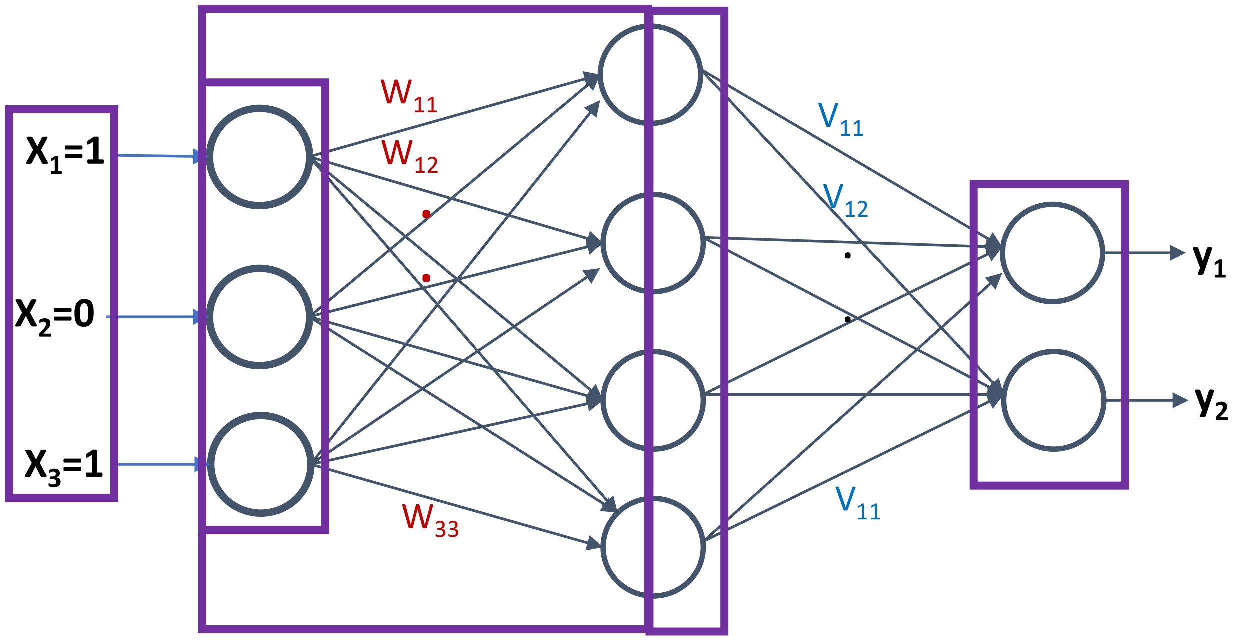
Lesson 11

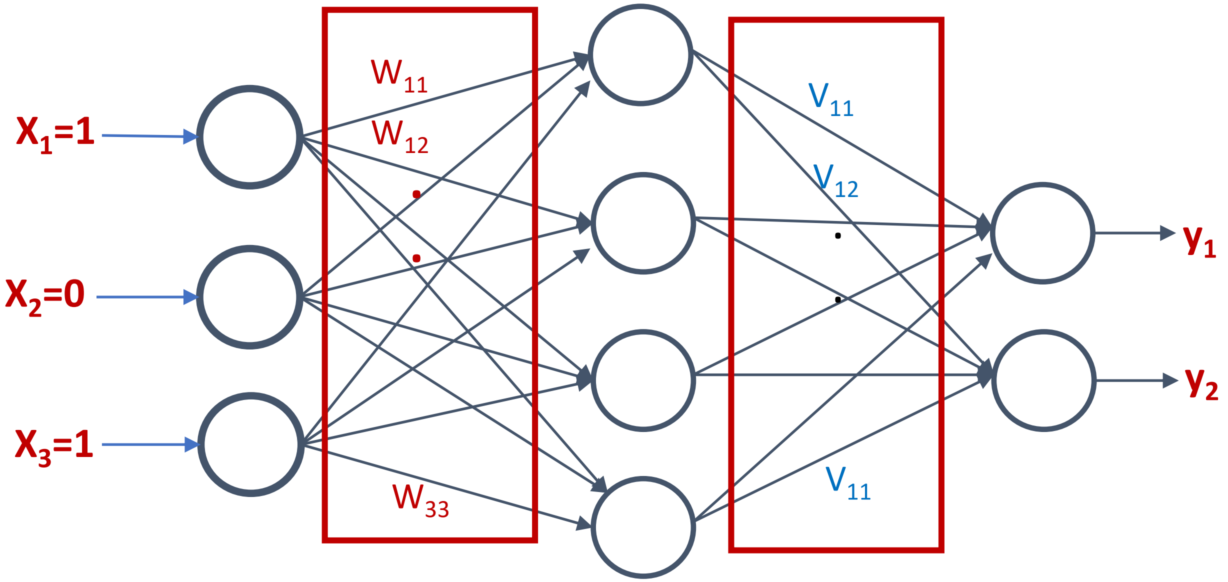
Learning the parameters
– Backpropagation through time

What are the parameters?



$$\bar{y}^T = f_{hidden}\left(f_{output}\left(\bar{x}^T \boldsymbol{W}\right)^T . \boldsymbol{V}\right)$$

What are the parameters?



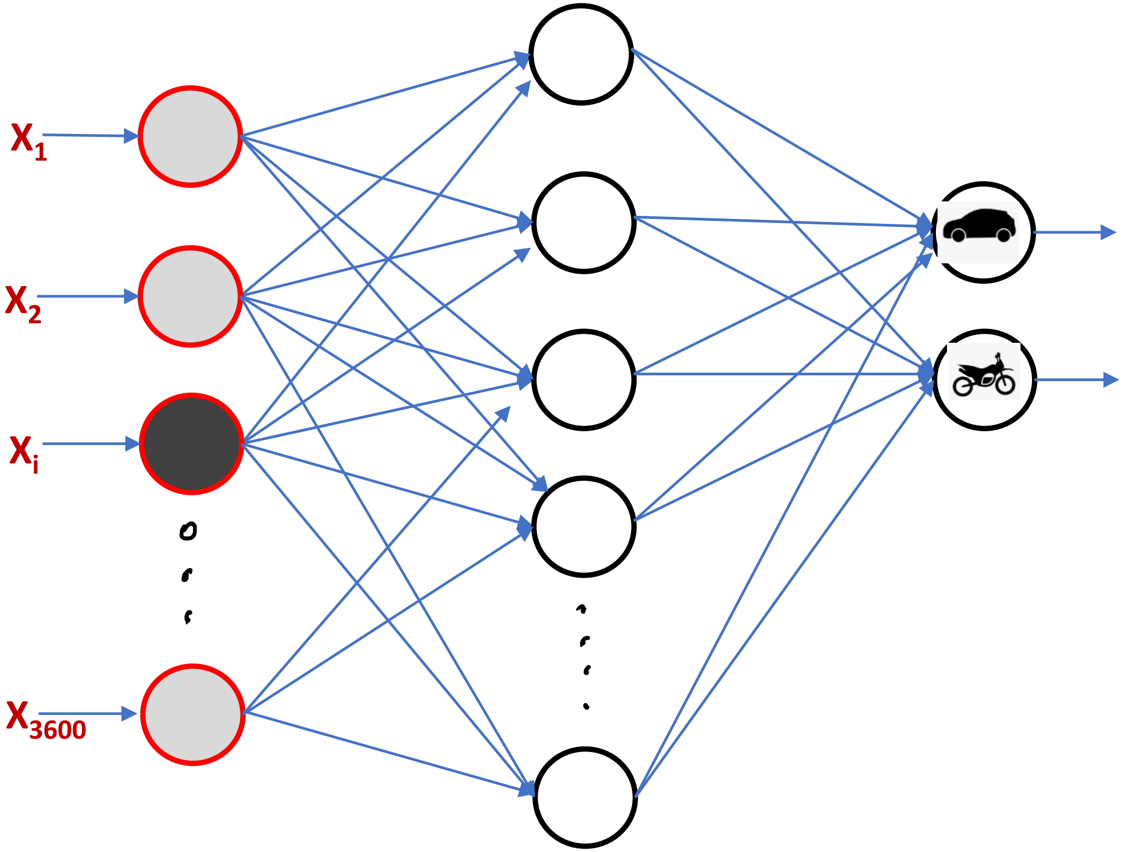
	h_1	h_2	h_3	h_4
i_1	W_{11}	W_{12}	W_{13}	W_{14}
i_2	W_{21}	W_{22}	W_{23}	W_{24}
i_3	W_{31}	W_{32}	W_{33}	W_{34}

