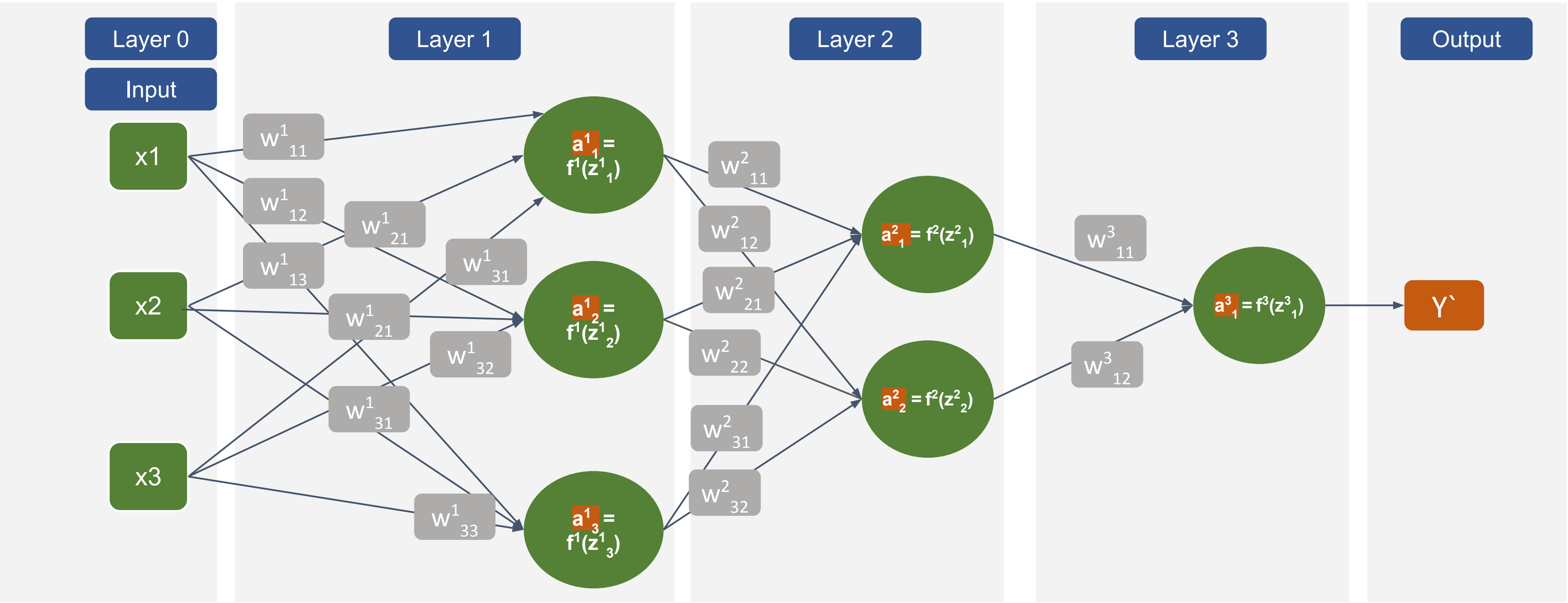
Deep Learning

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/



Layers L = 3

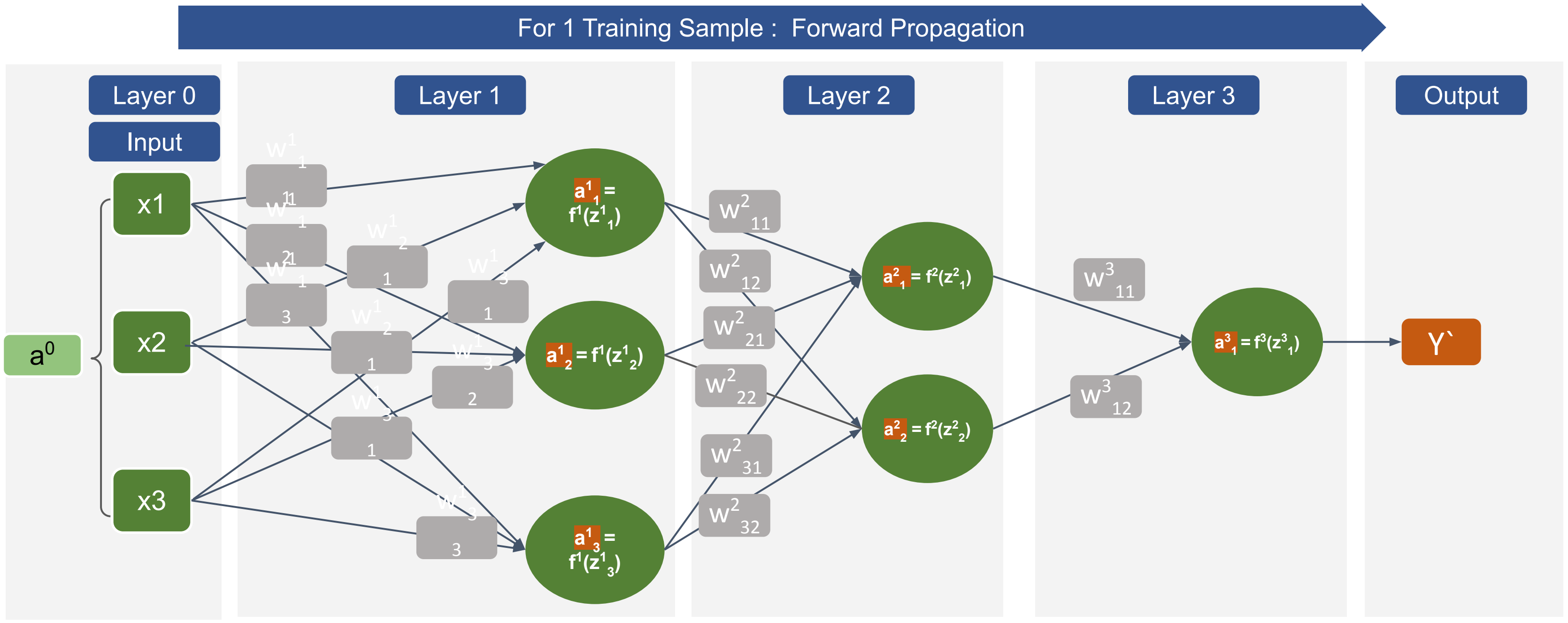
$n^{[L]}$ = # units in layer L $n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$b^{[L]}$ = biases for layer L

Arun Kumar A
analaarun.k@gmail.com
in https://www.linkedin.com/in/arun-kumar-anala-35760523/



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

$z^1 = w^1 x + b^1$

$a^1 = f^1 (z^1)$

$z^2 = w^2 a^1 + b^2$

$a^2 = f^2 (z^2)$

$z^3 = w^3 a^2 + b^3$

$a^3 = f^3 (z^3)$

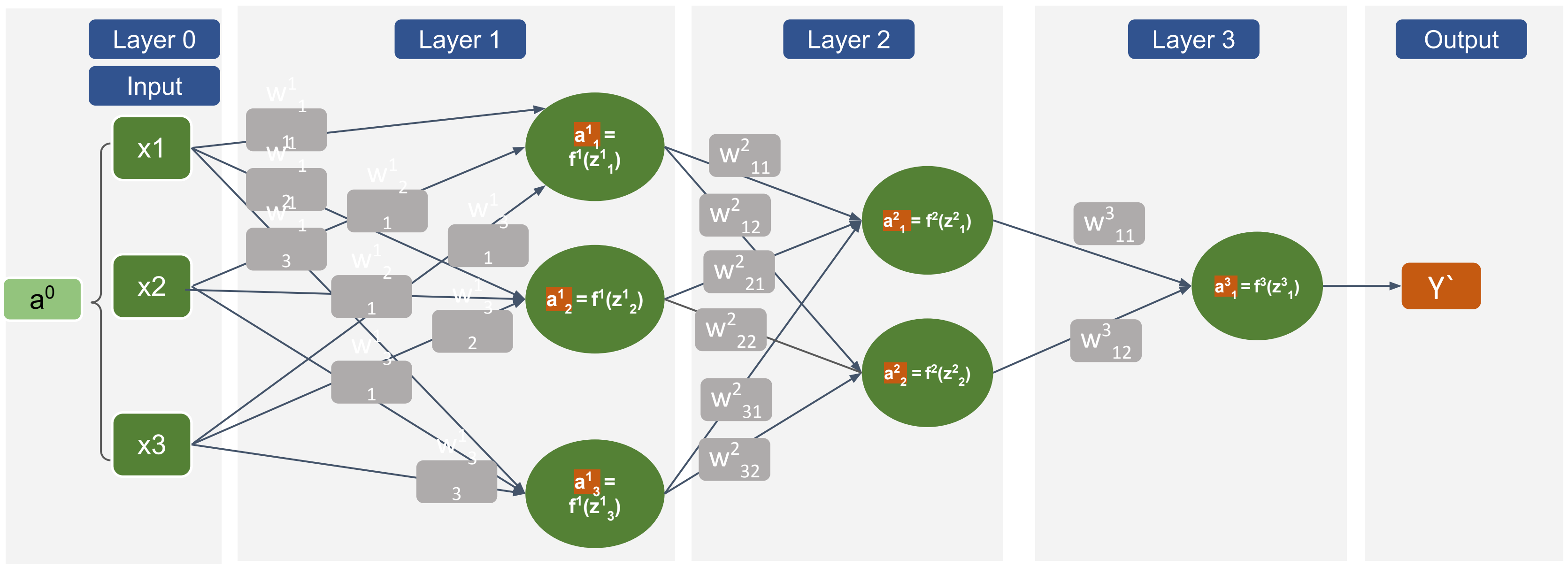
Y` $= a^3$

$$z^{[L]} = w^{[L]} a^{[L-1]} + b^{[L]}$$

$$a^{[L]} = f^{[L]} (z^{[L]})$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample : Forward Propagation



