Training a Multi-Layer Perceptron

Dr. Fani Deligianni,

fani.deligianni@glasgow.ac.uk

Lecturer (Assistant Professor)

Lead of the Computing Technologies for Healthcare Theme
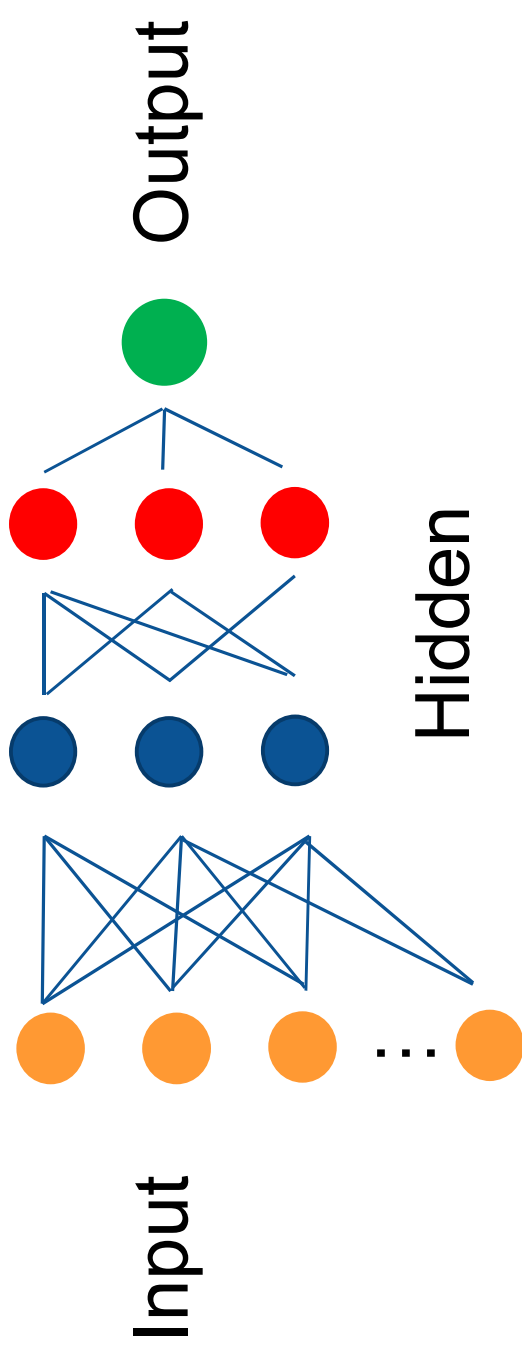
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