W

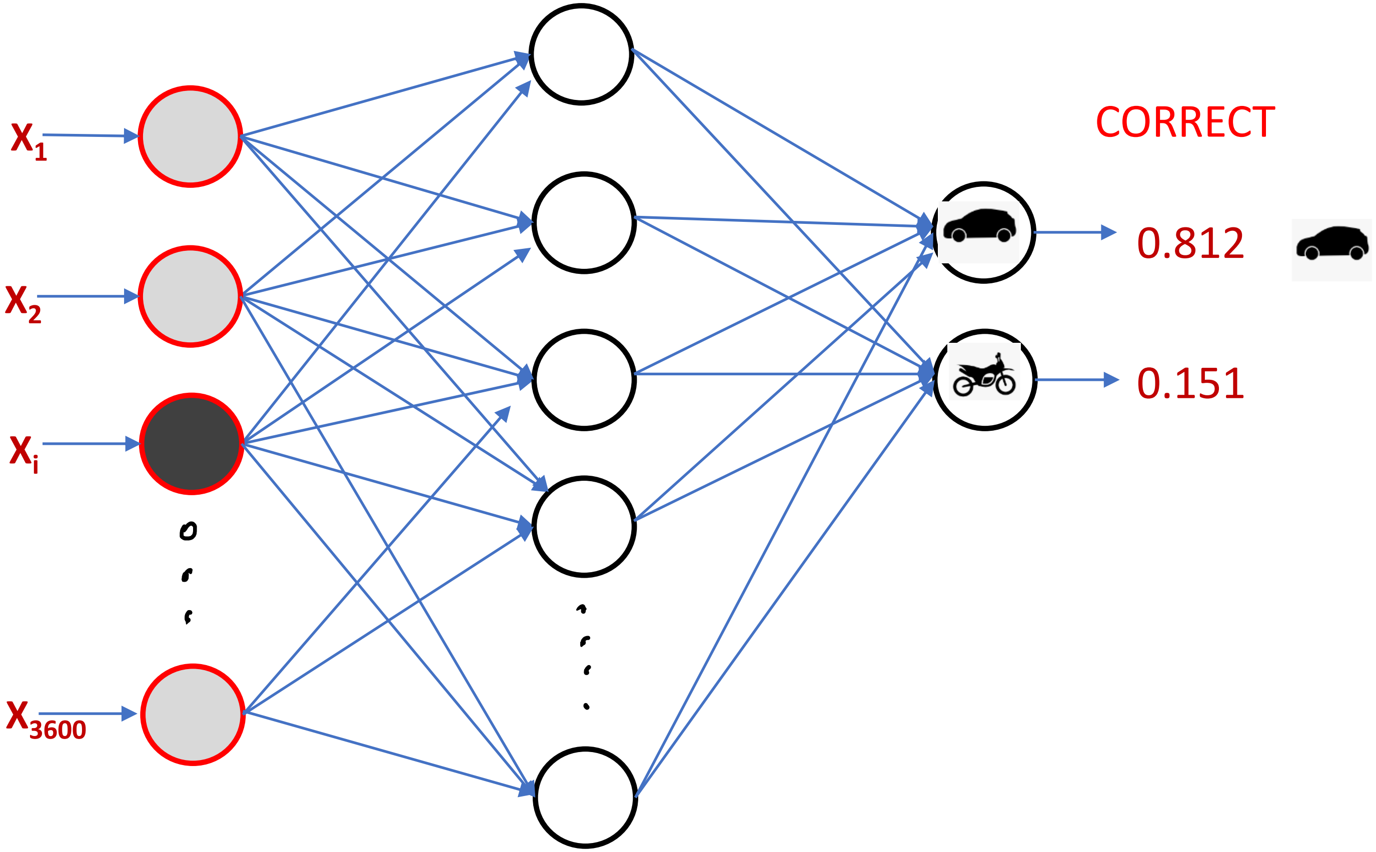
	o_1	o_2
h_1	V_{11}	V_{12}
h_2	V_{21}	V_{22}
h_3	V_{31}	V_{32}
h_4	V_{41}	V_{42}