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

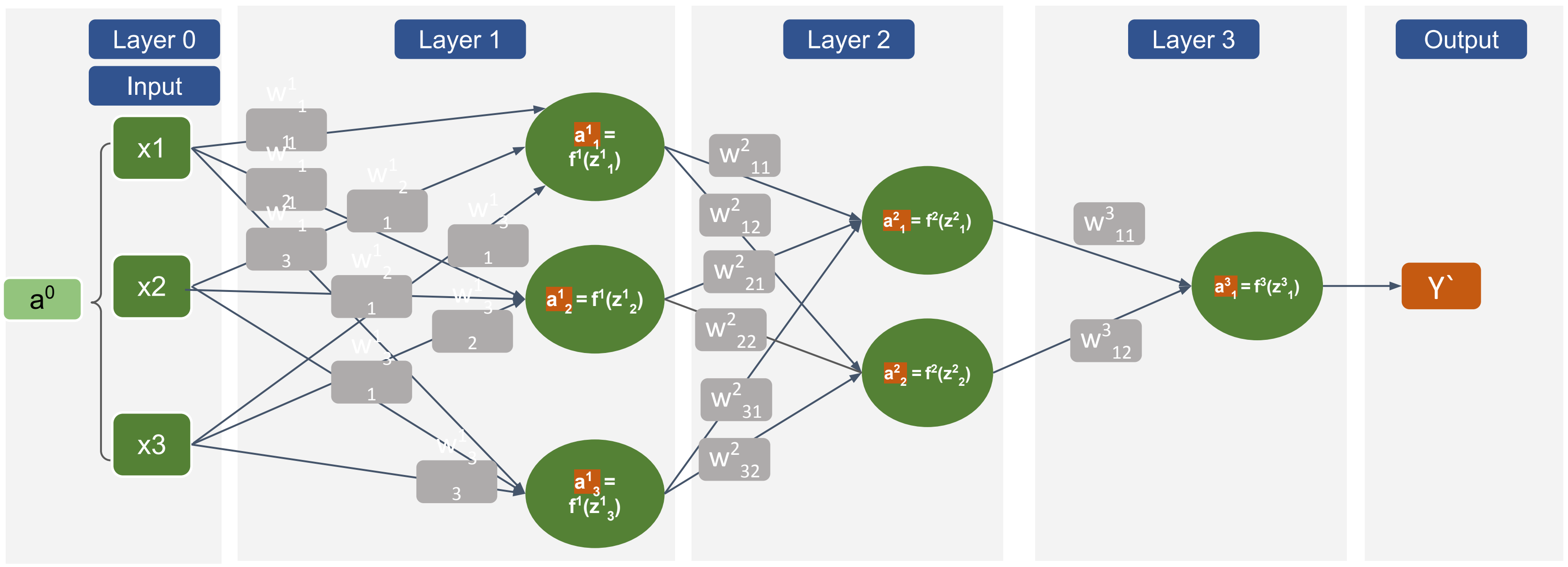
$b^{[L]}$ = biases for layer L

$$z^1 = w^1\ x + b^1$$

$$\begin{bmatrix} z^1{}_1 \\ z^1{}_2 \\ z^1{}_3 \end{bmatrix} = \begin{bmatrix} w^1{}_{11} & w^1{}_{21} & w^1{}_{31} \\ w^1{}_{12} & w^1{}_{22} & w^1{}_{32} \\ w^1{}_{13} & w^1{}_{23} & w^1{}_{33} \end{bmatrix} \times \begin{bmatrix} a^0{}_1 \\ a^0{}_2 \\ a^0{}_3 \end{bmatrix} + \begin{bmatrix} b^1{}_1 \\ b^1{}_2 \\ b^1{}_3 \end{bmatrix}$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample : Forward Propagation



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

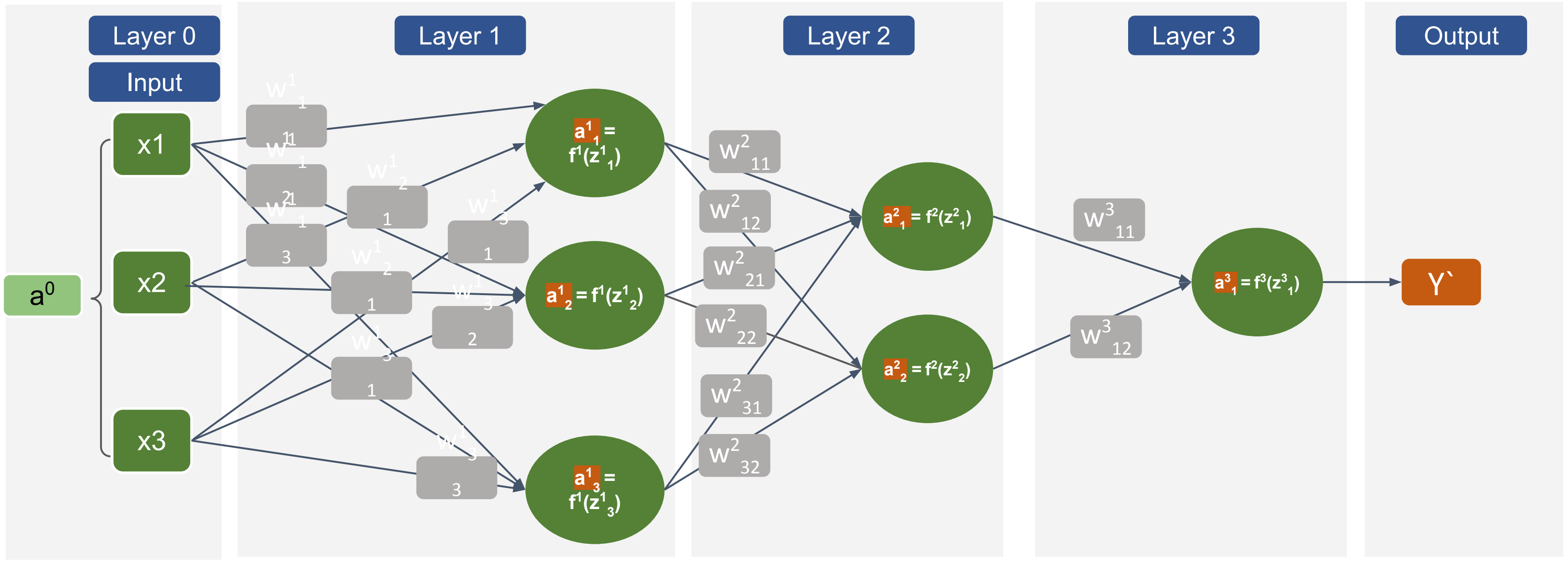
$z^1 = w^1\ x + b^1 \qquad z^1 = w^1\ a^0 + b^1$

$$\begin{bmatrix} \boxed{z^1_1} \\ z^1_2 \\ z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 \\ a^0_2 \\ a^0_3 \end{bmatrix} + \begin{bmatrix} \boxed{b^1_1} \\ b^1_2 \\ b^1_3 \end{bmatrix}$$

$z^1_1 = w^1_{11}\, a^0_1 + w^1_{21}\, a^0_2 + w^1_{31}\, a^0_3 + b^1_1$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample : Forward Propagation



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

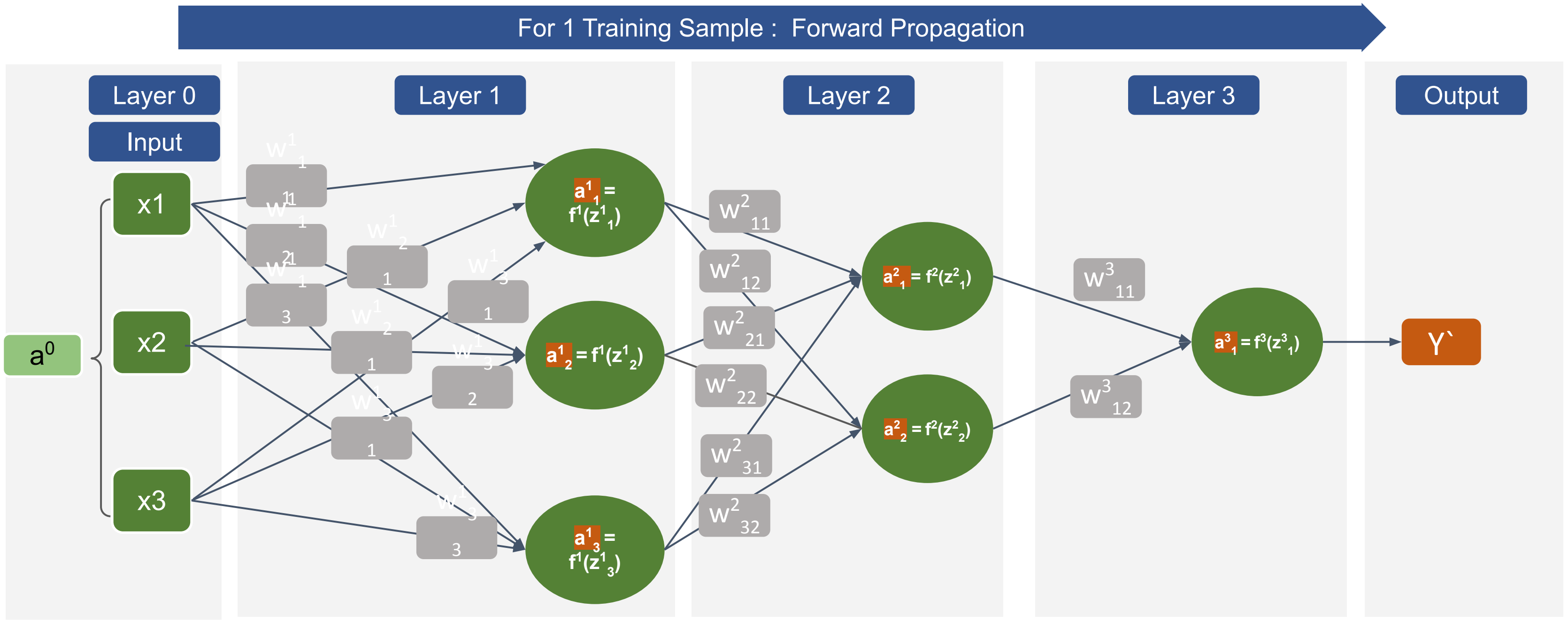
$b^{[L]}$ = biases for layer L

$$z^1 = w^1\, x + b^1 \qquad z^1 = w^1\, a^0 + b^1$$

$$\begin{bmatrix} z^1_1 \\ \boxed{z^1_2} \\ z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 \\ a^0_2 \\ a^0_3 \end{bmatrix} + \begin{bmatrix} b^1_1 \\ \boxed{b^1_2} \\ b^1_3 \end{bmatrix}$$

$$z^1_2 = w^1_{12}\, a^0_1 + w^1_{22}\, a^0_2 + w^1_{32}\, a^0_3 + b^1_2$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

