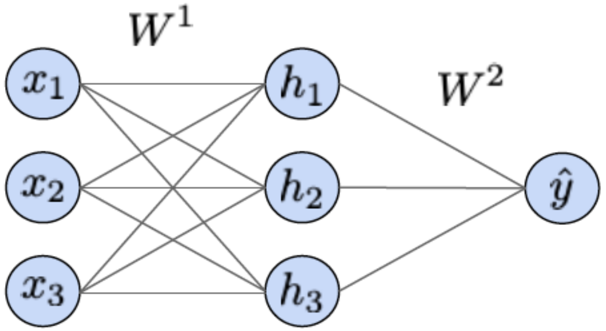
Multi-Layer Neural Networks

Shallow Neural Network (1-Hidden Layer)

Schematic Representation



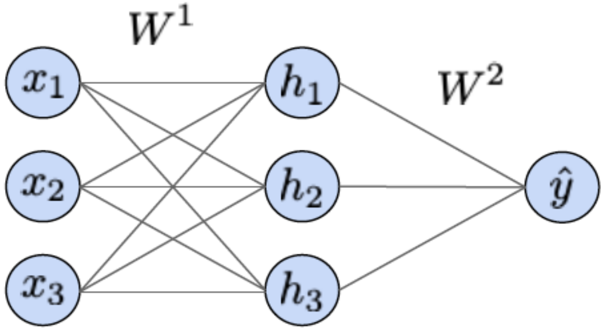
Representation of Parameters

$$W^1 = \begin{bmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 \\ w_{21}^1 & w_{22}^1 & w_{23}^1 \\ w_{31}^1 & w_{32}^1 & w_{33}^1 \end{bmatrix} \quad b^1 = \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \end{bmatrix}$$

$$W^2 = \begin{bmatrix} w_{11}^2 & w_{12}^2 & w_{13}^2 \end{bmatrix} \quad b^2 = \begin{bmatrix} b_1^2 \end{bmatrix}$$

Shallow Neural Network (1-Hidden Layer)

Schematic Representation



Representation of Parameters

$$W^1 = \begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \end{bmatrix} \quad b^1 = \begin{bmatrix} b^1_1 \\ b^1_2 \\ b^1_3 \end{bmatrix}$$

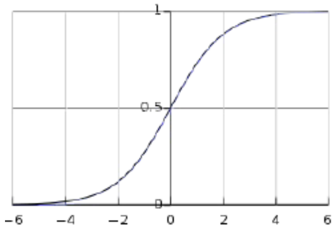
$$W^2 = \begin{bmatrix} w^2_{11} & w^2_{12} & w^2_{13} \end{bmatrix} \quad b^2 = \begin{bmatrix} b^2_1 \end{bmatrix}$$

Forward Pass:

$$h = \sigma(W^1 x + b^1)$$
$$\hat{y} = W^2 h + b^2$$

$$\hat{y} = W^2(\sigma(W^1 x + b^1)) + b^2$$

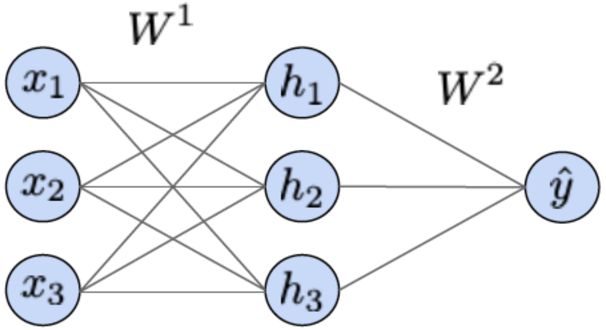
$$\sigma(a) = \frac{1}{1 + e^{-a}}$$



Activation Function
(Sigmoid)

Shallow Neural Network (1-Hidden Layer)

Schematic Representation



Parameters to be Learnt

$$W^1 = \begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \end{bmatrix} \quad b^1 = \begin{bmatrix} b^1_1 \\ b^1_2 \\ b^1_3 \end{bmatrix}$$

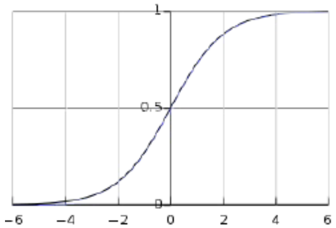
$$W^2 = \begin{bmatrix} w^2_{11} & w^2_{12} & w^2_{13} \end{bmatrix} \quad b^2 = \begin{bmatrix} b^2_1 \end{bmatrix}$$

