-}1}$

Outer Product

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

Backward Pass

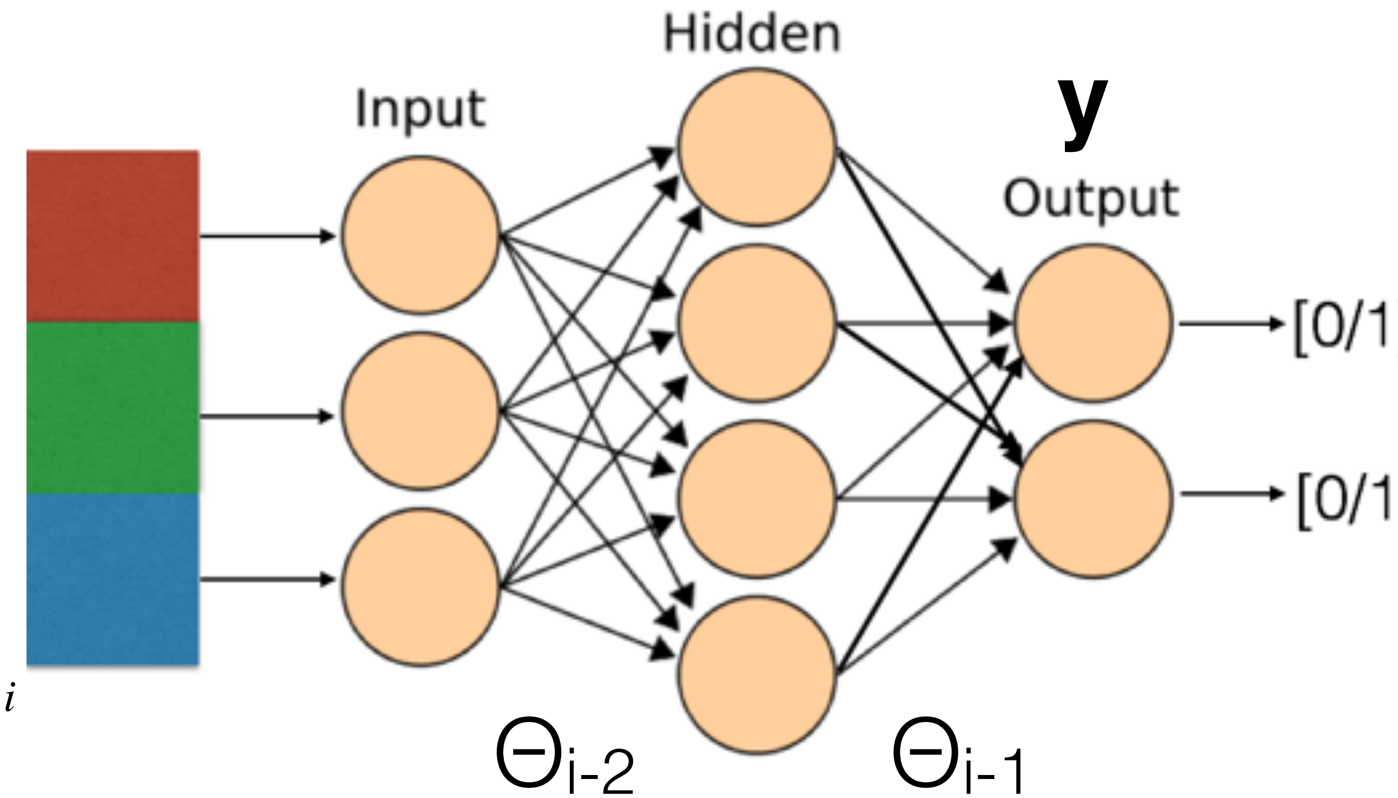


$$h(x) = \frac{1}{1+e^{-x}} \qquad \boldsymbol{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y} = E$$

Last Layer

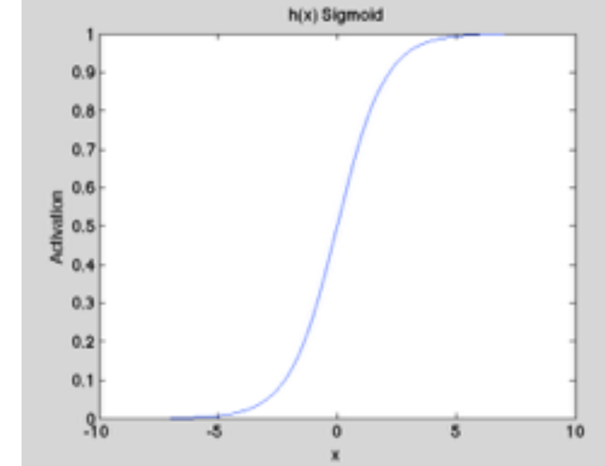
$$\boldsymbol{\delta}_{i-1} = \frac{\partial E}{\partial a_{i-1}} = \frac{\partial E}{\partial a_i}\frac{\partial a_i}{\partial a_{i-1}} = h'(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})\boldsymbol{\Theta}_{i-1}\boldsymbol{\delta}_i$$



l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\boldsymbol{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