V

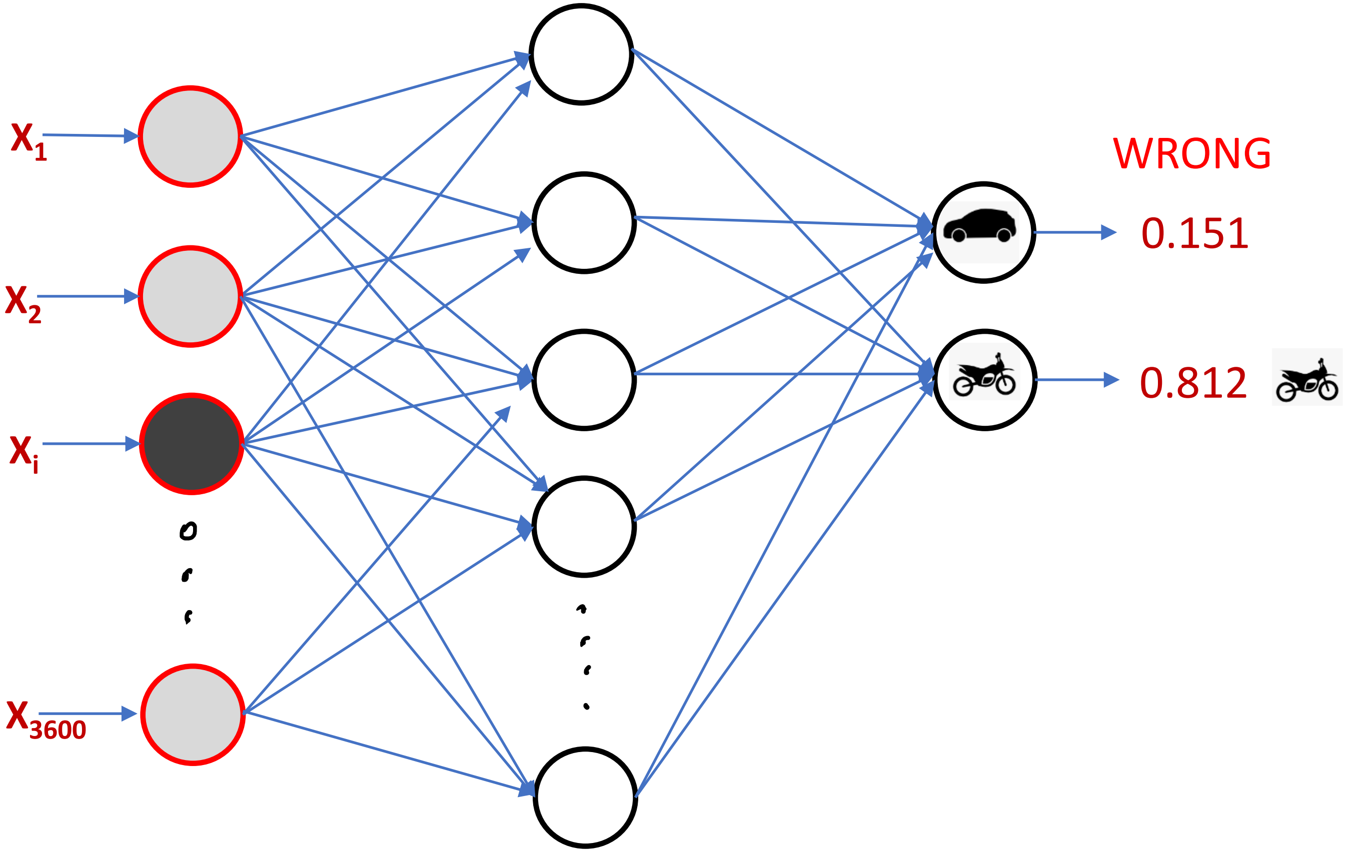
60x60=3600



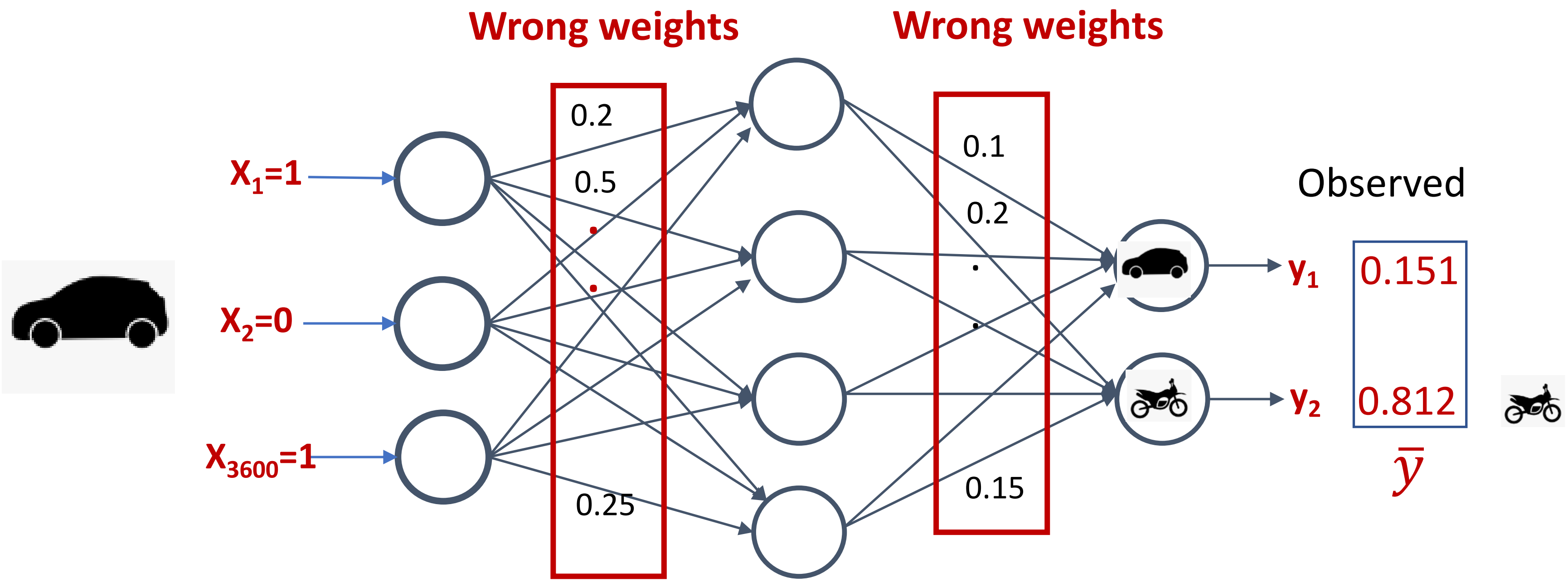
60x60=3600

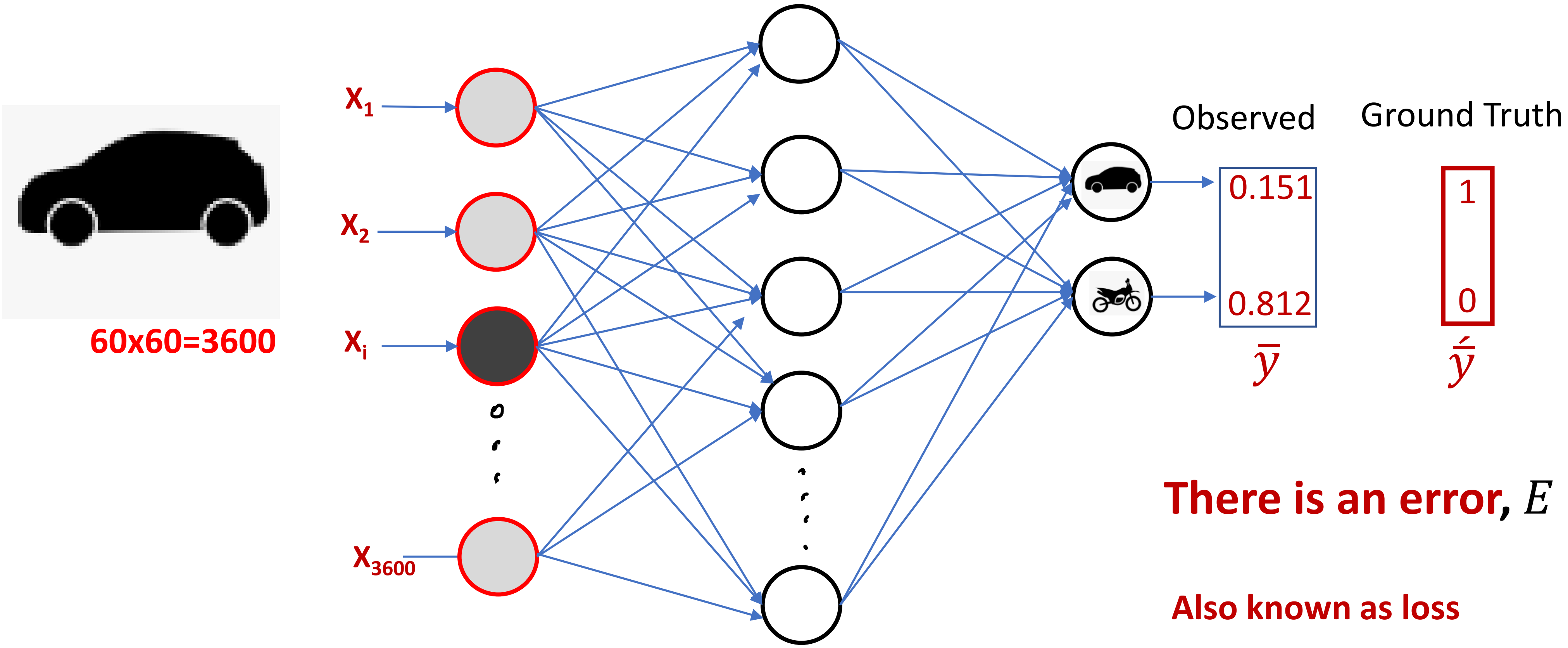


60x60=3600



Why there is an error in prediction?