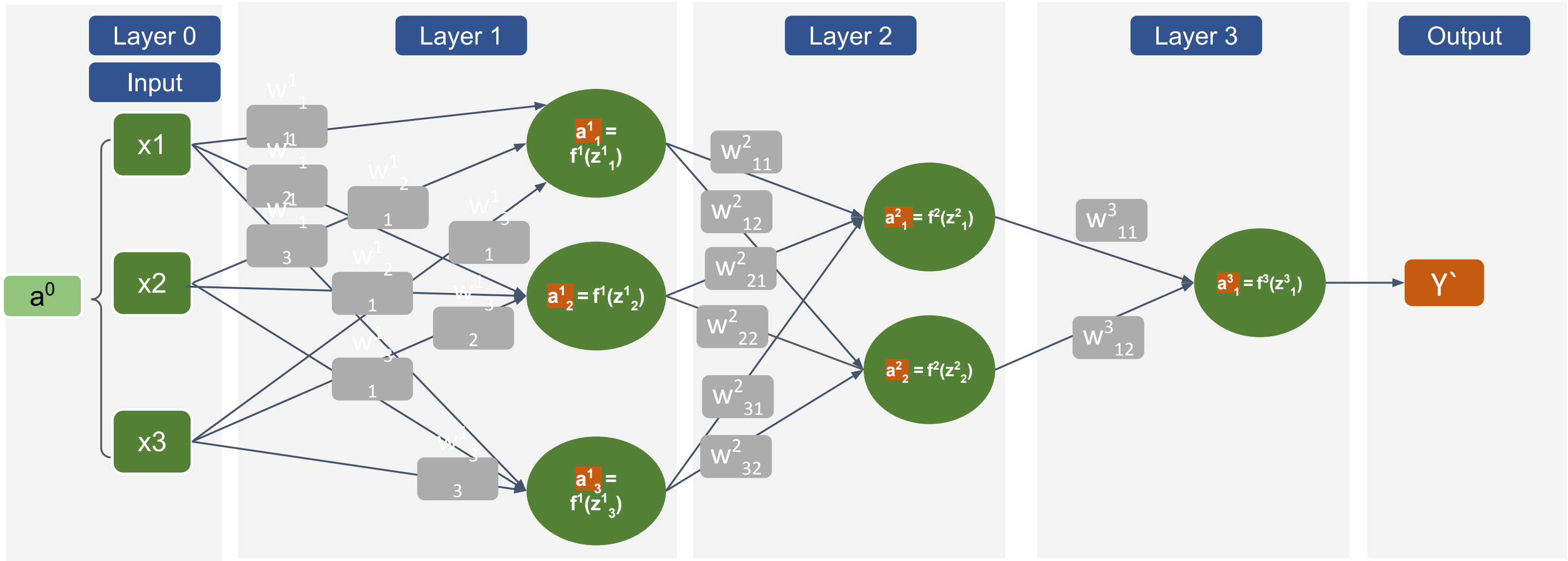
$z^1 = w^1 x + b^1 \qquad z^1 = w^1 a^0 + b^1$

$$\begin{bmatrix} z^1_1 \\ z^1_2 \\ \boxed{z^1_3} \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ \boxed{w^1_{13} \quad w^1_{23} \quad w^1_{33}} \end{bmatrix} \times \boxed{\begin{bmatrix} a^0_1 \\ a^0_2 \\ a^0_3 \end{bmatrix}} + \begin{bmatrix} b^1_1 \\ b^1_2 \\ \boxed{b^1_3} \end{bmatrix}$$

$z^1_3 = w^1_{13} a^0_1 + w^1_{23} a^0_2 + w^1_{33} a^0_3 + b^1_3$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample : Forward Propagation



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

$$z^1 = w^1\, x + b^1 \qquad z^1 = w^1\, a^0 + b^1$$

$$\begin{bmatrix} z^1_1 \\ z^1_2 \\ \boxed{z^1_3} \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 \\ a^0_2 \\ a^0_3 \end{bmatrix} + \begin{bmatrix} b^1_1 \\ b^1_2 \\ \boxed{b^1_3} \end{bmatrix}$$

$$\begin{bmatrix} a^1_1 \\ a^1_2 \\ a^1_3 \end{bmatrix} = f^1 \begin{bmatrix} z^1_1 \\ z^1_2 \\ z^1_3 \end{bmatrix}$$

$$z^2 = w^2\, a^1 + b^2$$

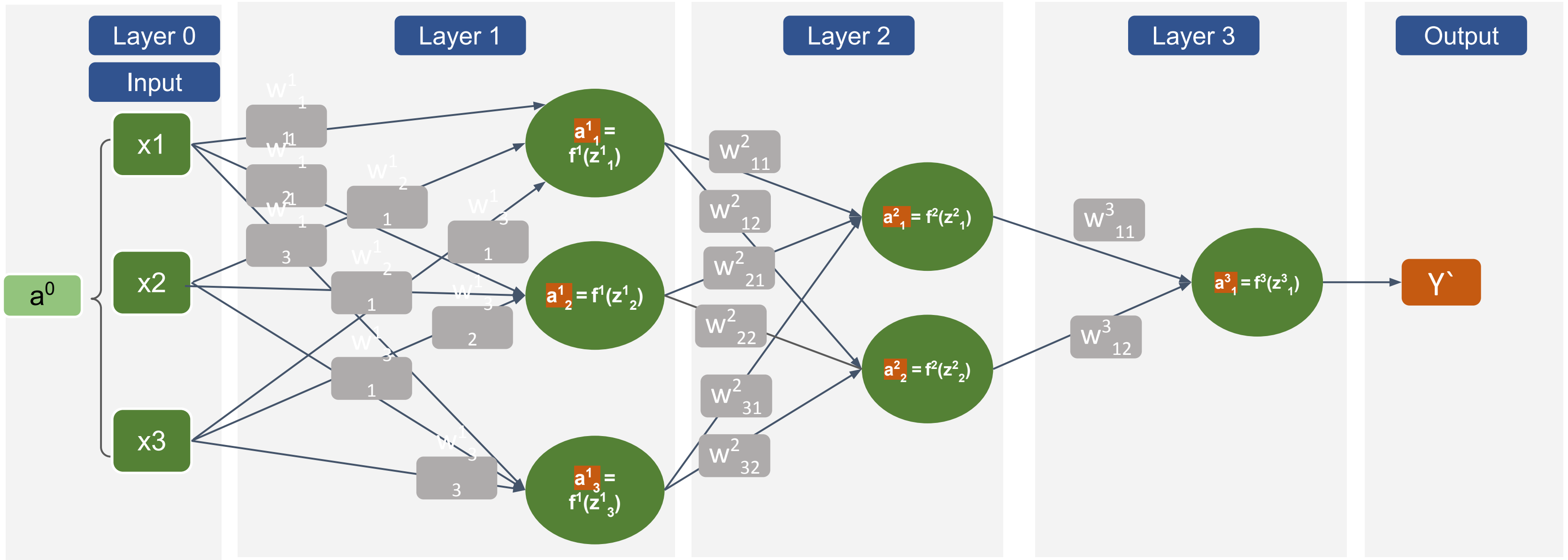
$$\begin{bmatrix} \boxed{z^2_1} \\ z^2_2 \end{bmatrix} = \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \times \begin{bmatrix} a^1_1 \\ a^1_2 \\ a^1_3 \end{bmatrix} + \begin{bmatrix} \boxed{b^2_1} \\ b^2_2 \end{bmatrix}$$

$$z^2_1 = w^2_{11}\, a^1_1 + w^2_{21}\, a^1_2 + w^2_{31}\, a^1_3 + b^2_1$$

$$z^2_2 = w^2_{12}\, a^1_1 + w^2_{22}\, a^1_2 + w^2_{32}\, a^1_3 + b^2_2$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample : Forward Propagation



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

Layer 1

$$z^1 = w^1\, x + b^1 \qquad z^1 = w^1\, a^0 + b^1$$

$$\begin{bmatrix} z^1_1 \\ z^1_2 \\ z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 \\ a^0_2 \\ a^0_3 \end{bmatrix} + \begin{bmatrix} b^1_1 \\ b^1_2 \\ b^1_3 \end{bmatrix}$$

(3,1) (3,3) (3,1) (3,1)

Layer 2

$$z^2 = w^2\, a^1 + b^2$$

$$\begin{bmatrix} z^2_1 \\ z^2_2 \end{bmatrix} = \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \times \begin{bmatrix} a^1_1 \\ a^1_2 \\ a^1_3 \end{bmatrix} + \begin{bmatrix} b^2_1 \\ b^2_2 \end{bmatrix}$$

(2,1) (2,3) (3,1) (2,1)

$w^{[L]} : (n^{[L]} , n^{[L-1]})$

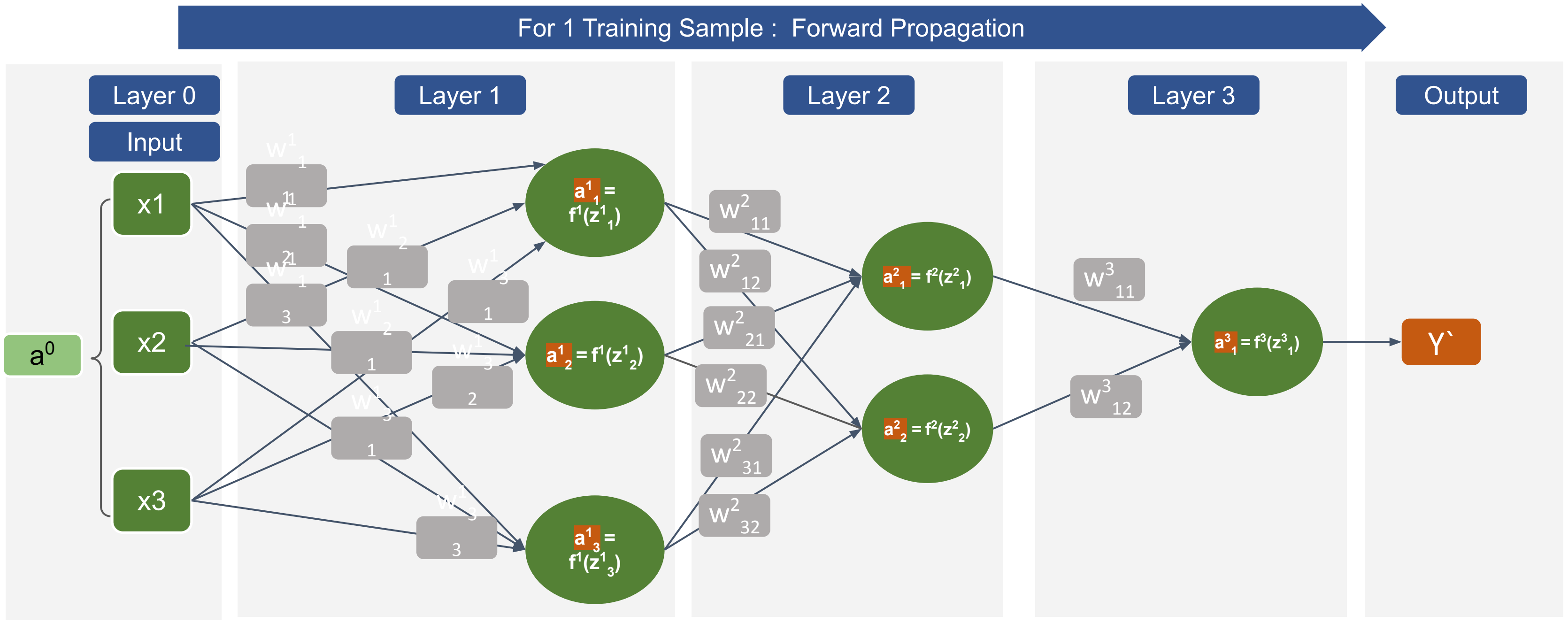
$b^{[L]} : (n^{[L]} , 1)$

$w^{[1]} : (3 , 3)$ $b^{[1]} : (3 , 1)$

$w^{[2]} : (2 , 3)$ $b^{[2]} : (2 , 1)$

$w^{[3]} : (1 , 2)$ $b^{[3]} : (1 , 1)$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