University | School of
of Glasgow | Computing Science

THE AWARDS 2020

UNIVERSITY OF THE YEAR

WORLD CHANGING GLASGOW

Training an MLP

- Forward Propagation
- Error Computation
- Backpropagation
- Parameter update

Output

Hidden
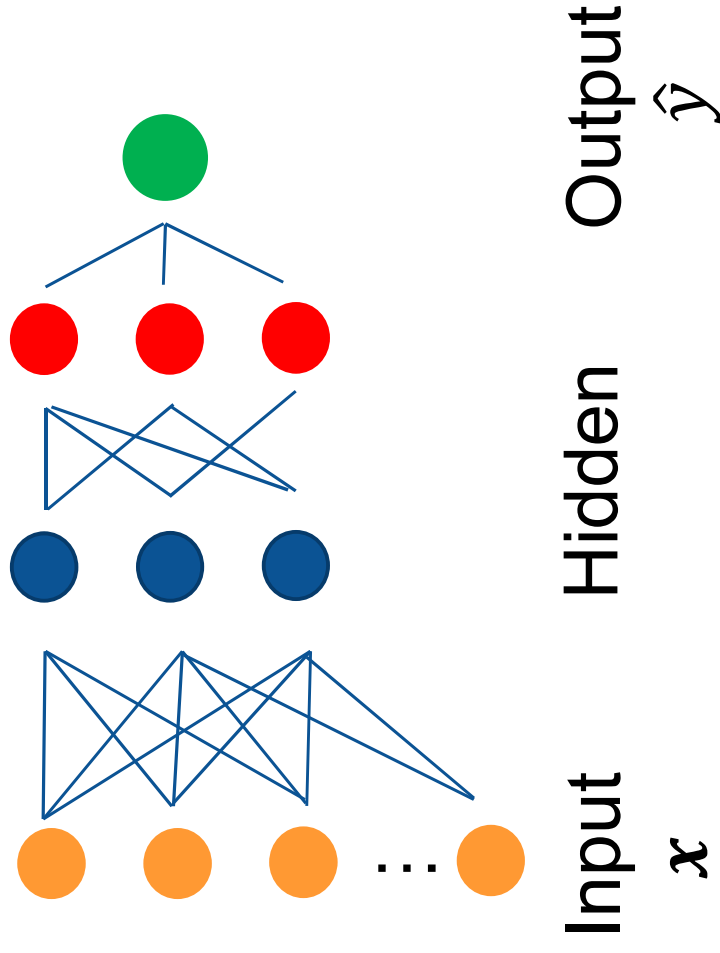
Input

Forward Propagation

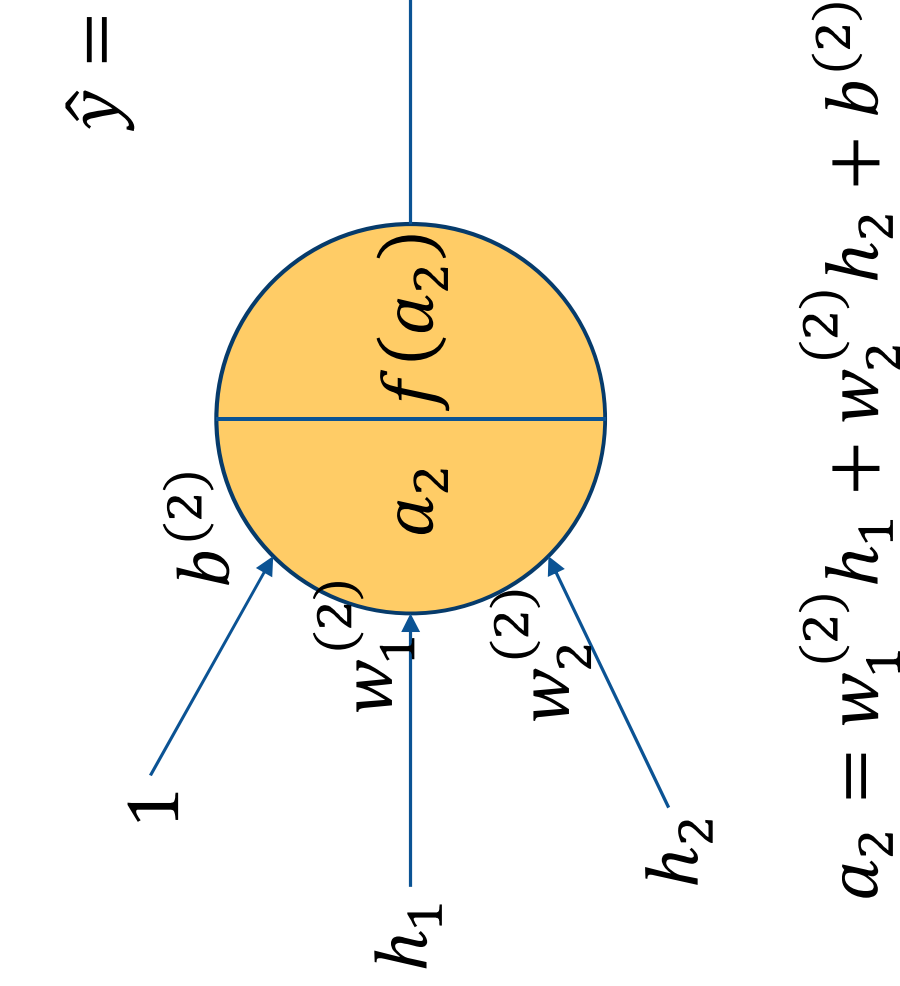$$f(x) = \frac{1}{1 - e^{-x}}$$

Activation Function