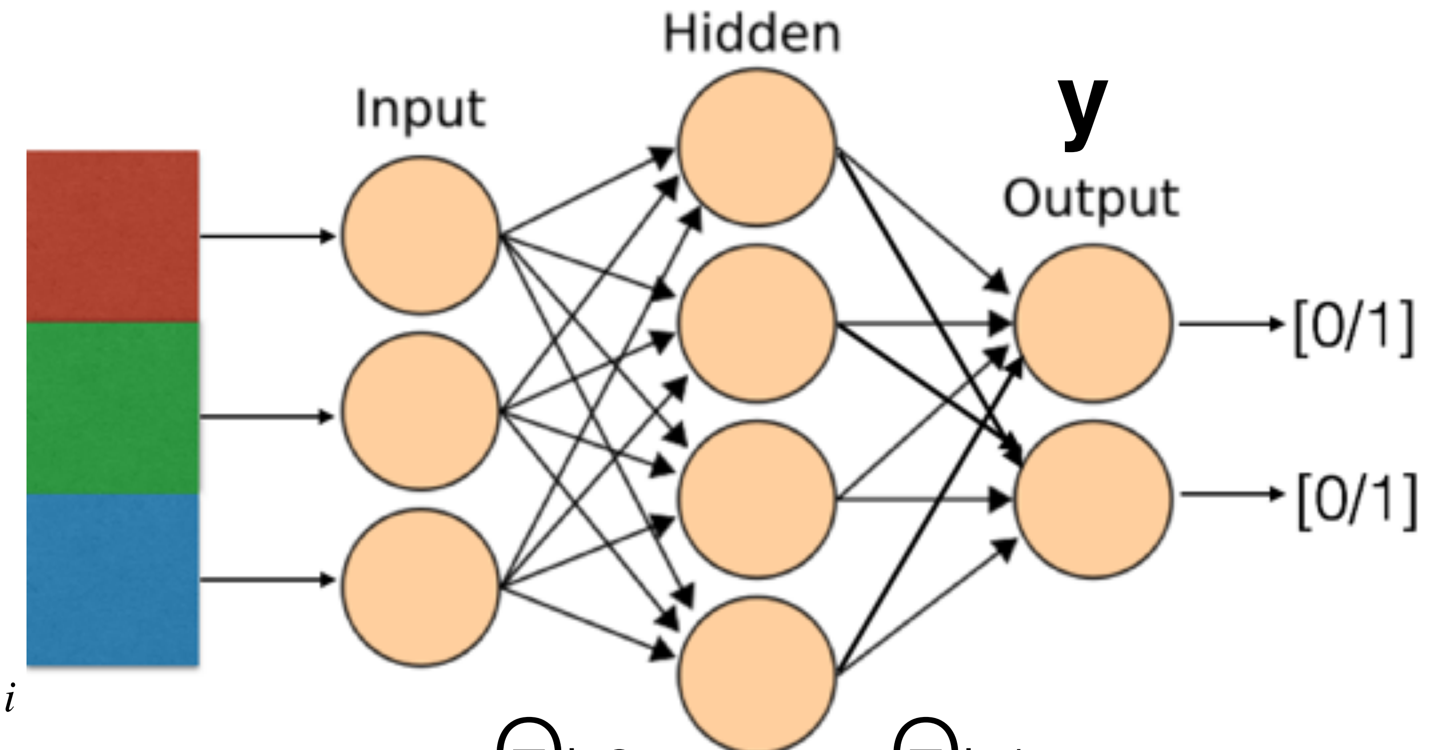
Backward Pass



$$h(x) = \frac{1}{1+e^{-x}} \qquad \boldsymbol{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y} = E$$

Last Layer



$$\boldsymbol{\delta}_{i-1} = \frac{\partial E}{\partial a_{i-1}} = \frac{\partial E}{\partial a_i}\frac{\partial a_i}{\partial a_{i-1}} = h'(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})\boldsymbol{\Theta}_{i-1}\boldsymbol{\delta}_i$$