$b^{[L]}$ = biases for layer L

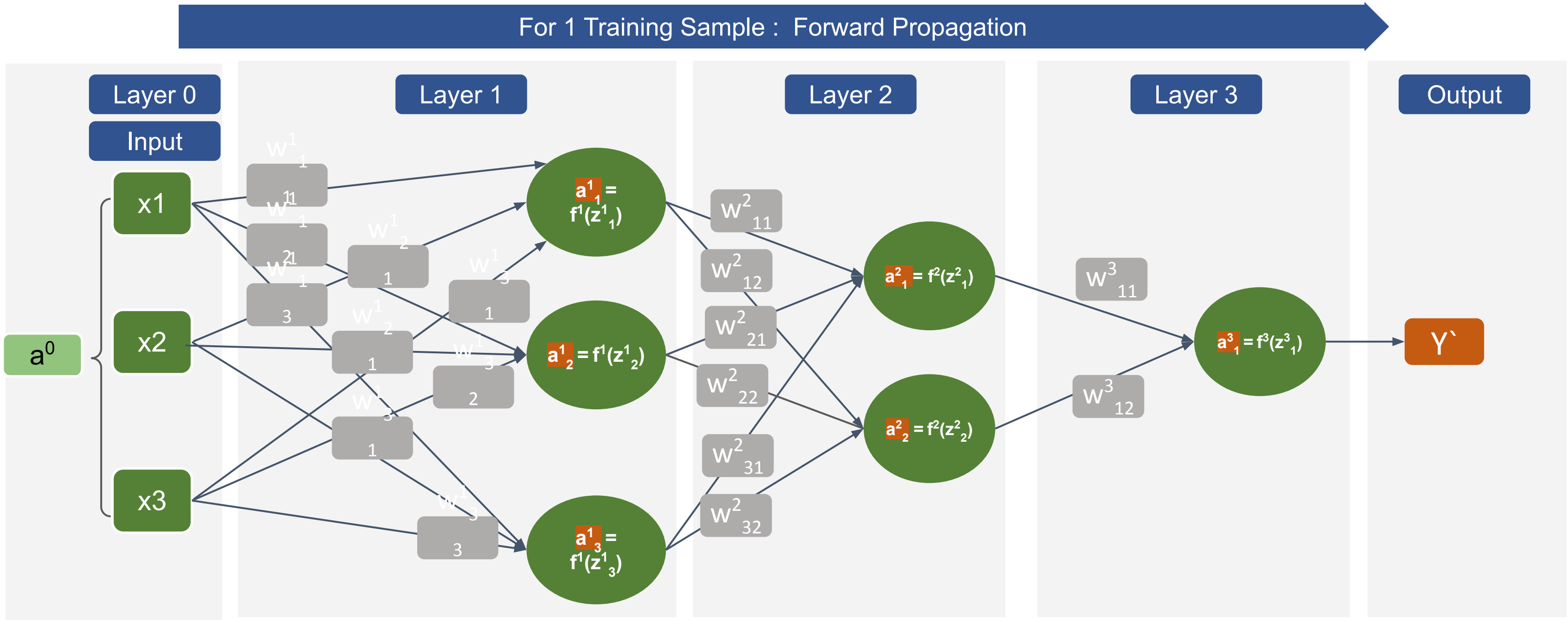
$$z^{[L]} = w^{[L]} a^{[L-1]} + b^{[L]}$$

$$a^{[L]} = f^{[L]} (z^{[L]})$$

$$w^{[L]} : (n^{[L]} , n^{[L-1]})$$

$$b^{[L]} : (n^{[L]} , 1)$$

Arun Kumar A
analaarun.k@gmail.com
in https://www.linkedin.com/in/arun-kumar-anala-35760523/



Layers L = 3

$n^{[L]}$ = # units in layer L

$a^{[L]}$ = # activations in layer L = $f^{[L]}(z^{[L]})$

$n^{[1]} = 3$, $n^{[2]} = 2$, $n^{[3]} = 1$

$w^{[L]}$ = weights for layer L

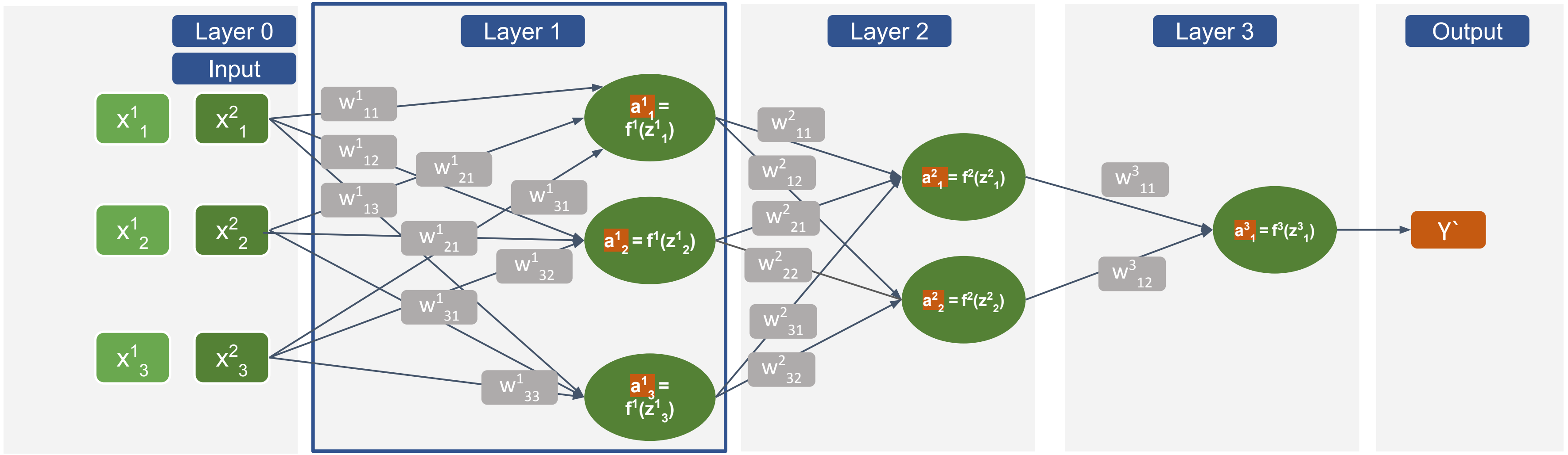
$b^{[L]}$ = biases for layer L

$$z^{[L]} = w^{[L]} a^{[L-1]} + b^{[L]}$$

$$a^{[L]} = f^{[L]}(z^{[L]})$$

$z^{[L]} : (n^{[L]} , 1)$	$\partial z^{[L]} : (n^{[L]} , 1)$
$a^{[L]} : (n^{[L]} , 1)$	$\partial a^{[L]} : (n^{[L]} , 1)$
$w^{[L]} : (n^{[L]} , n^{[L-1]})$	$\partial w^{[L]} : (n^{[L]} , n^{[L-1]})$
$b^{[L]} : (n^{[L]} , 1)$	$\partial b^{[L]} : (n^{[L]} , 1)$

For 2 Training Sample : Forward Propagation

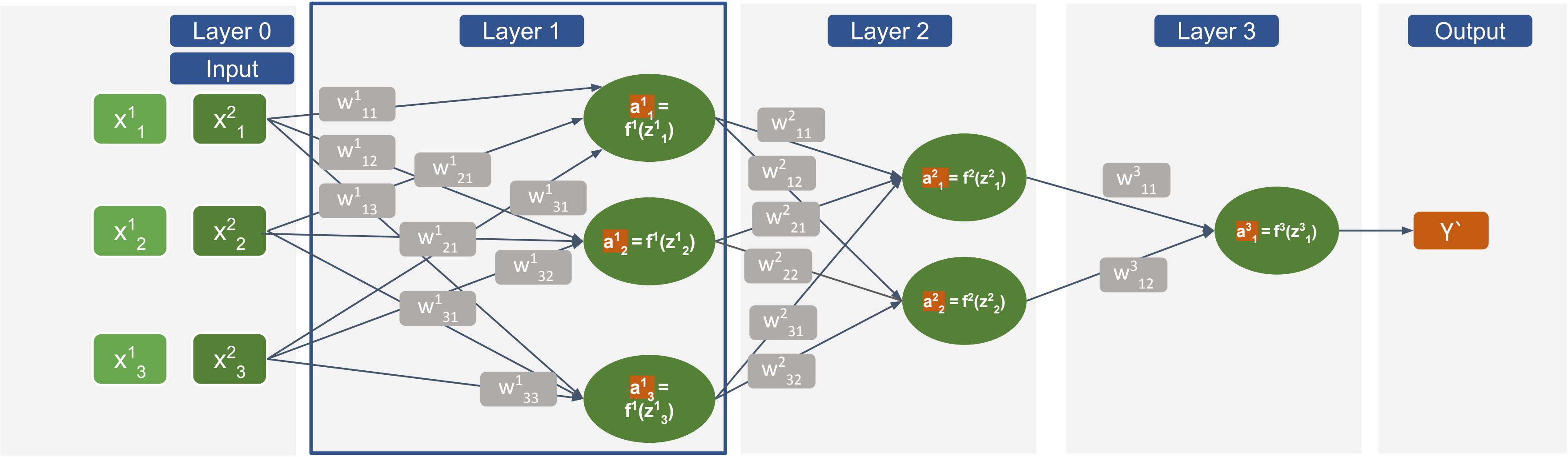


$$
\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}
=
\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}
\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}
\times
\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}
+
\begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}
$$

(3,2) ((3, 3) , 2) (3,2) (3,2)

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

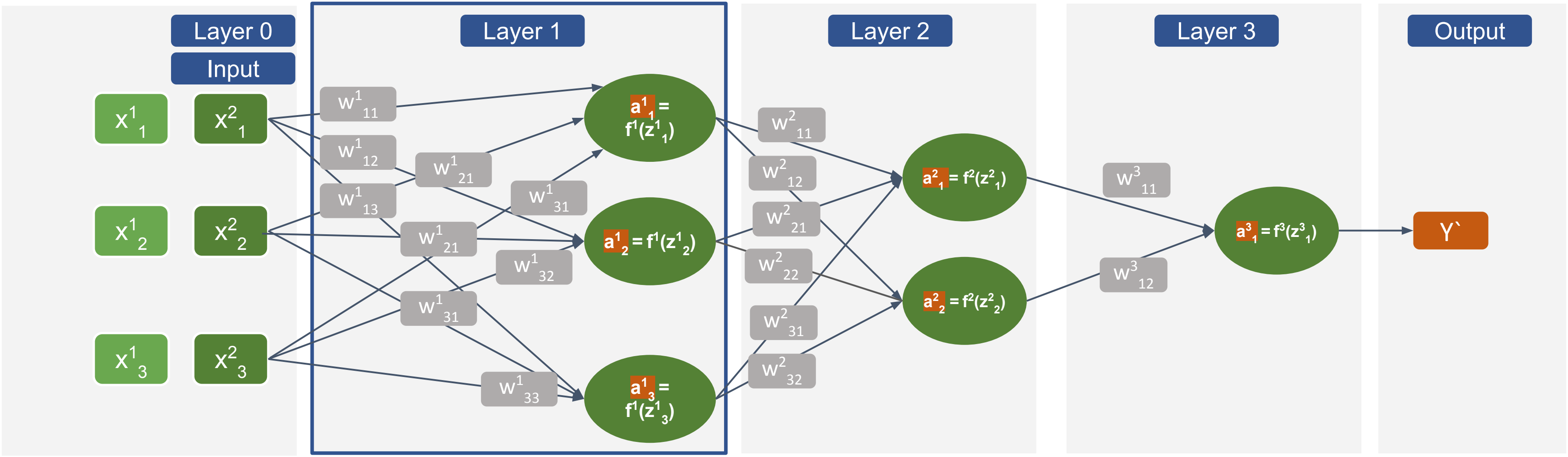


$$\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix} + \begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}$$