$$h_1 = f(W_1 x + b_1)$$
$$h_2 = f(W_2 h_1 + b_2)$$
$$\vdots$$
$$\hat{y} = h_n$$

Input
x

Hidden

Output
$\hat{y}$

Error Computation
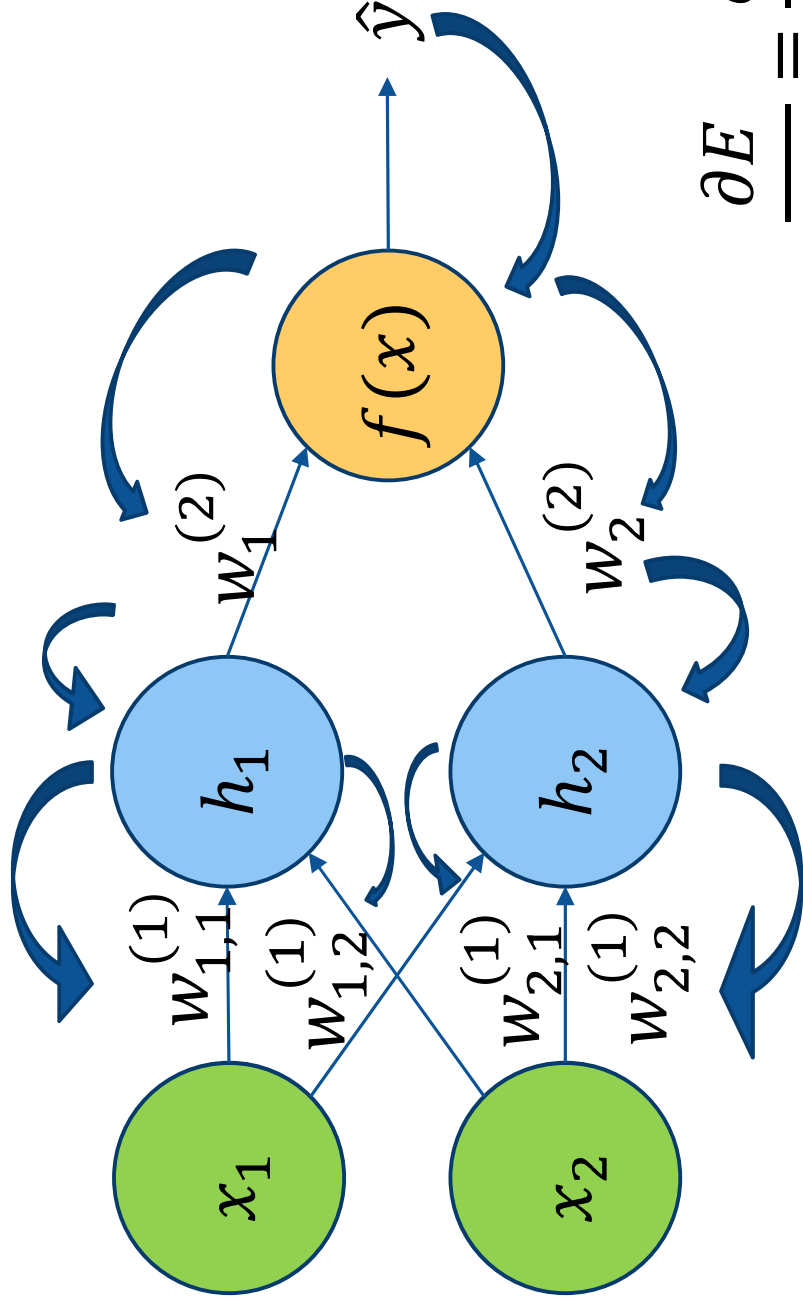

$$\hat{y} = f(a_2) = \frac{1}{1 - e^{-a_2}}$$

$$E_{out}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

Output Error

$$E(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{2k} \sum_{1=1}^{k} (\hat{y}_k - y_k)^2$$