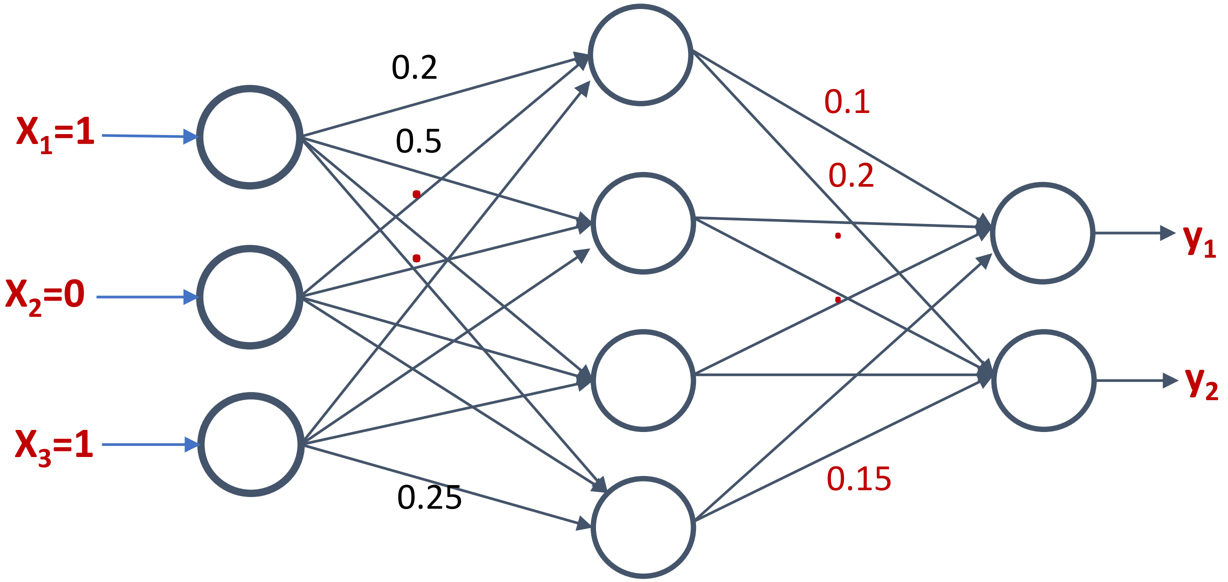




There is an error, $E = \|\bar{\boldsymbol{y}} - \acute{\bar{\boldsymbol{y}}}\|$

Also known as loss

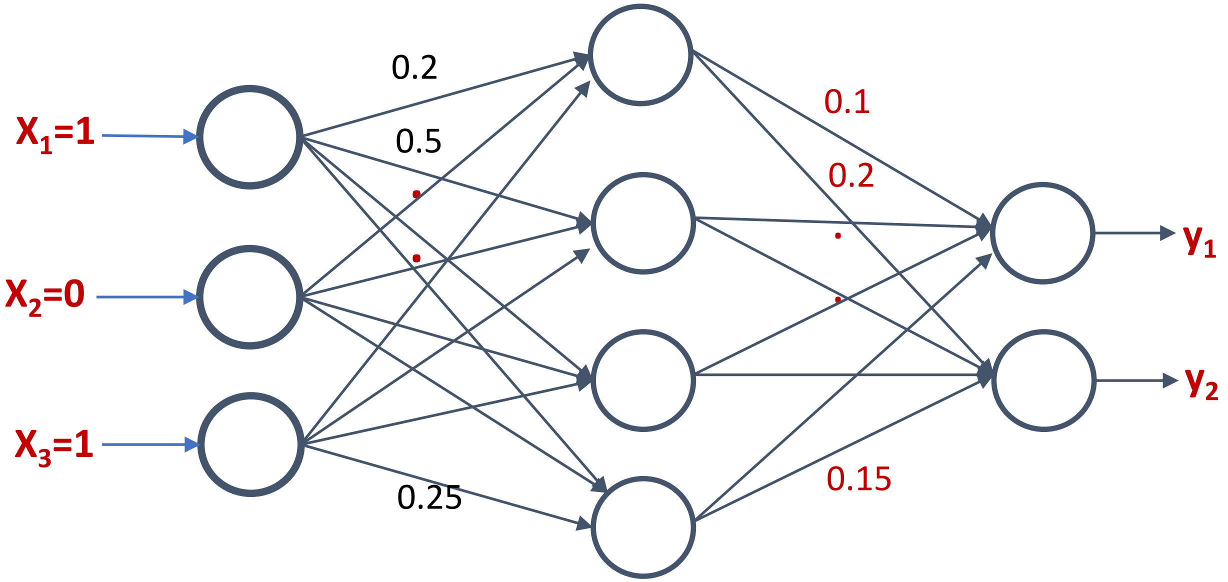
Backpropagation Through Time



Minimize the Loss function $E = \|\bar{y} - \acute{y}\|$

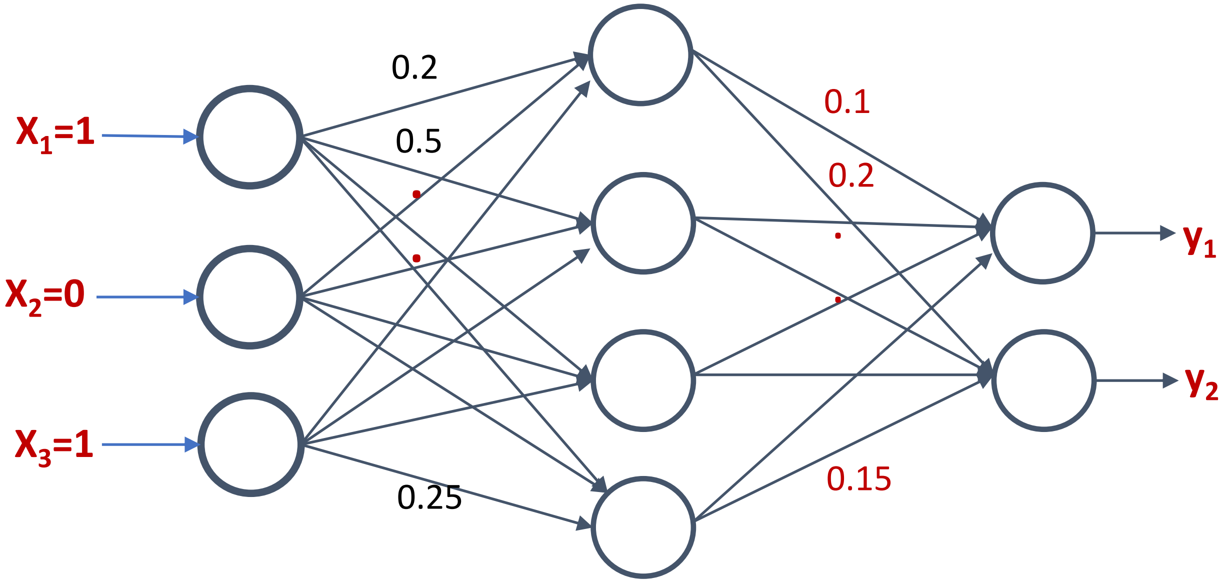
Backpropagation Through Time

Minimize the Loss function $E = \|\bar{y} - \acute{y}\|$



Backpropagation Through Time

Minimize the Loss function $E = \|\bar{y} - \acute{y}\|$

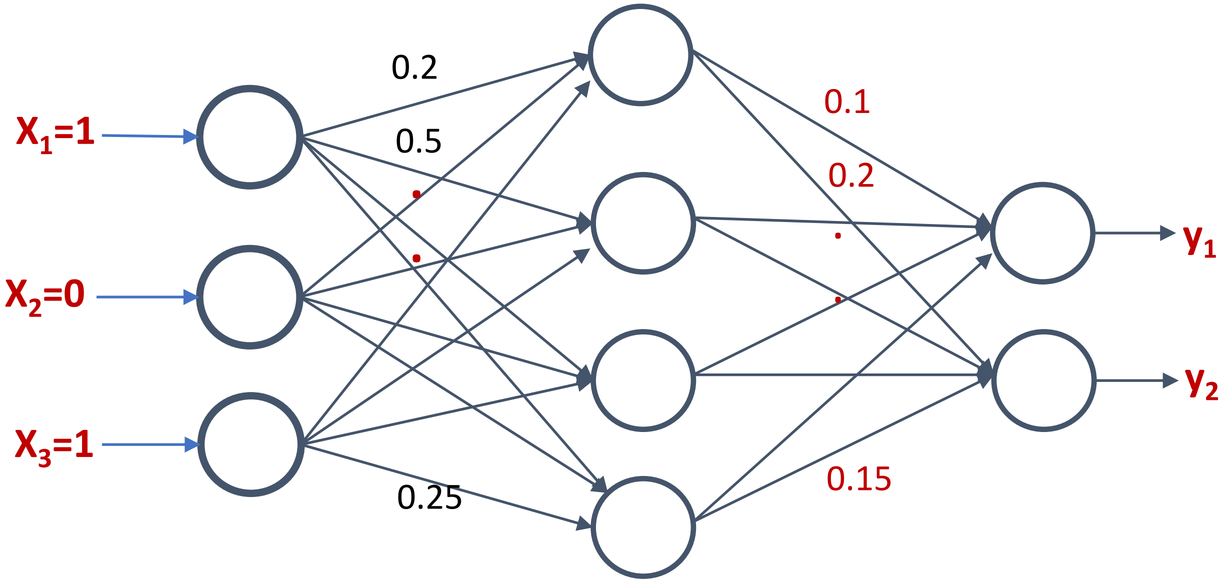


$$\nabla = \frac{\delta E}{\delta V} = 0$$

$$\nabla_{ij} = \frac{\delta E}{\delta V_{ij}} = \frac{\delta \|\bar{y} - \acute{y}\|}{\delta V_{ij}} = 0$$

Backpropagation Through Time

Minimize the Loss function $E = \|\bar{y} - \acute{y}\|$



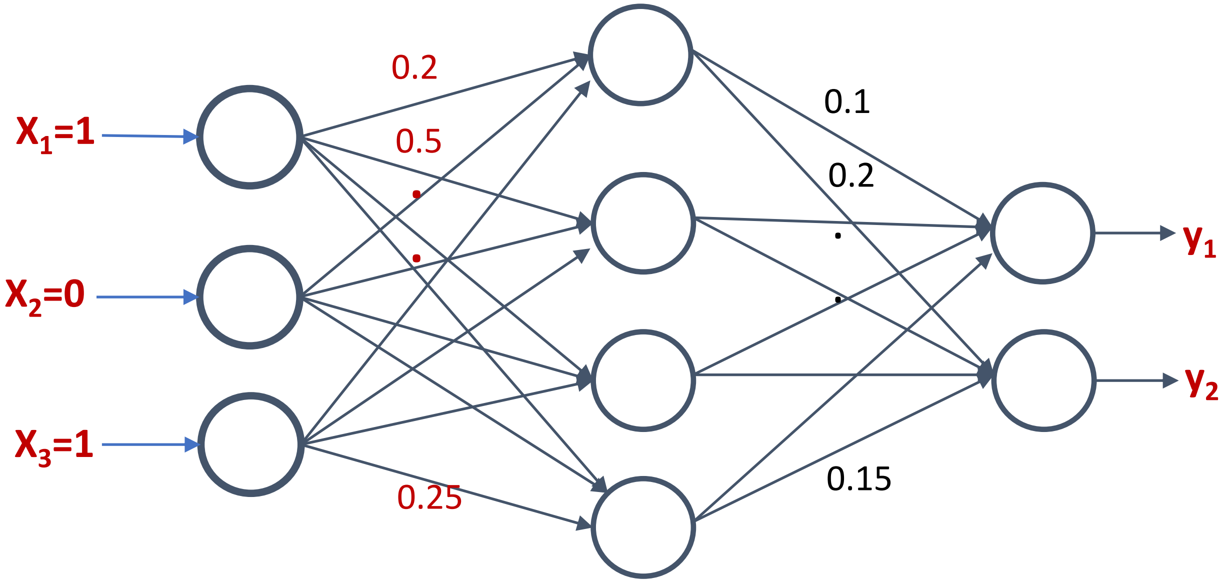
$$\nabla = \frac{\delta E}{\delta V} = 0$$

$$\nabla_{ij} = \frac{\delta E}{\delta V_{ij}} = \frac{\delta \|\bar{y} - \acute{y}\|}{\delta V_{ij}} = 0$$