(3,2) ((3, 3) , 2) (3,2) (3,2)

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

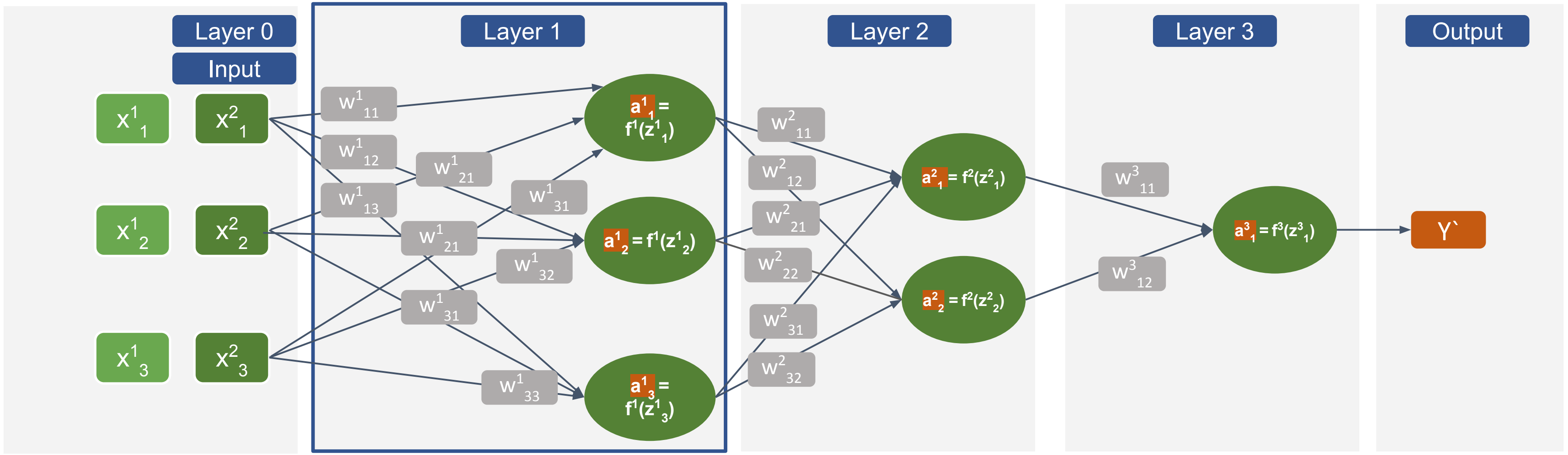


$$\underbrace{\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}}_{(3,2)} = \underbrace{\boxed{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}}_{(3,\ 3)} \times \underbrace{\boxed{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}}}_{(3,2)} + \underbrace{\begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}}_{(3,2)}$$

$$\underbrace{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}_{(2,\ (3,\ 3))}$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation



$$\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \boxed{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}} + \begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}$$

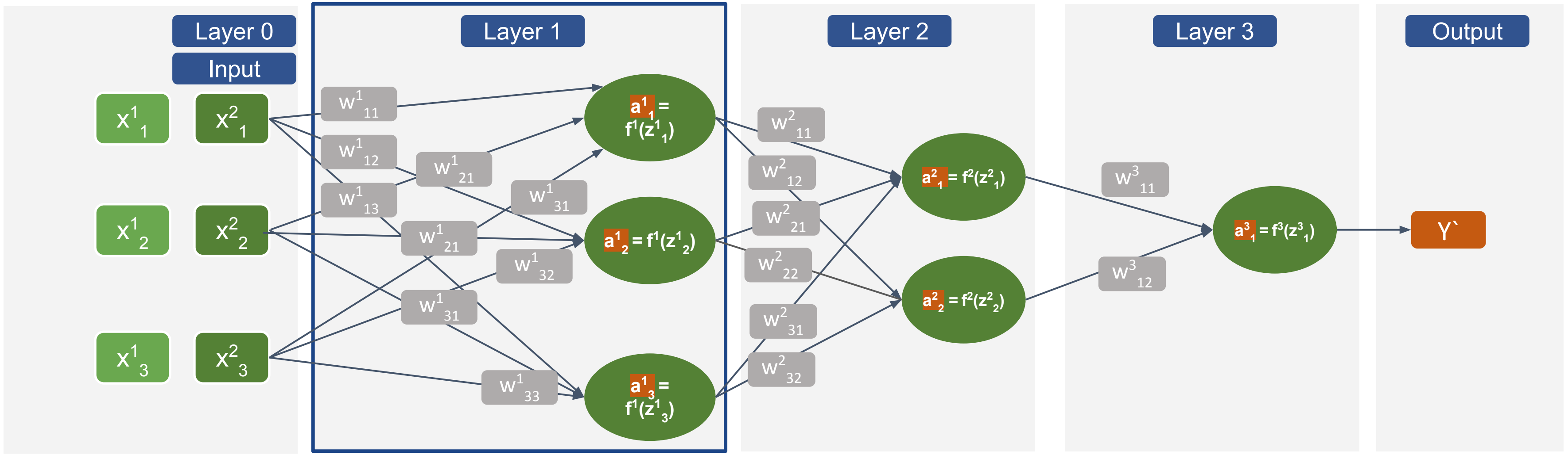
(3,2) (3, 3) (3,2) (3,2)

$$\boxed{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}$$

(2, (3, 3))

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

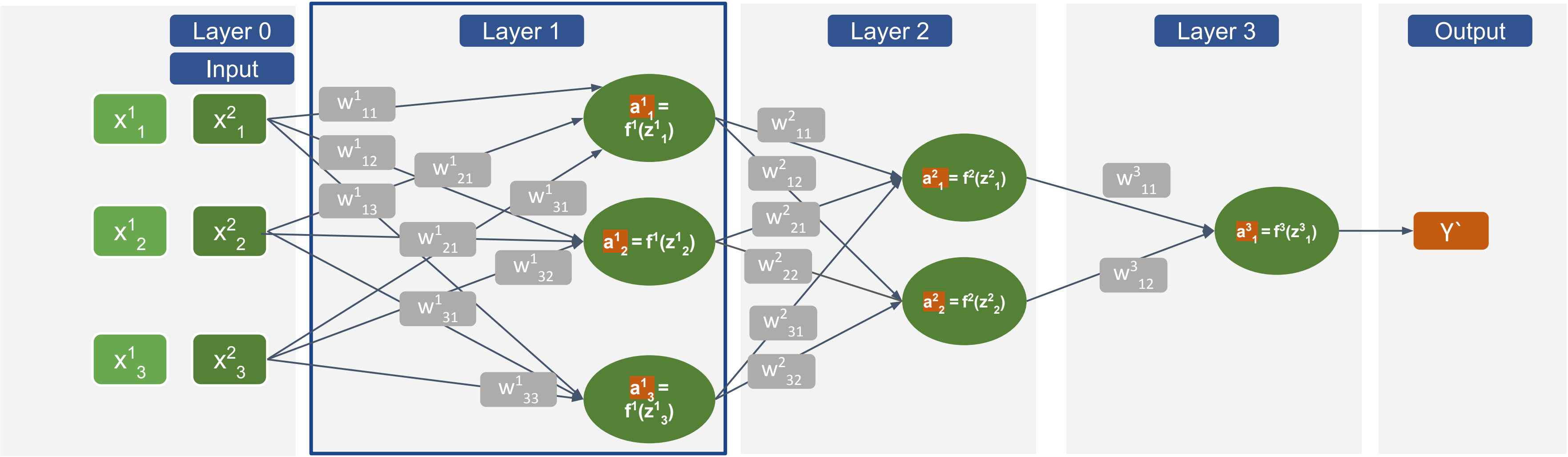
For 2 Training Sample : Forward Propagation



$$
\underset{(3,2)}{\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}} = \underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}} \times \underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}} + \underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}} \times \underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}} + \underset{(3,2)}{\begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}}
$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

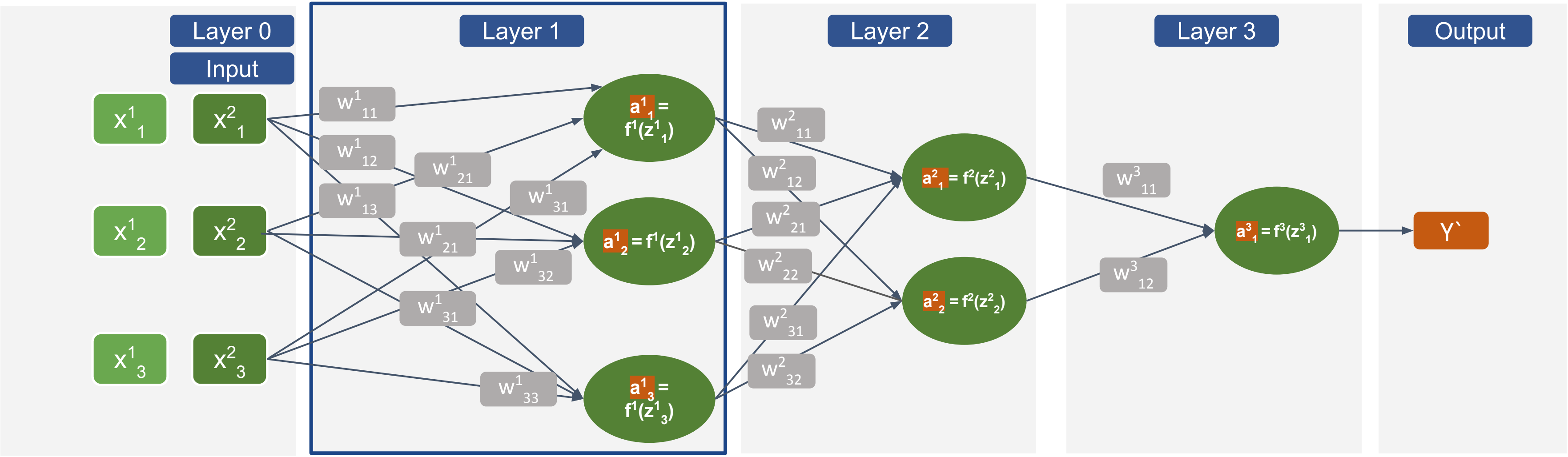
For 2 Training Sample : Forward Propagation



$$\underbrace{\begin{bmatrix} z^1_1 & z^1_1 \\ \boxed{z^1_2 \quad z^1_2} \\ z^1_3 & z^1_3 \end{bmatrix}}_{(3,2)} = \underbrace{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ \boxed{w^1_{12} \quad w^1_{22} \quad w^1_{32}} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}_{(3,3)} \times \underbrace{\boxed{\begin{matrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{matrix}}}_{(3,2)} + \underbrace{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ \boxed{w^1_{12} \quad w^1_{22} \quad w^1_{32}} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}_{(3,3)} \times \underbrace{\boxed{\begin{matrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{matrix}}}_{(3,2)} + \underbrace{\begin{bmatrix} b^1_1 & b^1_1 \\ \boxed{b^1_2 \quad b^1_2} \\ b^1_3 & b^1_3 \end{bmatrix}}_{(3,2)}$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