Forward Pass:

$$h = \sigma(W^1 x + b^1)$$
$$\hat{y} = W^2 h + b^2$$

$$\hat{y} = W^2(\sigma(W^1 x + b^1)) + b^2$$

$$\sigma(a) = \frac{1}{1 + e^{-a}}$$



Activation Function (Sigmoid)

Loss Function

$$L(W^1, b^1, W^2, b^2) = \frac{1}{m} \sum (y_i - \hat{y}_i)^2$$

Backpropagation (Chain Rule)

$$z = f(y)$$
$$y = g(x)$$
$$Then$$
$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

Shallow Neural Network Equations

Forward Pass:

$$Z_1 = W_1 X + b_1$$

$$A_1 = \sigma(Z_1)$$

$$\hat{y} = W_2 A_1 + b_2$$

$$L = (\hat{y} - y)(\hat{y} - y)^T$$

Shallow Neural Network Equations

Forward Pass:

$$Z_1 = W_1 X + b_1$$

$$A_1 = \sigma(Z_1)$$

$$\hat{y} = W_2 A_1 + b_2$$

$$L = (\hat{y} - y)(\hat{y} - y)^T$$

Gradients for Backward Pass:

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)/m$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial \hat{y}}{\partial b_2}\frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial \hat{y}}{\partial W_2}\frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial A_1} = \frac{\partial \hat{y}}{\partial A_1}\frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial Z_1} = \frac{\partial L}{\partial A_1}\sigma'(Z_1)$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial W_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial b_1}$$

Shallow Neural Network Equations

Forward Pass:

$$Z_1 = W_1 X + b_1$$

$$A_1 = \sigma(Z_1)$$

$$\hat{y} = W_2 A_1 + b_2$$

$$L = (\hat{y} - y)(\hat{y} - y)^T$$

Gradients for Backward Pass:

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)/m$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial \hat{y}}{\partial b_2}\frac{\partial L}{\partial \hat{y}} = sum(\frac{2(\hat{y} - y)}{m})$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial \hat{y}}{\partial W_2}\frac{\partial L}{\partial \hat{y}} = \frac{2(\hat{y} - y)}{m} A_1^T$$

$$\frac{\partial L}{\partial A_1} = \frac{\partial \hat{y}}{\partial A_1}\frac{\partial L}{\partial \hat{y}} = W_2^T \frac{2(\hat{y} - y)}{m}$$

$$\frac{\partial L}{\partial Z_1} = \frac{\partial L}{\partial A_1}\sigma'(Z_1) = \frac{\partial L}{\partial A_1} A_1(1 - A_1)$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial W_1} = \frac{\partial L}{\partial Z_1} X^T$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial b_1} = sum(\frac{\partial L}{\partial Z_1})$$

Shallow Neural Network Equations

Forward Pass:

$$Z_1 = W_1 X + b_1$$

$$A_1 = \sigma(Z_1)$$

$$\hat{y} = W_2 A_1 + b_2$$

$$L = (\hat{y} - y)(\hat{y} - y)^T$$

Gradients for Backward Pass:

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)/m$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial \hat{y}}{\partial b_2}\frac{\partial L}{\partial \hat{y}} = sum(\frac{2(\hat{y} - y)}{m})$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial \hat{y}}{\partial W_2}\frac{\partial L}{\partial \hat{y}} = \frac{2(\hat{y} - y)}{m} A_1^T$$

$$\frac{\partial L}{\partial A_1} = \frac{\partial \hat{y}}{\partial A_1}\frac{\partial L}{\partial \hat{y}} = W_2^T \frac{2(\hat{y} - y)}{m}$$

$$\frac{\partial L}{\partial Z_1} = \frac{\partial L}{\partial A_1}\sigma'(Z_1) = \frac{\partial L}{\partial A_1} A_1(1 - A_1)$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial W_1} = \frac{\partial L}{\partial Z_1} X^T$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1}\frac{\partial Z_1}{\partial b_1} = sum(\frac{\partial L}{\partial Z_1})$$

Gradients Descent Step:

$$W_1 = W_1 - \alpha \frac{\partial L}{\partial W_1}$$

$$b_1 = b_1 - \alpha \frac{\partial L}{\partial b_1}$$

$$W_2 = W_2 - \alpha \frac{\partial L}{\partial W_2}$$

$$b_2 = b_2 - \alpha \frac{\partial L}{\partial b_2}$$