$$V_{ij}^{t} = V_{ij}^{t-1} + \eta \, \nabla_{ij}^{t}$$

$$\boldsymbol{\eta = [0, 1]}$$

Backpropagation Through Time

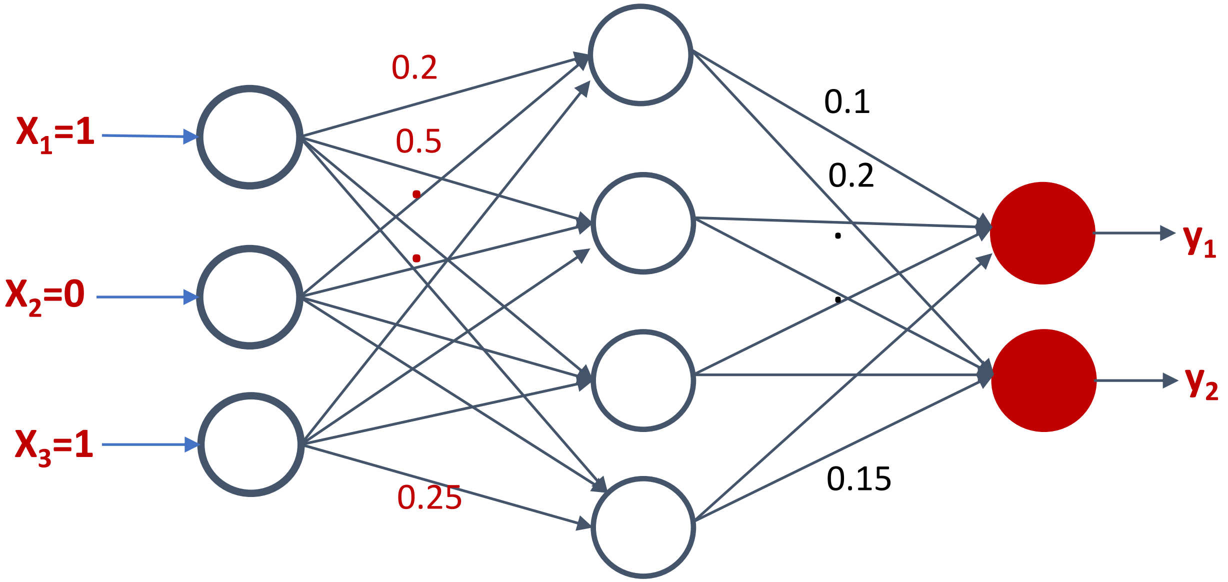


$$E = \|\bar{y} - \acute{\bar{y}}\| = 0$$

or

$$E = \|\bar{y} - \acute{\bar{y}}\| \leq \epsilon$$

Backpropagation Through Time

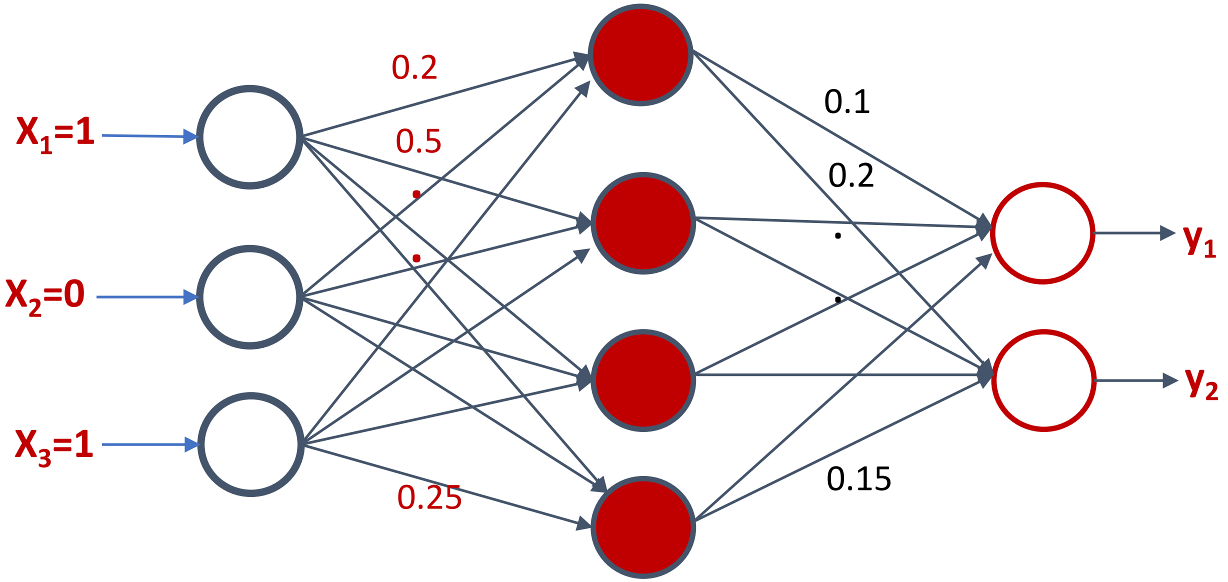


$$E = \|\bar{y} - \acute{\bar{y}}\| = 0$$

or

$$E = \|\bar{y} - \acute{\bar{y}}\| \leq \epsilon$$

Backpropagation Through Time

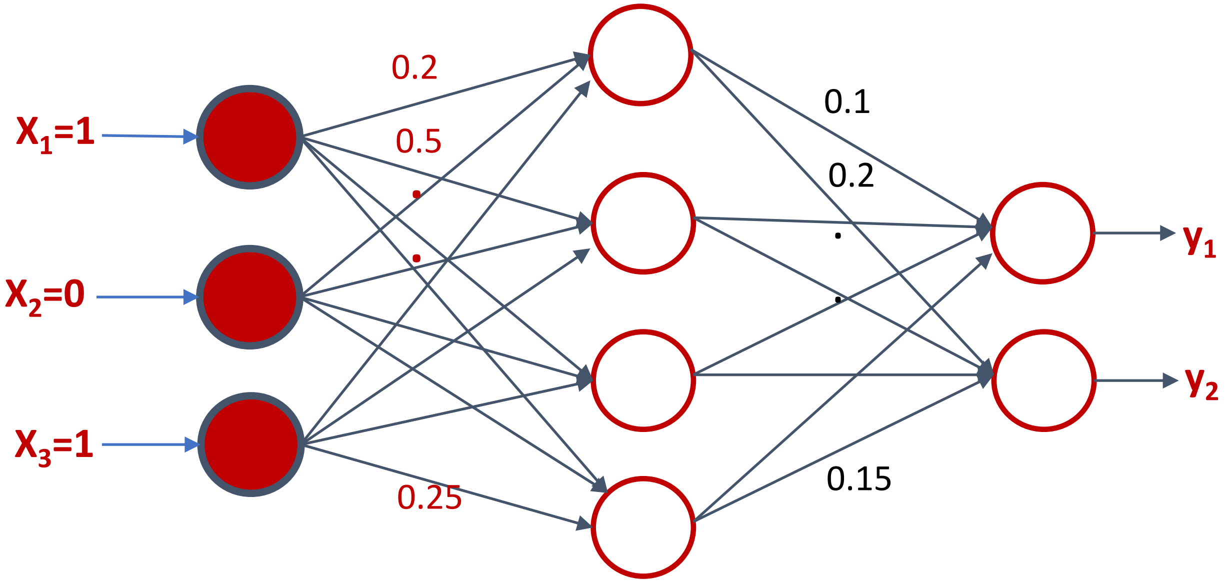


$E = \|\bar{y} - \acute{\bar{y}}\| = 0$

or

$E = \|\bar{y} - \acute{\bar{y}}\| \leq \epsilon$

Backpropagation Through Time

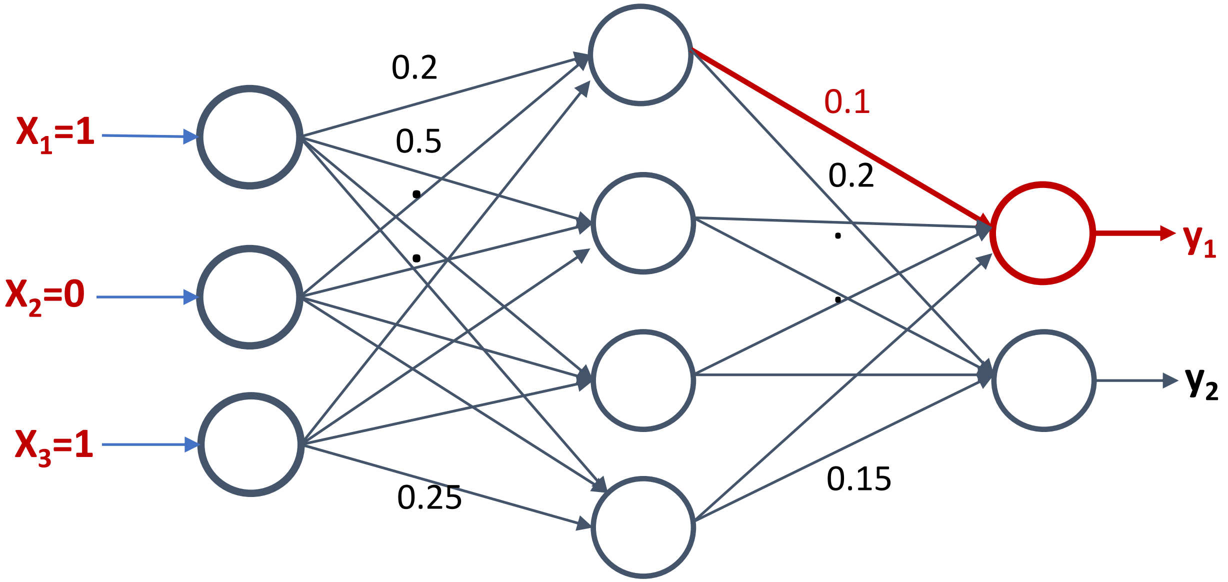