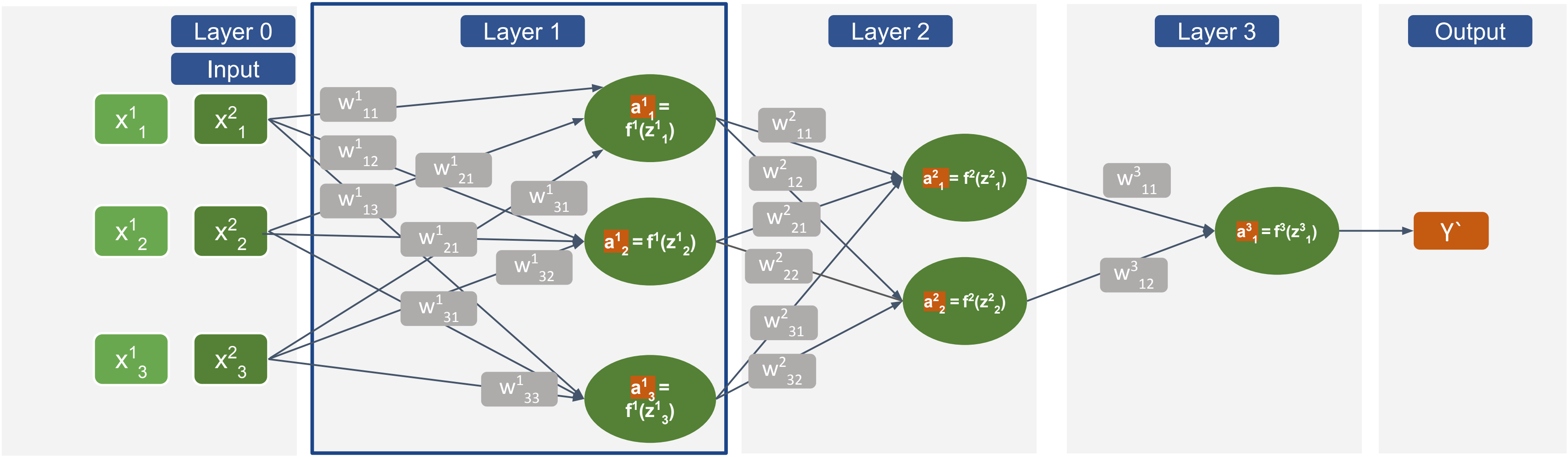
For 2 Training Sample : Forward Propagation



$$
\underset{(3,2)}{\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ \boxed{z^1_3} & \boxed{z^1_3} \end{bmatrix}}
=
\underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}
\ \mathrm{X}\
\underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}}
+
\underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}
\ \mathrm{X}\
\underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}}
+
\underset{(3,2)}{\begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}}
$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

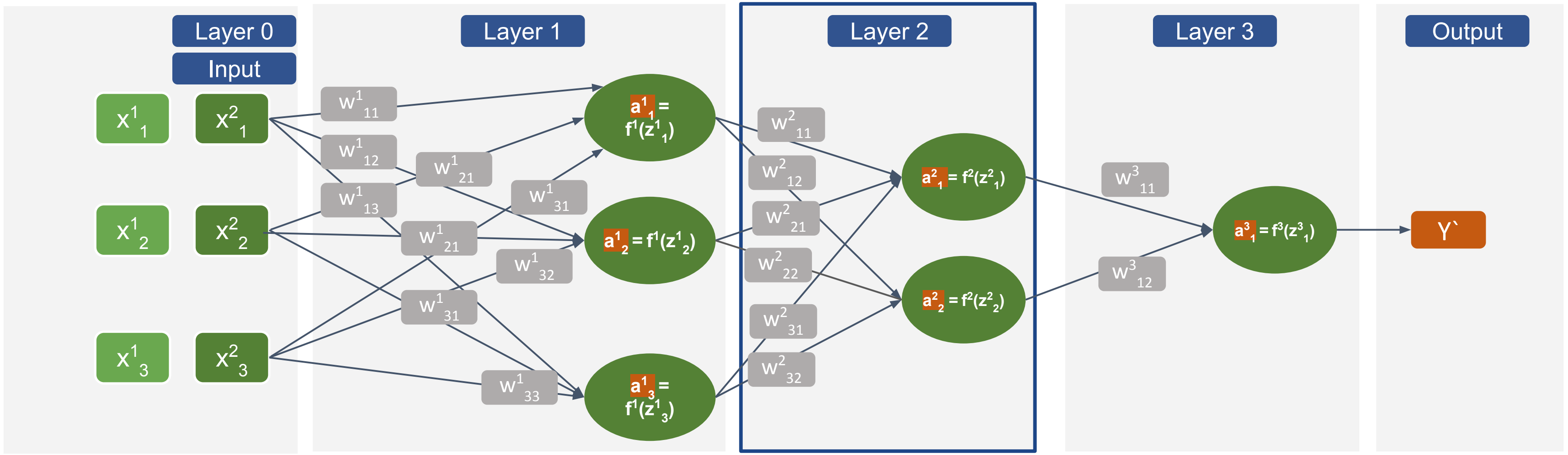


$$
\underset{(3,2)}{\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}}
=
\underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}
\times
\underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}}
+
\underset{(3,\ 3)}{\begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix}}
\times
\underset{(3,2)}{\begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix}}
+
\underset{(3,2)}{\begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}}
$$

$$
\underset{(3,2)}{\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix}}
= f\left(
\underset{(3,2)}{\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}}
\right)
$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

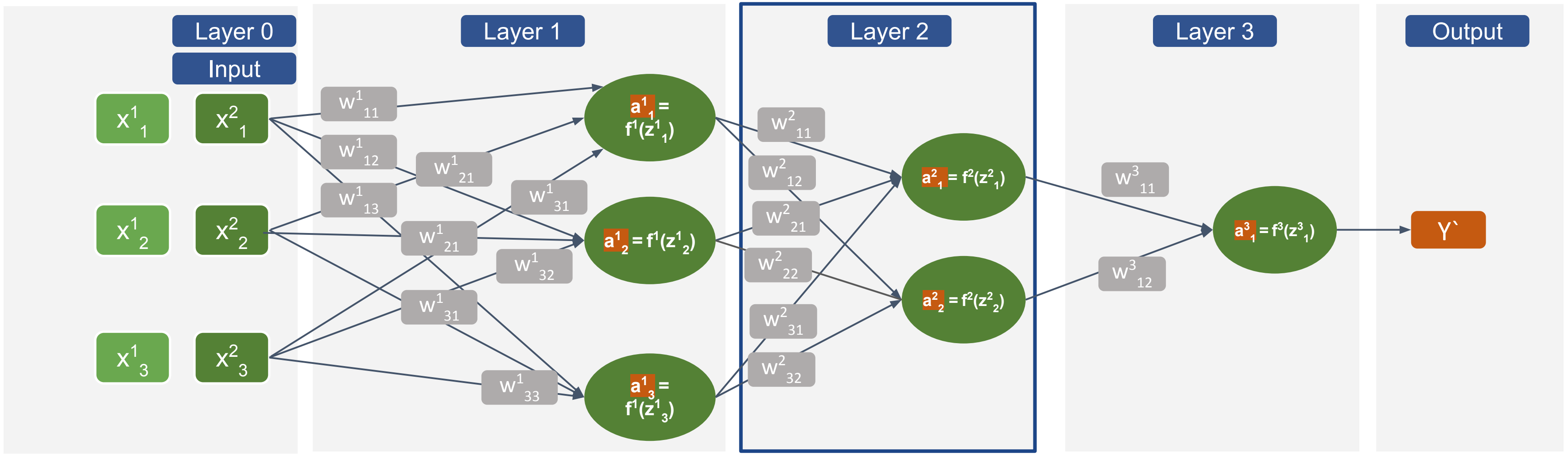


$$
\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix}
=
\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}
\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}
\times
\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix}
+
\begin{bmatrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{bmatrix}
$$

(2,2) ((2, 3) , 2) (3,2) (3,2)

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

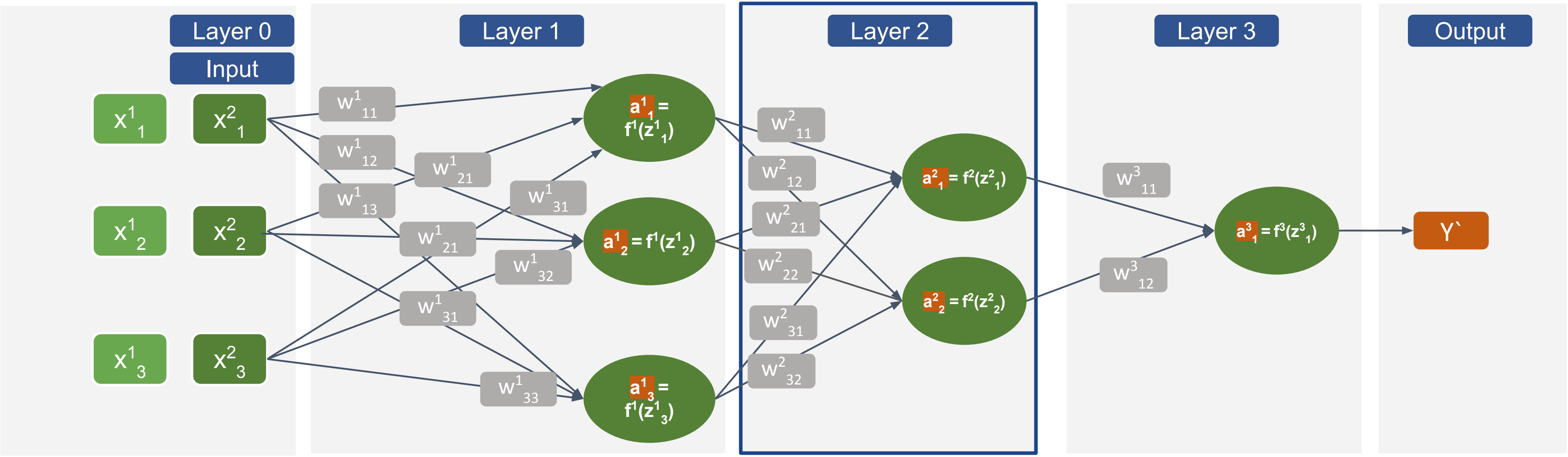


$$\underbrace{\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix}}_{(2,2)} = \underbrace{\boxed{\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}}}_{(2,3)} \times \underbrace{\boxed{\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix}}}_{(3,2)} + \underbrace{\begin{bmatrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{bmatrix}}_{(2,2)}$$

$$\underbrace{\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}}_{(2,\ (2,\ 3))}$$

Arun Kumar A
analaarun.k@gmail.com
in https://www.linkedin.com/in/arun-kumar-anala-35760523/