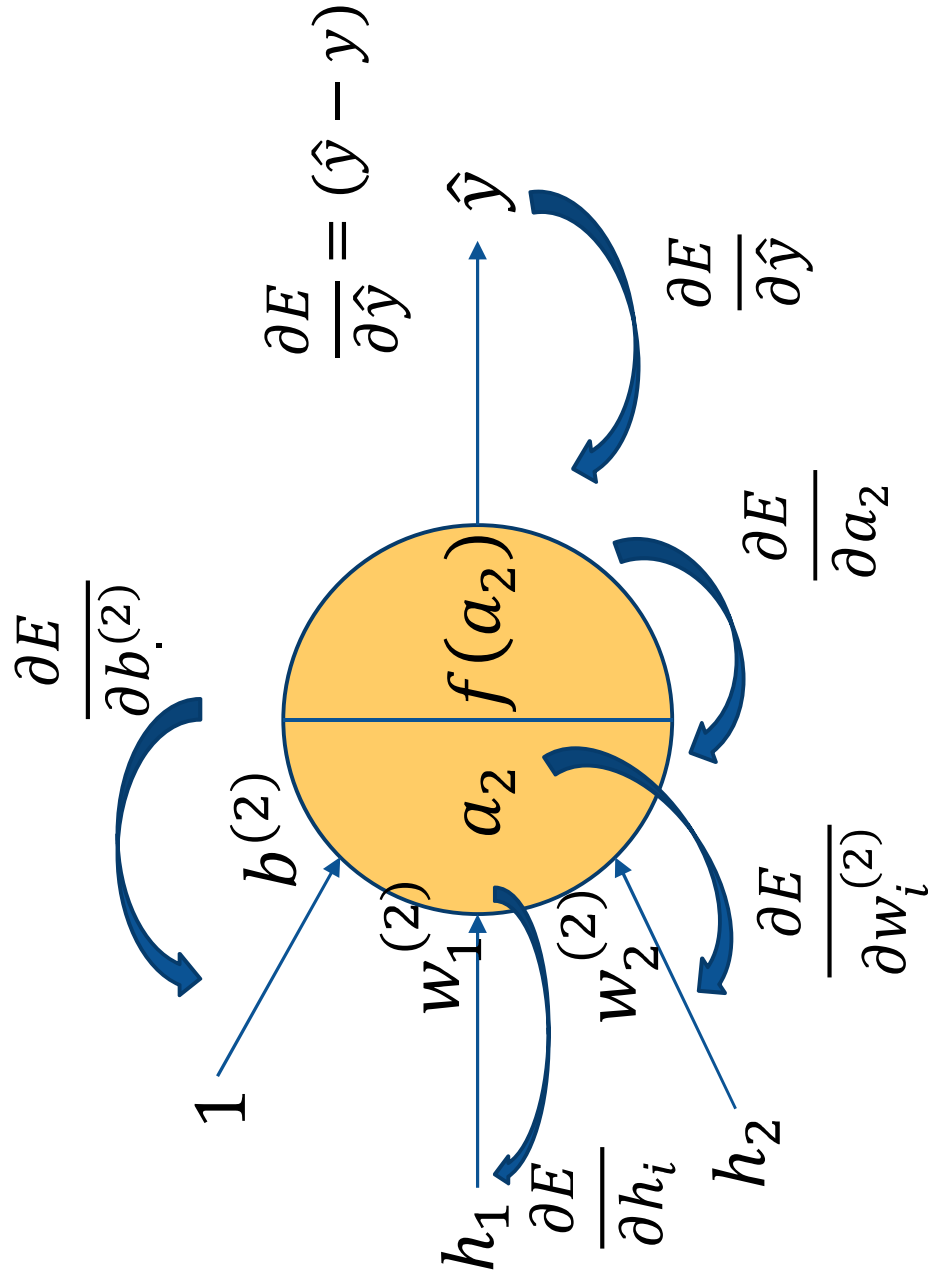$$a_2 = w_1^{(2)} h_1 + w_2^{(2)} h_2 + b^{(2)}$$