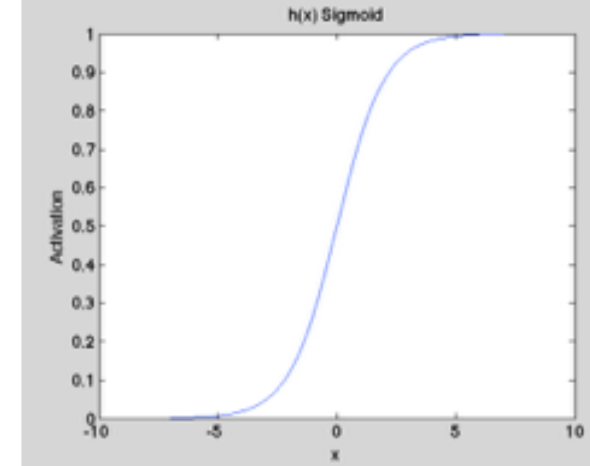
$$\frac{\partial E}{\partial \Theta_{i-2}} = \frac{\partial E}{\partial a_{i-1}}\frac{\partial a_{i-1}}{\partial \Theta_{i-2}} = h'(\mathbf{a}_{i-2}\boldsymbol{\Theta}_{i-2})\boldsymbol{\Theta}_{i-1}\boldsymbol{\delta}_i h(\mathbf{a}_{i-3}\boldsymbol{\Theta}_{i-3}) = h(\mathbf{a}_{i-3}\boldsymbol{\Theta}_{i-3})\boldsymbol{\delta}_{i-1}$$

Θ_{i-2} Θ_{i-1}

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\boldsymbol{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