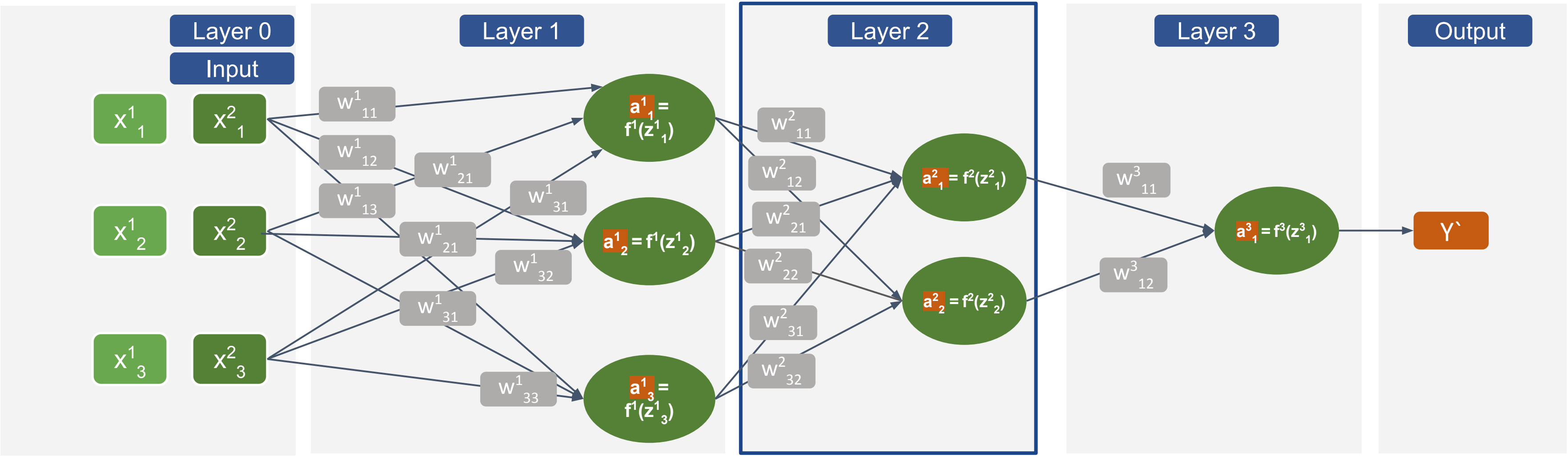
For 2 Training Sample : Forward Propagation



$$\underset{(2,2)}{\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix}} = \underset{(2,3)}{\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}} \times \underset{(3,2)}{\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix}} + \underset{(2,3)}{\begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix}} \times \underset{(3,2)}{\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix}} + \underset{(2,2)}{\begin{bmatrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{bmatrix}}$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

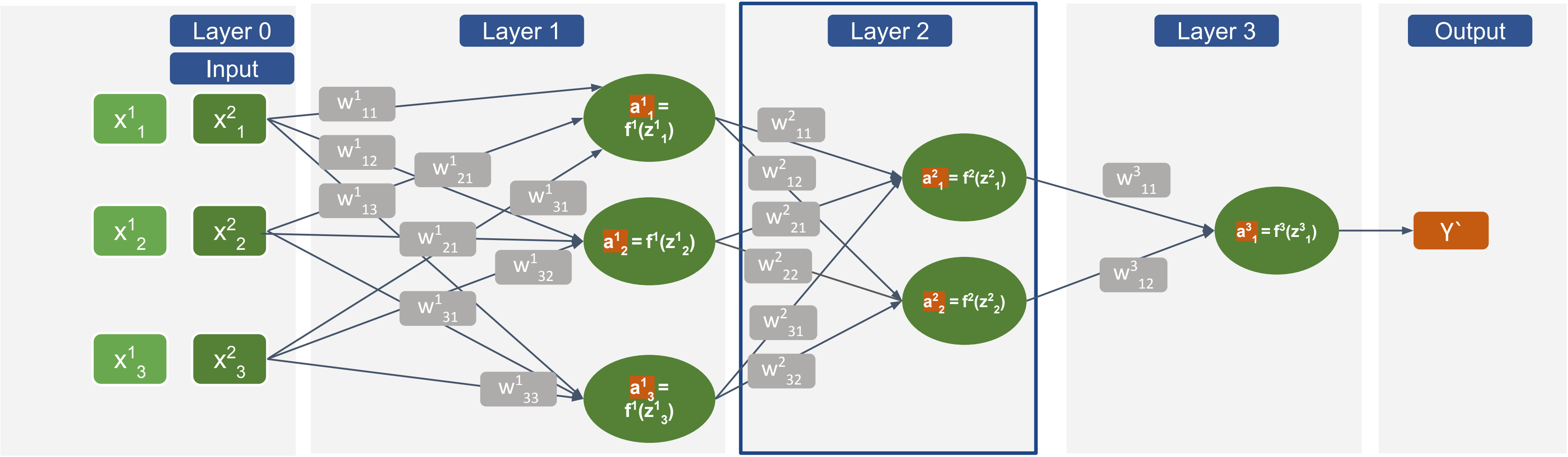
For 2 Training Sample : Forward Propagation



$$
\underset{(2,2)}{\begin{matrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{matrix}}
=
\underset{(2,3)}{\begin{matrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{matrix}}
\times
\underset{(3,2)}{\begin{matrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{matrix}}
+
\underset{(2,3)}{\begin{matrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{matrix}}
\times
\underset{(3,2)}{\begin{matrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{matrix}}
+
\underset{(2,2)}{\begin{matrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{matrix}}
$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation



$$\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix} = \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \times \begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix} + \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \times \begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix} + \begin{bmatrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{bmatrix}$$

(2,2) (2,3) (3,2) (2,3) (3,2) (2,2)

$$\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \end{bmatrix} = f\left(\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix}\right)$$

(2,2) (2,2)

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 2 Training Sample : Forward Propagation

$$\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix} = \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \begin{bmatrix} w^1_{11} & w^1_{21} & w^1_{31} \\ w^1_{12} & w^1_{22} & w^1_{32} \\ w^1_{13} & w^1_{23} & w^1_{33} \end{bmatrix} \times \begin{bmatrix} a^0_1 & a^0_1 \\ a^0_2 & a^0_2 \\ a^0_3 & a^0_3 \end{bmatrix} + \begin{bmatrix} b^1_1 & b^1_1 \\ b^1_2 & b^1_2 \\ b^1_3 & b^1_3 \end{bmatrix}$$

(3,2) ((3, 3) , 2) (3,2) (3,2)

Z^1 W^1 A^0 B^1

$$\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix} = f\left(\begin{bmatrix} z^1_1 & z^1_1 \\ z^1_2 & z^1_2 \\ z^1_3 & z^1_3 \end{bmatrix}\right)$$

(3,2) (3,2)

A^1 Z^1

$Z^1 = (W^1)^T A^0 + B^1$

$$\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix} = \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \begin{bmatrix} w^2_{11} & w^2_{21} & w^2_{31} \\ w^2_{12} & w^2_{22} & w^2_{32} \end{bmatrix} \times \begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \\ a^1_3 & a^1_3 \end{bmatrix} + \begin{bmatrix} b^2_1 & b^2_1 \\ b^2_2 & b^2_2 \end{bmatrix}$$

(2,2) ((2, 3) , 2) (3,2) (2,2)

Z^2 W^2 A^1 B^2

$$\begin{bmatrix} a^1_1 & a^1_1 \\ a^1_2 & a^1_2 \end{bmatrix} = f\left(\begin{bmatrix} z^2_1 & z^2_1 \\ z^2_2 & z^2_2 \end{bmatrix}\right)$$

(2,2) (2,2)

A^2 Z^2

$Z^2 = (W^2)^T A^1 + B^2$

$Z^{[L]} = (W^{[L]})^T A^{[L-1]} + B^{[L]}$

$A^{[L]} = f^{[L]} (Z^{[L]})$

$Z^{[L]} : (n^{[L]} , m)$

$A^{[L]} : (n^{[L]} , m)$

$\partial Z^{[L]} : (n^{[L]} , m)$

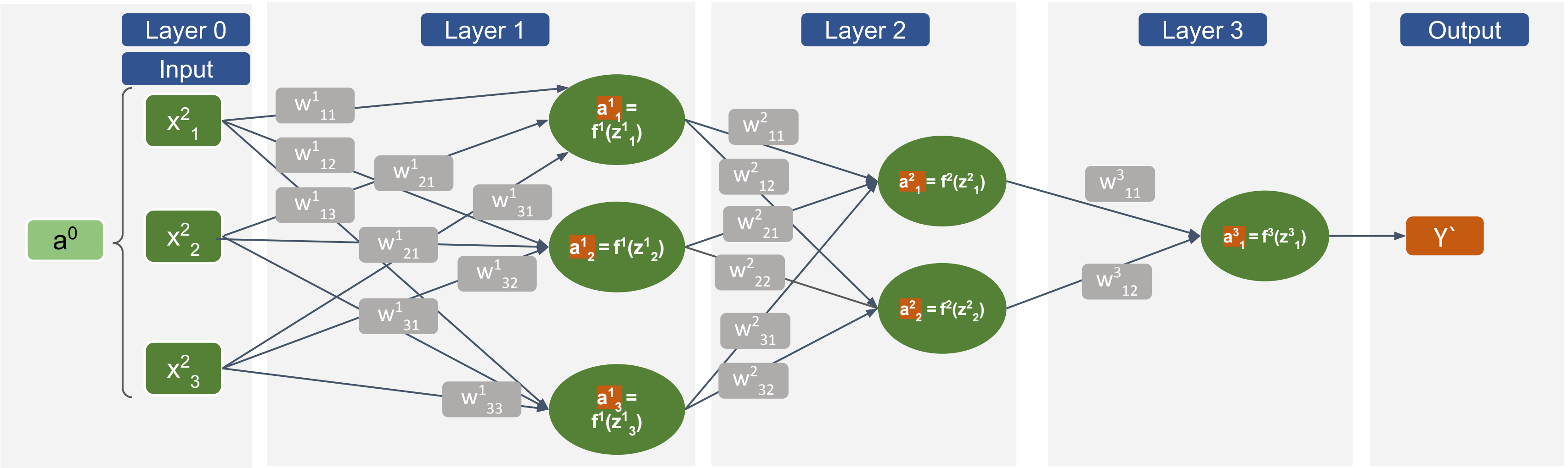
$\partial A^{[L]} : (n^{[L]} , m)$

$W^{[L]} : (n^{[L]} , n^{[L-1]} , m)$

$B^{[L]} : (n^{[L]} , m)$

$\partial W^{[L]} : (n^{[L]} , n^{[L-1]}, m)$

$\partial B^{[L]} : (n^{[L]} , m)$



a^0 → Given: a^0, w^1, b^1; Compute: z^1, a^1 → a^1 → Given: a^1, w^2, b^2; Compute: z^2, a^2 → a^2 → Given: a^2, w^3, b^3; Compute: z^3, a^3 → a^3 → Compute Cross Entropy Loss