Backpropagation

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial W}$$

Backpropagation

$$\frac{\partial E}{\partial a_2} = \frac{\partial E}{\partial \hat{y}} f' a_2$$

$$\frac{\partial E}{\partial w_i^{(2)}} = \frac{\partial E}{\partial a_2} h_i$$

$$\frac{\partial E}{\partial b^{(2)}} = \frac{\partial E}{\partial a_2}$$

$$\frac{\partial E}{\partial h_i} = \frac{\partial E}{\partial a_2} w_i^{(2)}$$

$$\frac{\partial E}{\partial \hat{y}} = (\hat{y} - y)$$

$f(a_2)$

a_2

$b^{(2)}$

$w_1^{(2)}$

$w_2^{(2)}$

h_1

h_2

$\hat{y}$

$\frac{\partial E}{\partial b^{(2)}}$

$\frac{\partial E}{\partial \hat{y}}$

$\frac{\partial E}{\partial a_2}$

$\frac{\partial E}{\partial w_i^{(2)}}$

$\frac{\partial E}{\partial h_i}$

1

Parameters' Update

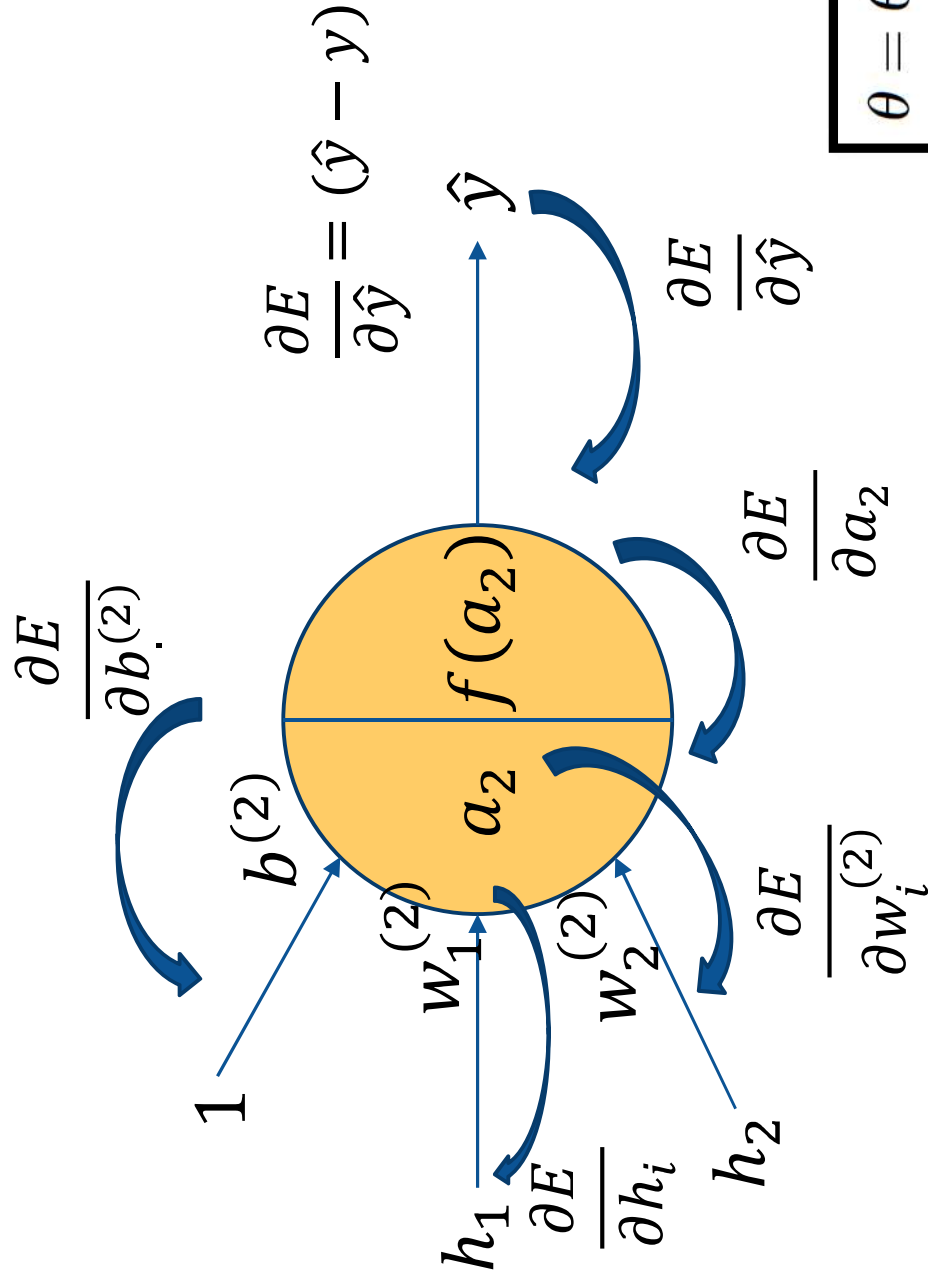$$\frac{\partial E}{\partial a_2} = \frac{\partial E}{\partial \hat{y}} f' a_2$$

$$\frac{\partial E}{\partial w_i^{(2)}} = \frac{\partial E}{\partial a_2} h_i$$

$$\frac{\partial E}{\partial b^{(2)}} = \frac{\partial E}{\partial a_2}$$

$$\frac{\partial E}{\partial h_i} = \frac{\partial E}{\partial a_2} w_i^{(2)}$$

$$\theta = \theta - \alpha \nabla_\theta E$$

$$\frac{\partial E}{\partial \hat{y}} = (\hat{y} - y)$$

$$\frac{\partial E}{\partial \hat{y}}$$

$$\hat{y}$$

$$f(a_2)$$

$$a_2$$

$$\frac{\partial E}{\partial a_2}$$

$$\frac{\partial E}{\partial b^{(2)}}$$

$$b^{(2)}$$

$$1$$

$$w_1^{(2)}$$

$$w_2^{(2)}$$

$$h_1$$

$$\frac{\partial E}{\partial h_i}$$

$$\frac{\partial E}{\partial w_i^{(2)}}$$

$$h_2$$

Summary

- Training a multi-layer perceptron involves four steps
  - Forward propagation of the input values to the output via the connections to the activation functions
  - Error computation
  - Backpropagation based on the chain rule
  - Parameters' update

References

- Ravi et al. Deep Learning for Health Informatics, IEEE Journal of Biomedical and Health Informatics, 21(1), 2017

- Kamath, Deep Learning for NLP Applications, Springer, 2019

- Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, O'Reilly, 2019