$$E = \|\bar{y} - \acute{\bar{y}}\| = 0$$

or

$$E = \|\bar{y} - \acute{\bar{y}}\| \le \epsilon$$

How are the Derivatives performed



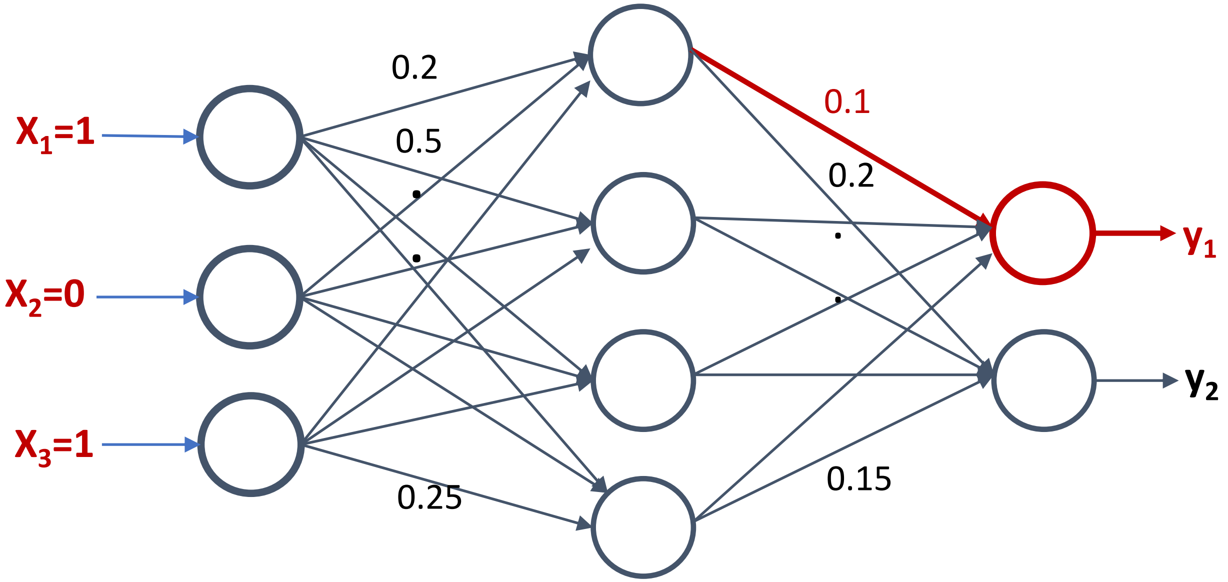
Loss function $E = \|\bar{y} - \acute{y}\|$

$$\nabla = \frac{\delta E}{\delta V} = 0$$

$$\nabla_{11} = \frac{\delta E}{\delta V_{11}} = 0$$

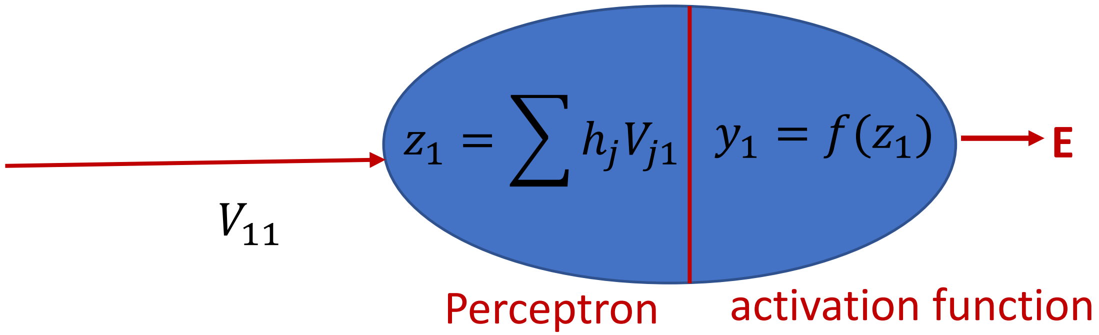
How are the Derivatives performed

Loss function $E = \|\bar{y} - \acute{y}\|$



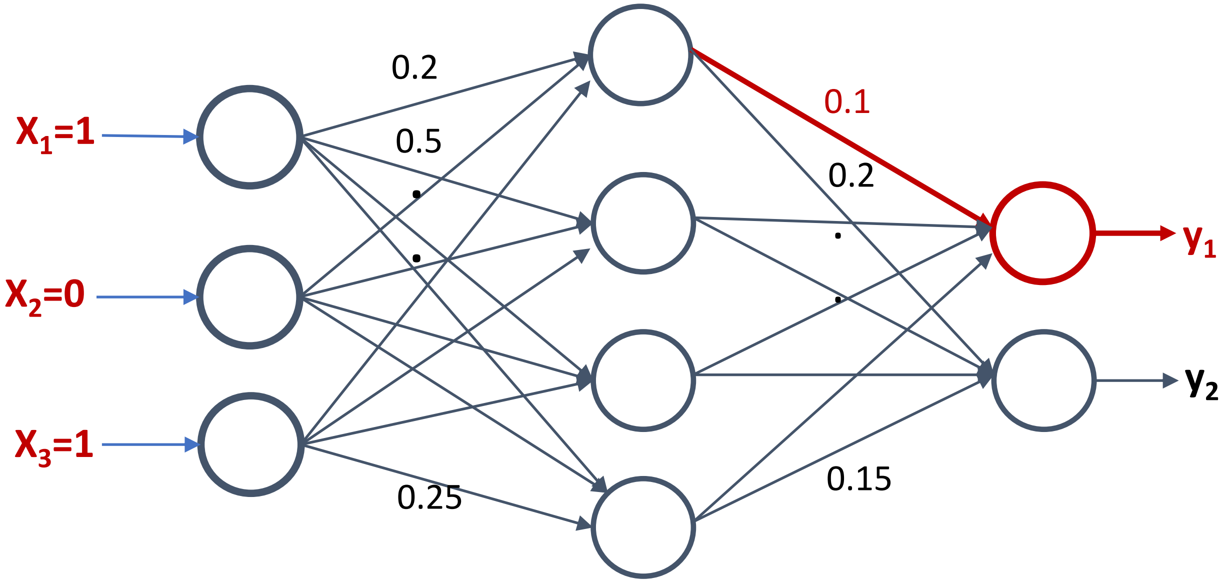
$$\nabla = \frac{\delta E}{\delta V} = 0$$

$$\nabla_{11} = \frac{\delta E}{\delta V_{11}} = 0$$



How are the Derivatives performed

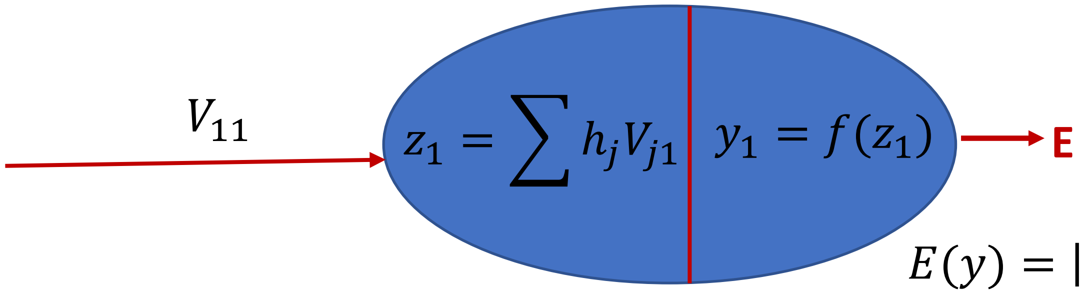
Loss function $E = \|\bar{y} - \acute{y}\|$