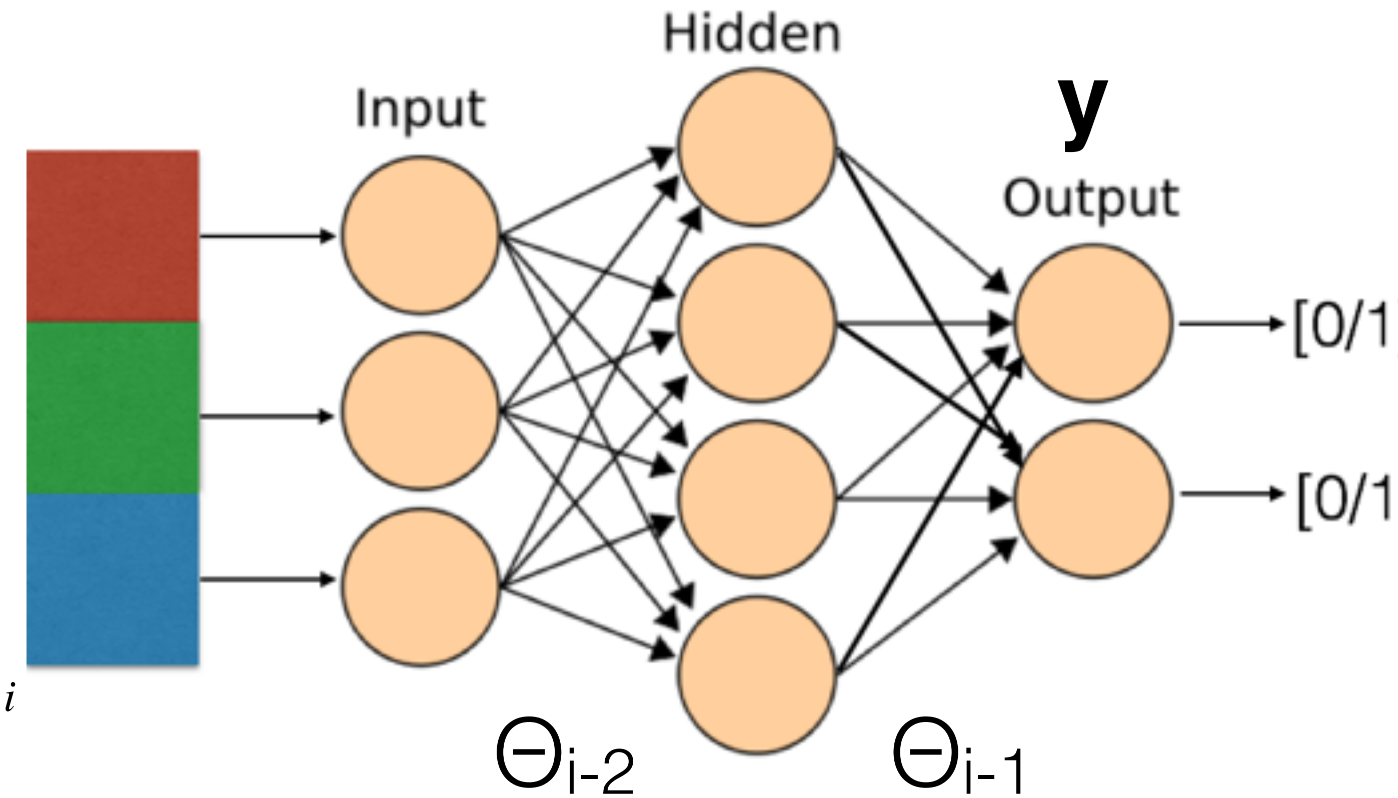
Backward Pass



$$h(x) = \frac{1}{1+e^{-x}} \qquad \mathbf{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y} = E$$

Last Layer



$$\boldsymbol{\delta}_{i-1} = \frac{\partial E}{\partial a_{i-1}} = \frac{\partial E}{\partial a_i}\frac{\partial a_i}{\partial a_{i-1}} = h'(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})\mathbf{\Theta}_{i-1}\boldsymbol{\delta}_i$$

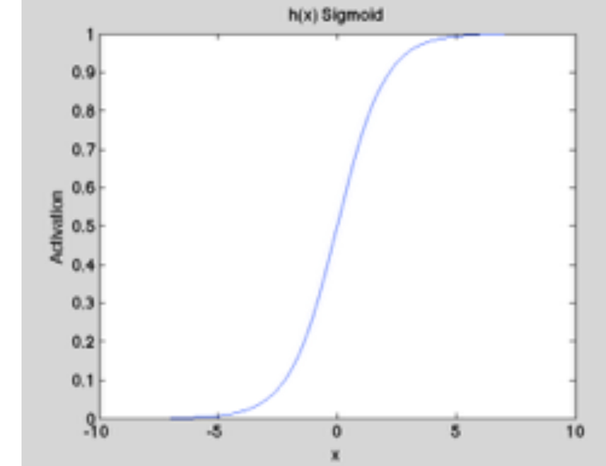
Outer Product

$$\frac{\partial E}{\partial \Theta_{i-2}} = \frac{\partial E}{\partial a_{i-1}}\frac{\partial a_{i-1}}{\partial \Theta_{i-2}} = h'(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})\mathbf{\Theta}_{i-1}\boldsymbol{\delta}_i h(\mathbf{a}_{i-3}\mathbf{\Theta}_{i-3}) = h(\mathbf{a}_{i-3}\mathbf{\Theta}_{i-3})\boldsymbol{\delta}_{i-1}$$