Compute Cross Entropy Loss
L(Y`, Y) =
$-Y \log(a^3) + (1-Y) \log(1-a^3)$

$\partial a^3 = -(Y/a^3) + (1-Y)/(1-a^3)$

∂a^3 ← Given: ∂a^3; Compute: ∂z^3, ∂w^3, ∂b^3, ∂a^2 ← ∂a^2 ← Given: ∂a^2; Compute: ∂z^2, ∂w^2, ∂b^2, ∂a^1 ← ∂a^1 ← Given: ∂a^1; Compute: ∂z^1, ∂w^1, ∂b^1, ∂a^0 ← ∂a^0

Given: ∂a^1
Compute: ∂z^1, ∂w^1, ∂b^1, ∂a^0

$w^1 = w^1 - \eta\, \partial w^1$
$b^1 = b^1 - \eta\, \partial b^1$

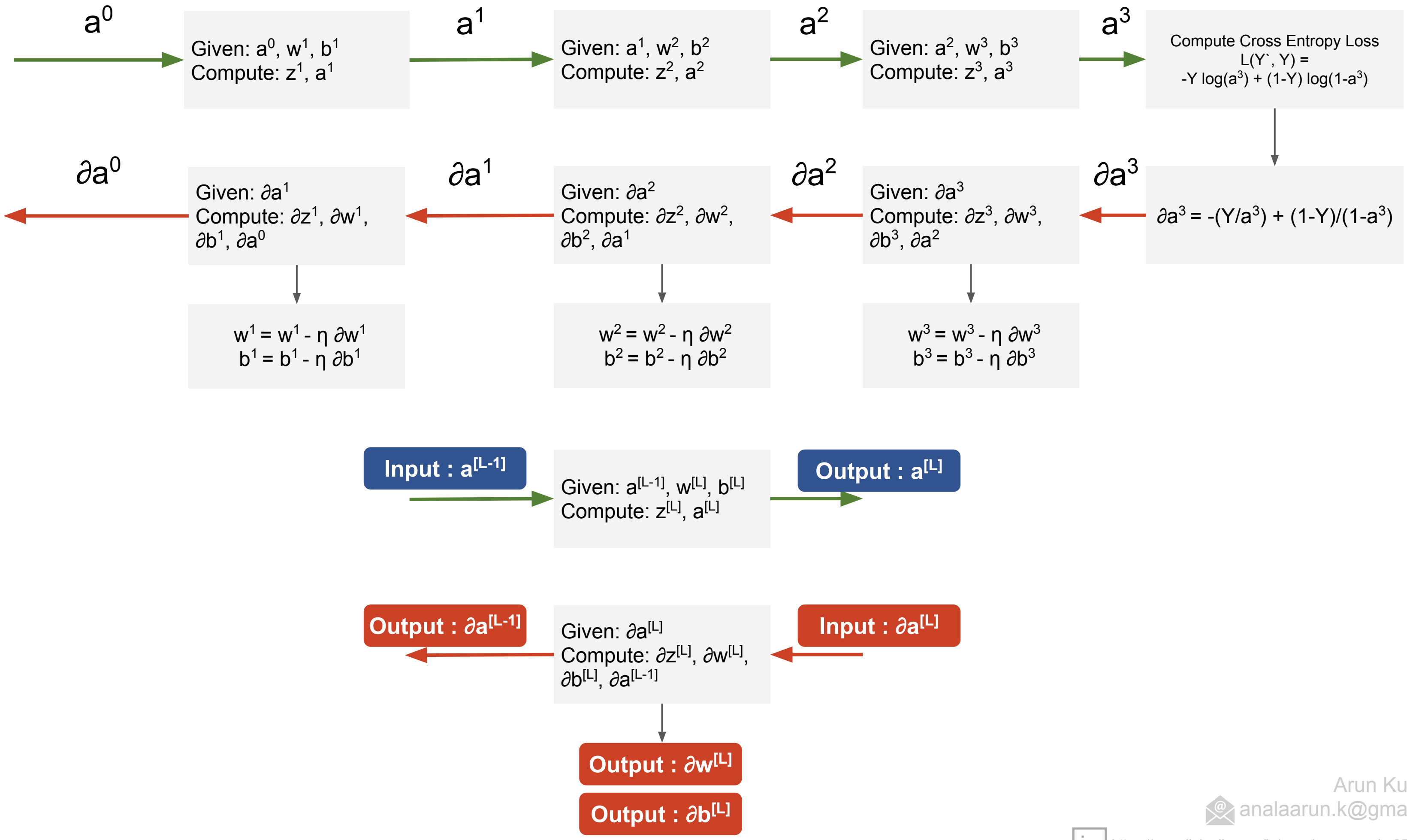
Given: ∂a^2
Compute: ∂z^2, ∂w^2, ∂b^2, ∂a^1

$w^2 = w^2 - \eta\, \partial w^2$
$b^2 = b^2 - \eta\, \partial b^2$

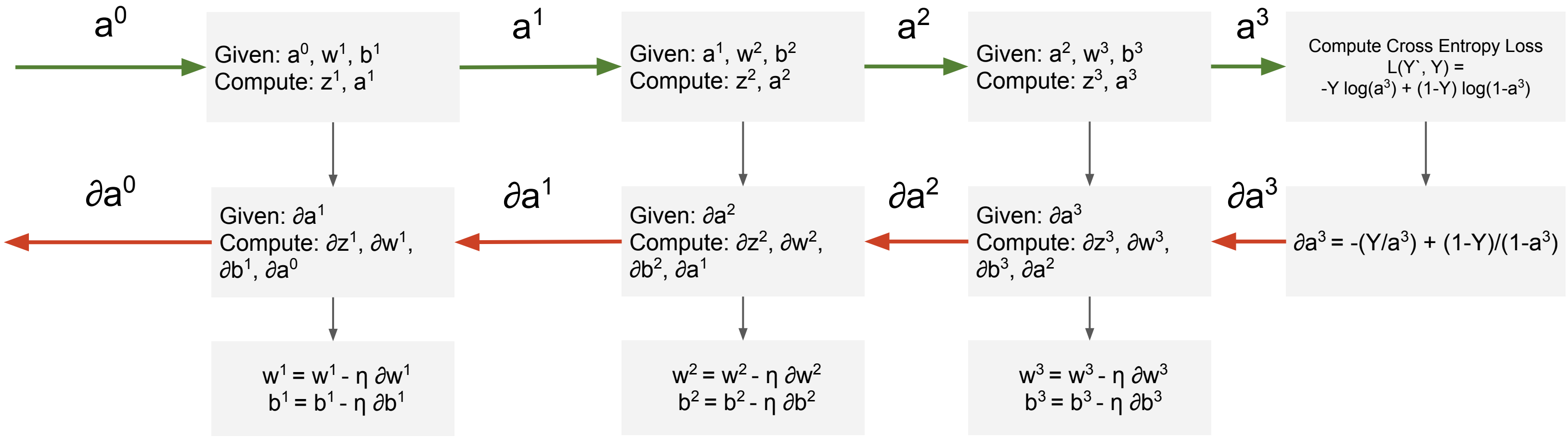
Given: ∂a^3
Compute: ∂z^3, ∂w^3, ∂b^3, ∂a^2

$w^3 = w^3 - \eta\, \partial w^3$
$b^3 = b^3 - \eta\, \partial b^3$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/



Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/



$$a^{[L]} = f^{[L]}(z^{[L]})$$

$$z^{[L]} = w^{[L]} a^{[L-1]} + b^{[L]}$$

$$\partial z^{[L]} = \partial a^{[L]} * f^{[L]`}(z^{[L]})$$

$$\partial w^{[L]} = \partial z^{[L]} * a^{[L-1]}$$

$$\partial b^{[L]} = \partial z^{[L]}$$

$$\partial a^{[L-1]} = w^{[L]T} * \partial z^{[L]}$$

$$\partial z^{[L]} = w^{[L+1]T} * \partial z^{[L+1]} * f^{[L]`}(z^{[L]})$$

$$\partial Z^{[L]} = \partial A^{[L]} * f^{[L]`}(Z^{[L]})$$

$$\partial w^{[L]} = 1/m * \partial Z^{[L]} * A^{[L-1]T}$$

$$\partial b^{[L]} = 1/m * \Sigma(\partial z^{[L]})$$

$$\partial A^{[L-1]} = W^{[L]T} * \partial Z^{[L]}$$

Arun Kumar A
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

For 1 Training Sample

Forward Prop		Backward Prop	
Input : $a^{[L-1]}$	Output : $a^{[L]}$	Input : $\partial a^{[L]}$	Output : $\partial a^{[L-1]}$

Forward Prop	Backward Prop
$z^{[L]} = w^{[L]}\, a^{[L-1]} + b^{[L]}$	$\partial z^{[L]} = \partial a^{[L]} * f^{[L]`}(z^{[L]})$
$a^{[L]} = f^{[L]}\, (z^{[L]})$	$\partial w^{[L]} = \partial z^{[L]} * a^{[L-1]}$
	$\partial b^{[L]} = \partial z^{[L]}$
	$\partial a^{[L-1]} = w^{[L]T} * \partial z^{[L]}$

Forward Prop	Backward Prop
$z^{[L]}$: ($n^{[L]}$, 1)	$\partial z^{[L]}$: ($n^{[L]}$, 1)
$a^{[L]}$: ($n^{[L]}$, 1)	$\partial a^{[L]}$: ($n^{[L]}$, 1)
$w^{[L]}$: ($n^{[L]}$, $n^{[L-1]}$)	$\partial w^{[L]}$: ($n^{[L]}$, $n^{[L-1]}$)
$b^{[L]}$: ($n^{[L]}$, 1)	$\partial b^{[L]}$: ($n^{[L]}$, 1)

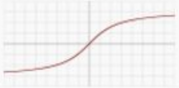
For m Training Samples

Also called Vectorized Implementation

Forward Prop		Backward Prop	
Input : $a^{[L-1]}$	Output : $a^{[L]}$	Input : $\partial A^{[L]}$	Output : $\partial A^{[L-1]}$

Forward Prop	Backward Prop
$Z^{[L]} = (W^{[L]})^T\, A^{[L-1]} + B^{[L]}$	$\partial Z^{[L]} = \partial A^{[L]} * f^{[L]`}(Z^{[L]})$
$A^{[L]} = f^{[L]}\, (Z^{[L]})$	$\partial w^{[L]} = 1/m * \partial Z^{[L]} * A^{[L-1]T}$
	$\partial b^{[L]} = 1/m * \Sigma(\partial z^{[L]})$
	$\partial A^{[L-1]} = W^{[L]T} * \partial Z^{[L]}$

Forward Prop	Backward Prop
$Z^{[L]}$: ($n^{[L]}$, m)	$\partial Z^{[L]}$: ($n^{[L]}$, m)
$A^{[L]}$: ($n^{[L]}$, m)	$\partial A^{[L]}$: ($n^{[L]}$, m)
$W^{[L]}$: ($n^{[L]}$, $n^{[L-1]}$, m)	$\partial W^{[L]}$: ($n^{[L]}$, $n^{[L-1]}$, m)
$B^{[L]}$: ($n^{[L]}$, m)	$\partial B^{[L]}$: ($n^{[L]}$, m)

Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parameteric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

Arun Kumar A
alaarun.k@gmail.com
/in/arun-kumar-anala-35760523/