$$\nabla = \frac{\delta E}{\delta V} = 0$$

$$\nabla_{11} = \frac{\delta E}{\delta V_{11}} = 0$$

$$z_1(V_{11}) = h_1V_{11} + h_2V_{21} + h_3V_{31} + h_4V_{41}$$

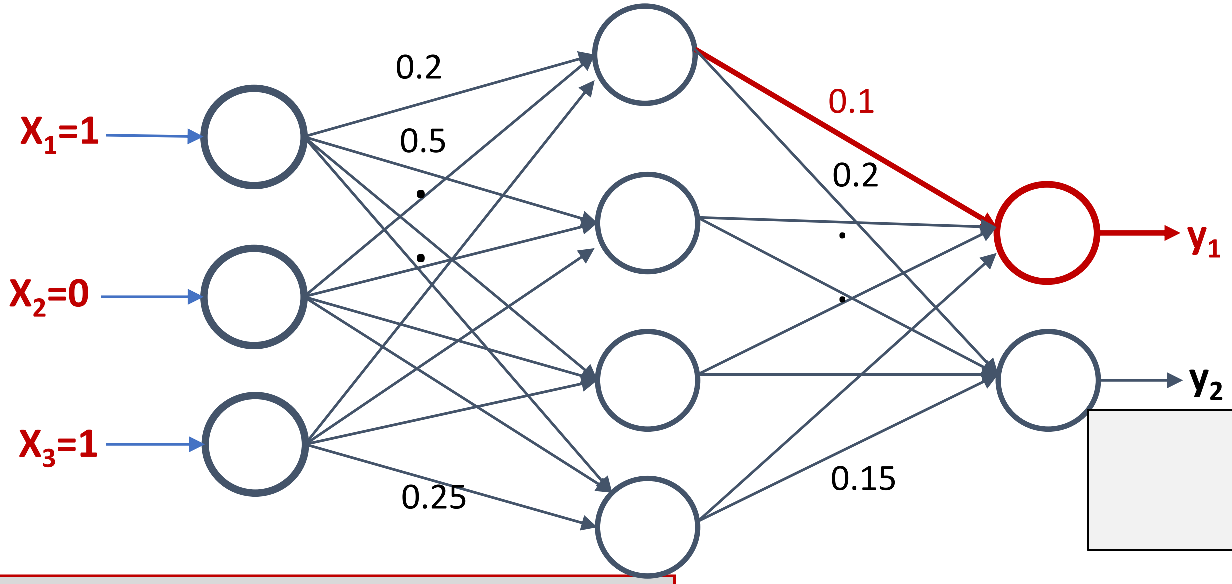


$$E(y) = \|\bar{y} - \acute{y}\|$$

$$y_1(z_1) = Sigmoid(z_1)$$

How are the Derivatives performed

Loss function $E = \|\bar{y} - \acute{y}\|$



$$\nabla = \frac{\delta E}{\delta V} = 0$$

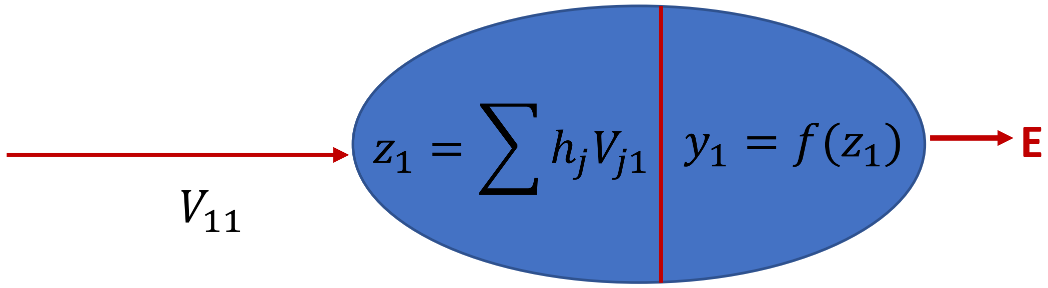
$$\nabla_{ij} = \frac{\delta E}{\delta V_{ij}} = 0$$

$$\frac{\delta E}{\delta V_{11}} = \frac{\delta z_1}{\delta V_{11}} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$

$$\nabla_{11} = \frac{\delta E}{\delta V_{11}} = \frac{\delta z_1}{\delta V_{11}} \cdot \frac{\delta y_1}{\delta z_1} \cdot \frac{\delta E}{\delta y_1}$$

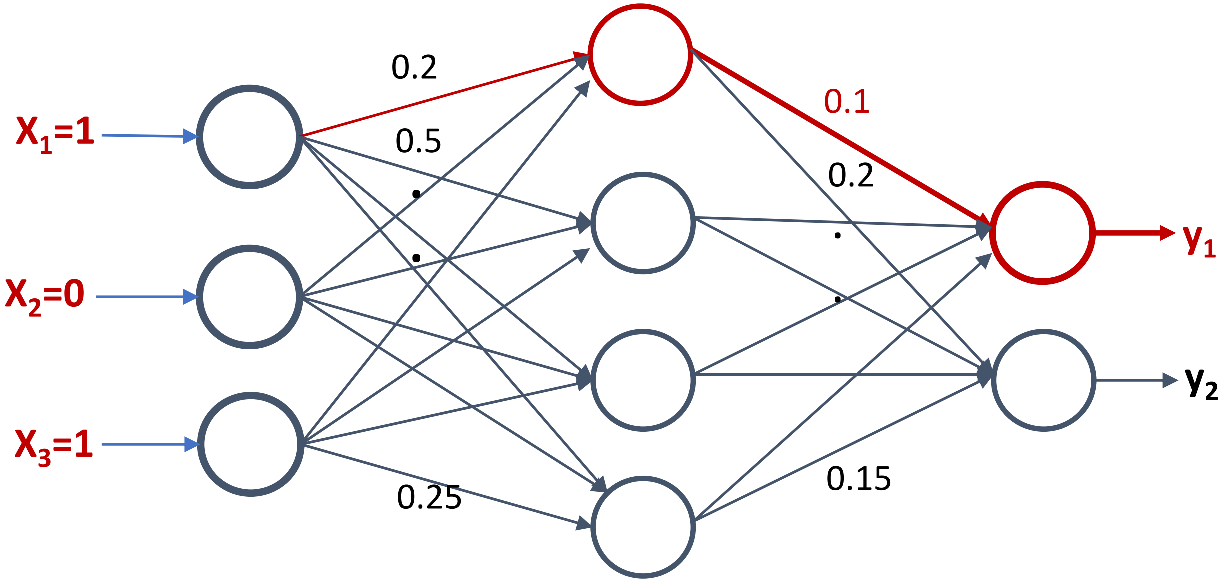
$$z_1 = h_1 V_{11} + h_2 V_{21} + h_3 V_{31} + h_4 V_{41}$$

$$y_1 = Sigmoid(\boldsymbol{z_1})$$



Backpropagation

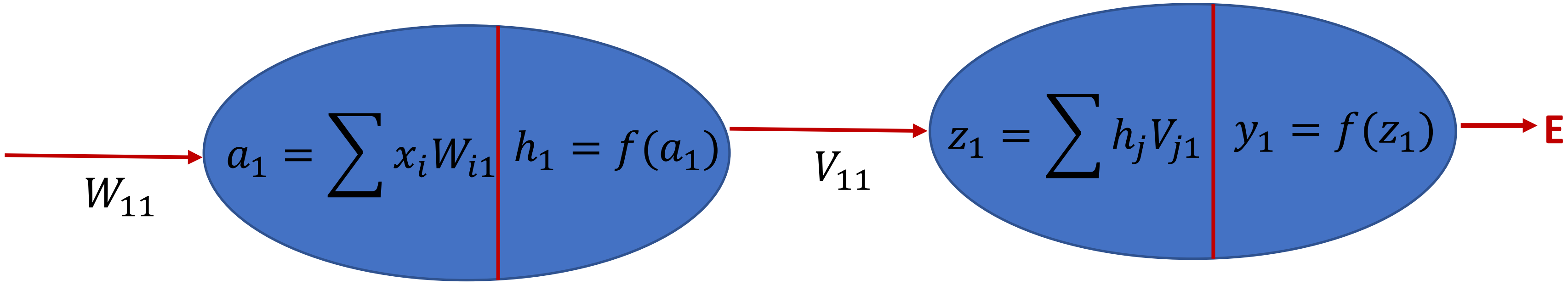
Loss function $E = \|\bar{y} - \acute{y}\|$



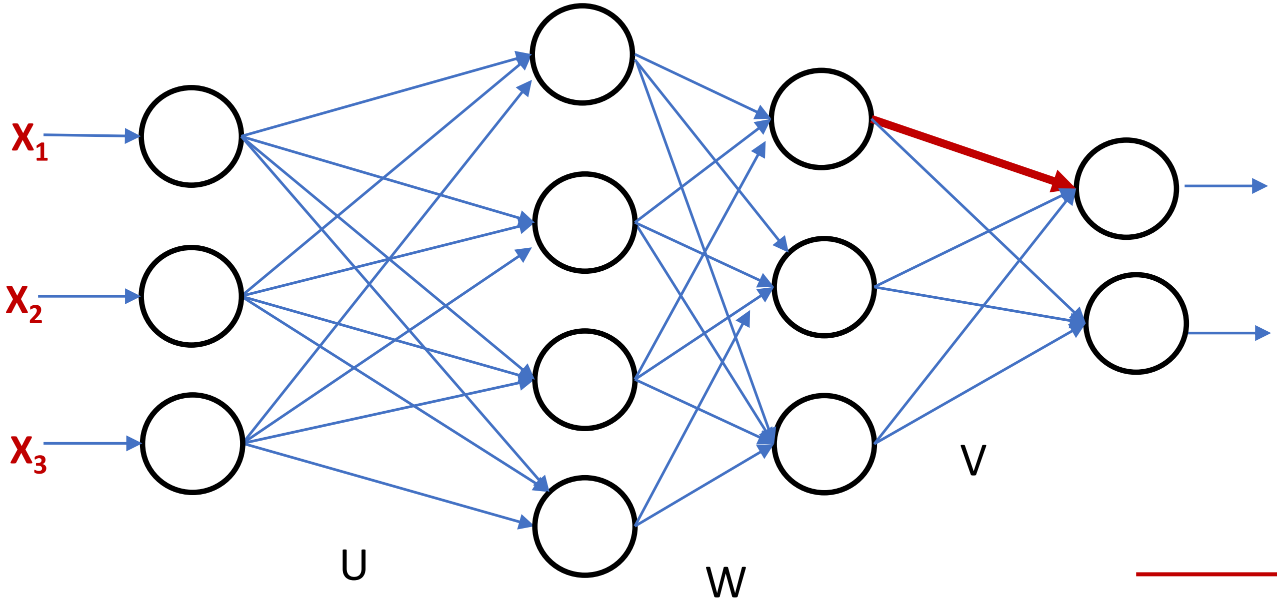
$$\nabla = \frac{\delta E}{\delta W} = 0$$