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

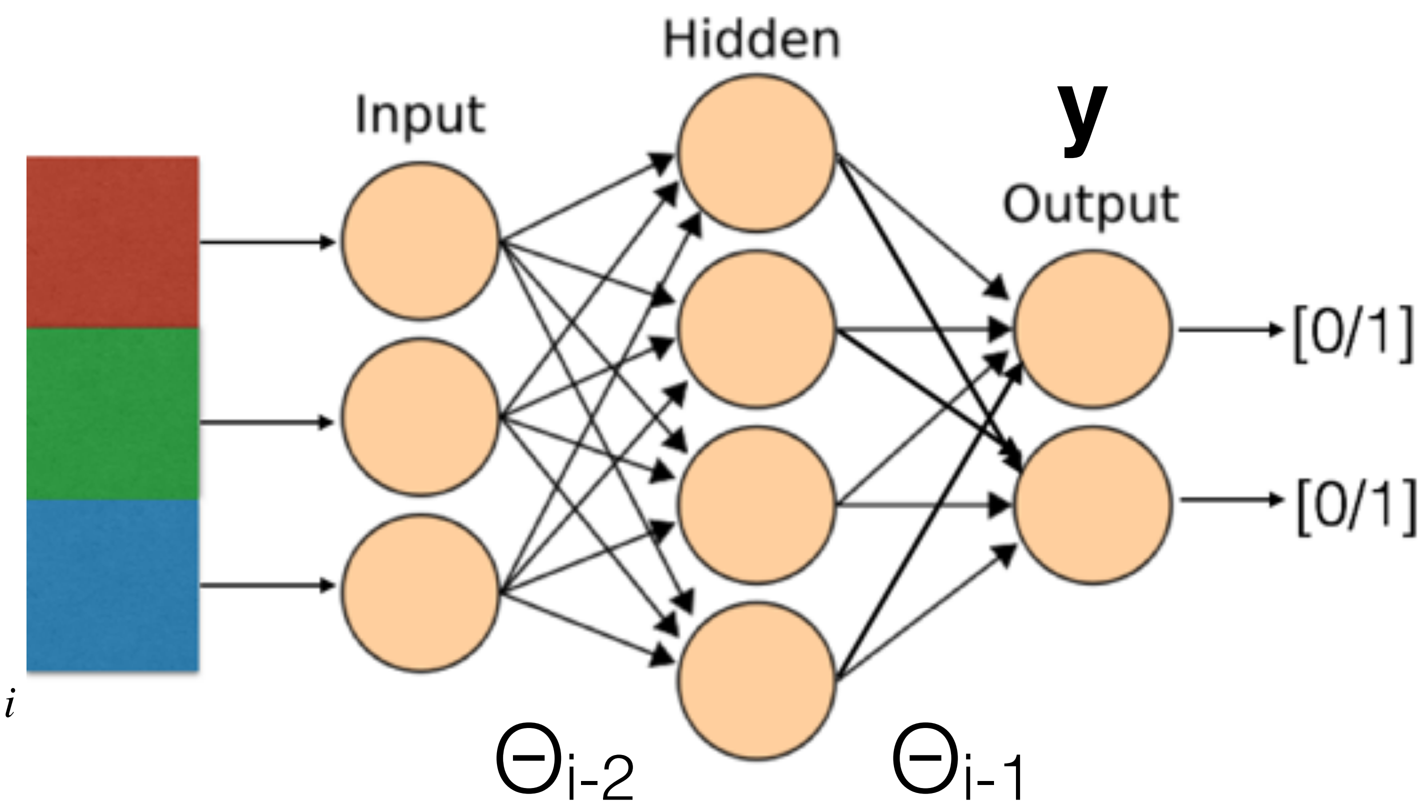
$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \qquad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

Backward Pass



$$h(x) = \frac{1}{1+e^{-x}} \qquad \mathbf{\Theta}_{i-1} = h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})$$

$$\boldsymbol{\delta}_i = \mathbf{a}_i - \mathbf{y} = E$$ Last Layer



$$\boldsymbol{\delta}_{i-1} = \frac{\partial E}{\partial a_{i-1}} = \frac{\partial E}{\partial a_i}\frac{\partial a_i}{\partial a_{i-1}} = h'(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})\mathbf{\Theta}_{i-1}\boldsymbol{\delta}_i$$

Outer Product

$$\frac{\partial E}{\partial \Theta_{i-2}} = \frac{\partial E}{\partial a_{i-1}}\frac{\partial a_{i-1}}{\partial \Theta_{i-2}} = h'(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})\mathbf{\Theta}_{i-1}\boldsymbol{\delta}_i h(\mathbf{a}_{i-3}\mathbf{\Theta}_{i-3}) = h(\mathbf{a}_{i-3}\mathbf{\Theta}_{i-3})\boldsymbol{\delta}_{i-1}$$

$$\mathbf{\Theta}_{i-1} \coloneqq \mathbf{\Theta}_{i-1} - \alpha h(\mathbf{a}_{i-2}\mathbf{\Theta}_{i-2})\boldsymbol{\delta}_i$$

l_{i-2}	l_{i-1}	l_i
k_{i-2}	k_{i-1}	k_i
$\mathbf{a}_{i-2}$	$\mathbf{a}_{i-1}$	$\mathbf{a}_i$

$$\mathbf{\Theta}_{i-2} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k_{i-1}} \\ \vdots & \ddots & \vdots \\ \theta_{k_{i-2}1} & \cdots & \theta_{k_{i-2}k_{i-1}} \end{bmatrix} \quad \mathbf{a}_{i-2} = \begin{bmatrix} a_1 & \cdots & a_{k_{i-2}} \end{bmatrix}$$

Practical Advice

- Start with a solid foundation. Vectorize everything you can, but do not try to vectorize over the training sample loop.
- To avoid overfitting, check with a validation set in-between iterations of backpropagation
- Language of choice: MATLAB
- Try a general implementation
- MNIST Dataset
- It is more difficult to implement a committee of machines than to just make a bunch of neural networks due to error distribution.

Thank You

MNIST Dataset: 86%