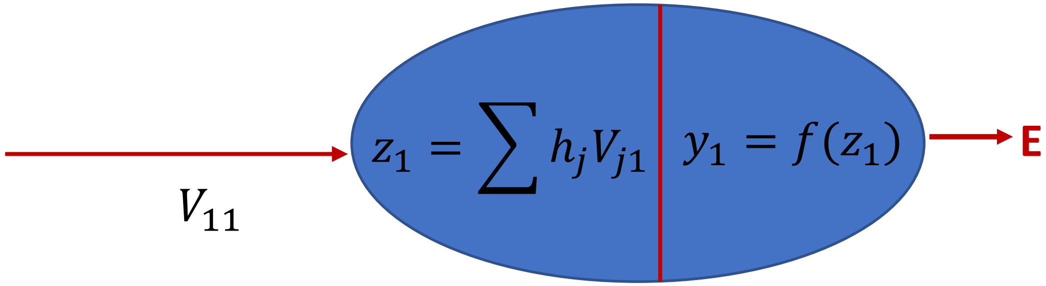
$$\nabla_{ij} = \frac{\delta E}{\delta W_{ij}} = 0$$

$$\frac{\delta E}{\delta W_{11}} = \frac{\delta a_1}{\delta W_{11}} \times \frac{\delta h_1}{\delta a_1} \times \qquad \frac{\delta z_1}{\delta h_1} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$



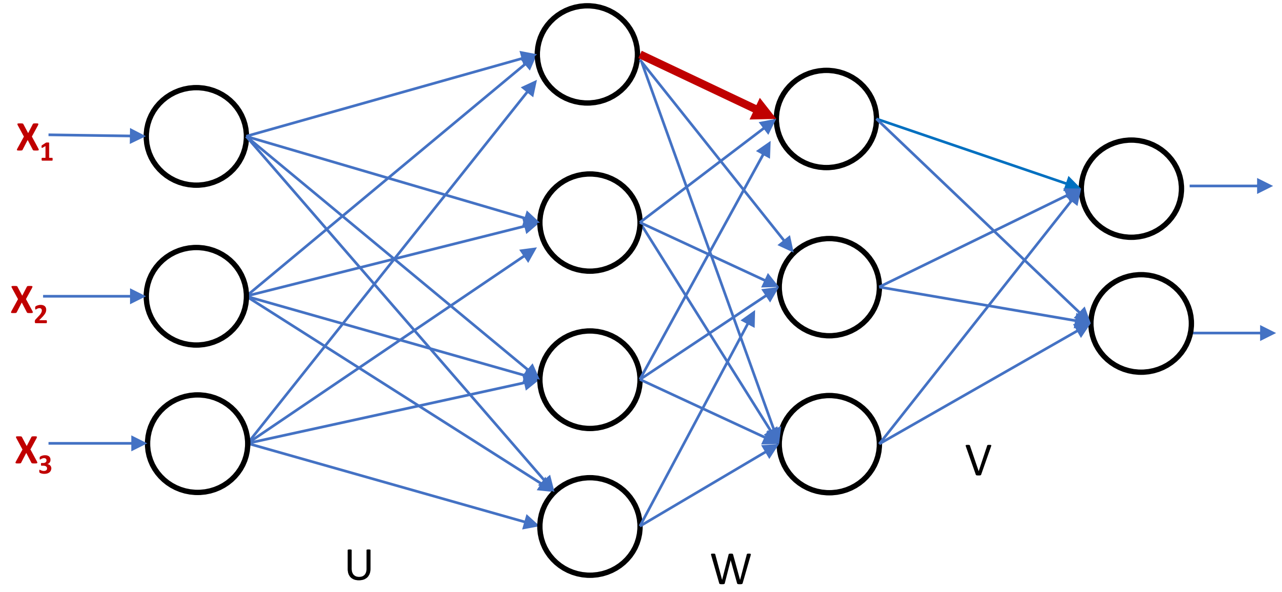
We may have multiple layers.



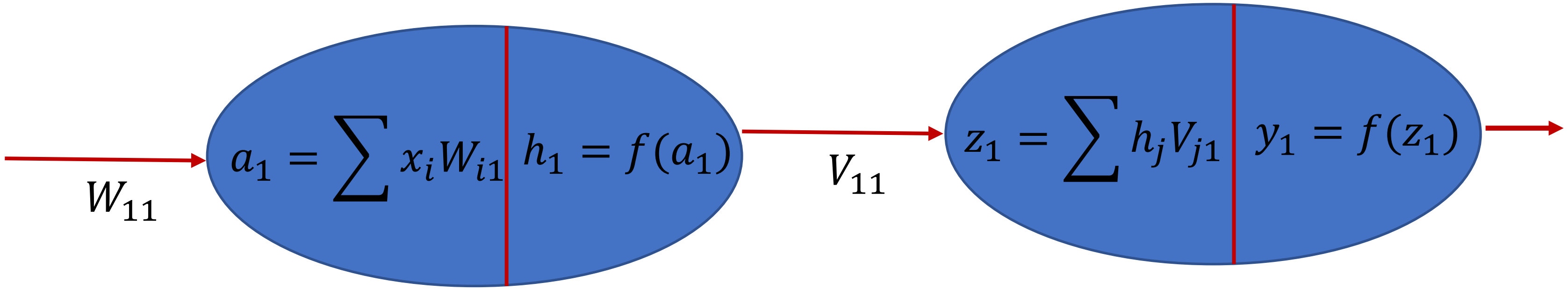


$$\frac{\delta E}{\delta V_{11}} = \frac{\delta z_1}{\delta V_{11}} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$

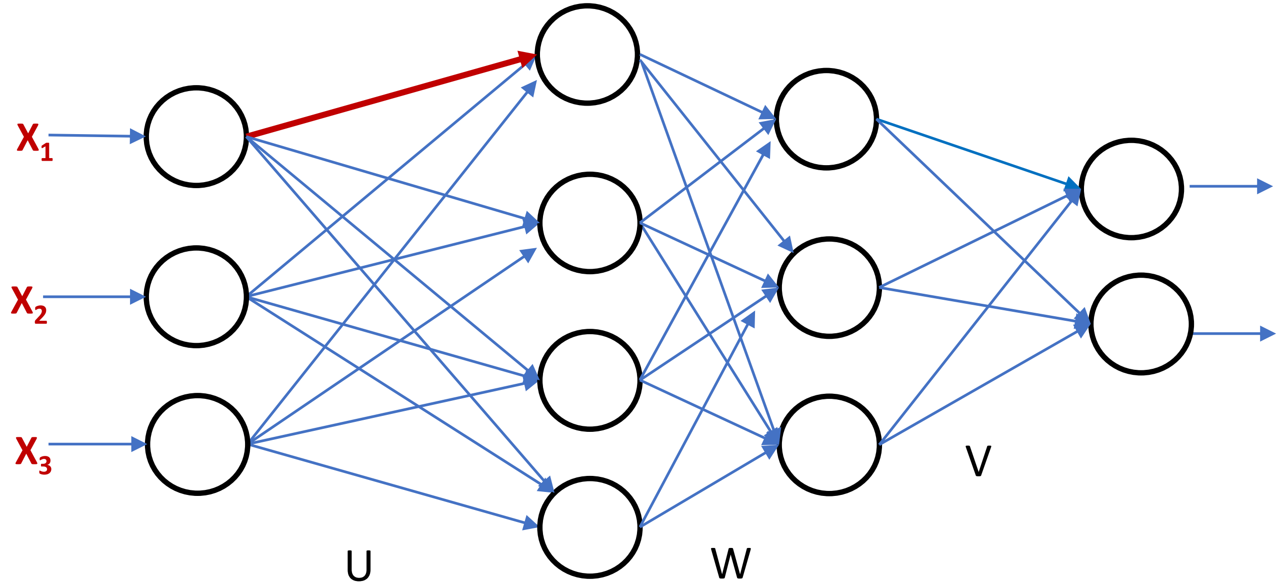
We may have multiple layers



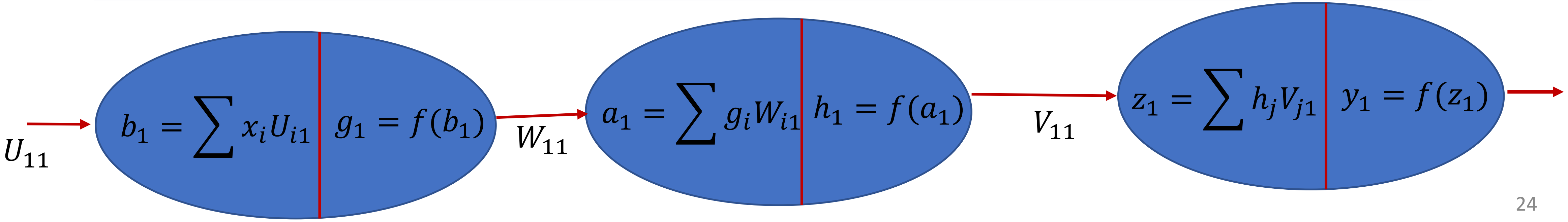
$$\frac{\delta E}{\delta W_{11}} = \frac{\delta a_1}{\delta W_{11}} \times \frac{\delta h_1}{\delta a_1} \times \quad \frac{\delta z_1}{\delta h_1} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$



We may have multiple layers



$$\frac{\delta E}{\delta U_{11}} = \frac{\delta b_1}{\delta U_{11}} \times \frac{\delta g_1}{\delta b_1} \times \frac{\delta a_1}{\delta g_1} \times \frac{\delta h_1}{\delta a_1} \times \frac{\delta z_1}{\delta h_1} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta E}{\delta y_1}$$



Summary

- What is loss function?
- What are the parameters of a multilayer perceptron neural network?
- How to estimate parameters using backpropagation through time?