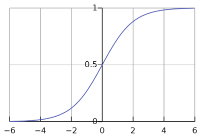
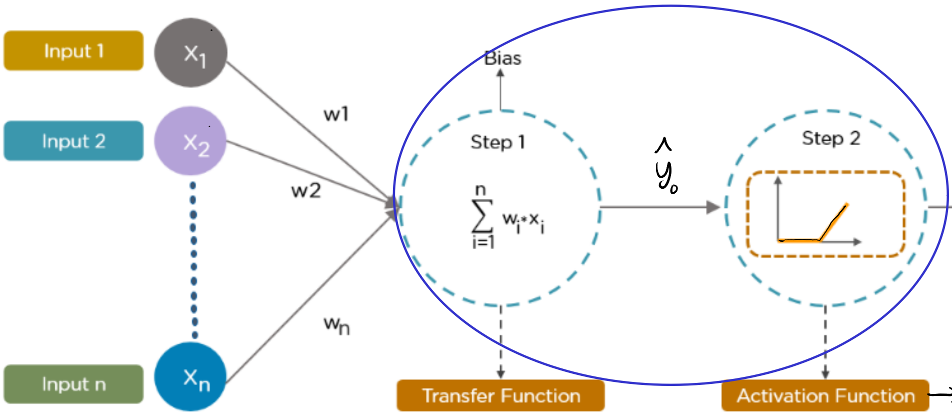
Forward Pass



Tanh
Sigmoid σ
Relu
Leaky Relu
Gelu

2.1 Neuron or perceptron



$\rightarrow g(\cdot)$

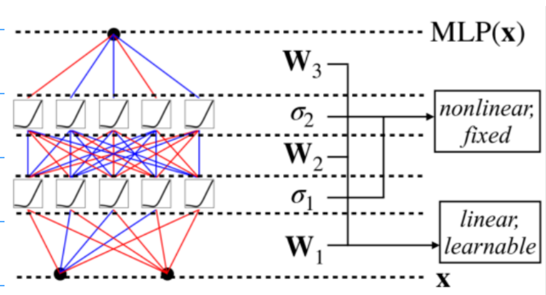
$$\hat{y}_o = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b = \sum_{i=1}^{n} w_i x_i + b$$

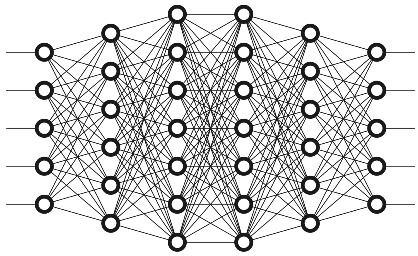
$$\hat{y}_o = w_1x_1 + b$$

$$y_o = mx + c$$

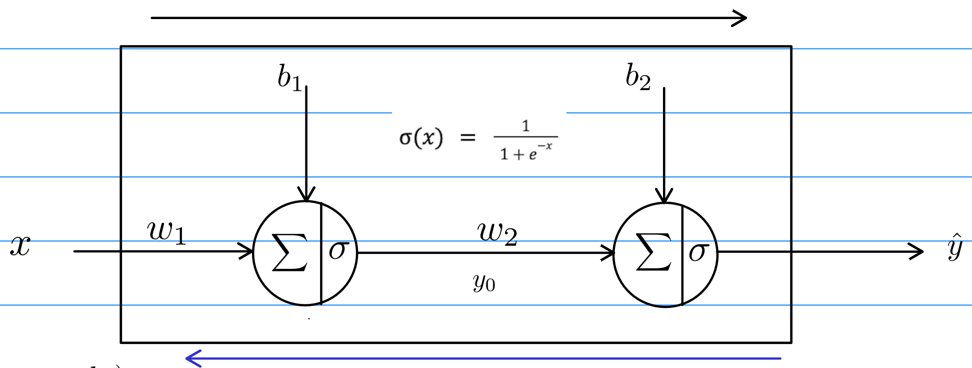
$$\hat{Y} = g\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

2.2 Neural Network or Multi Layer Perceptron





$$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$$

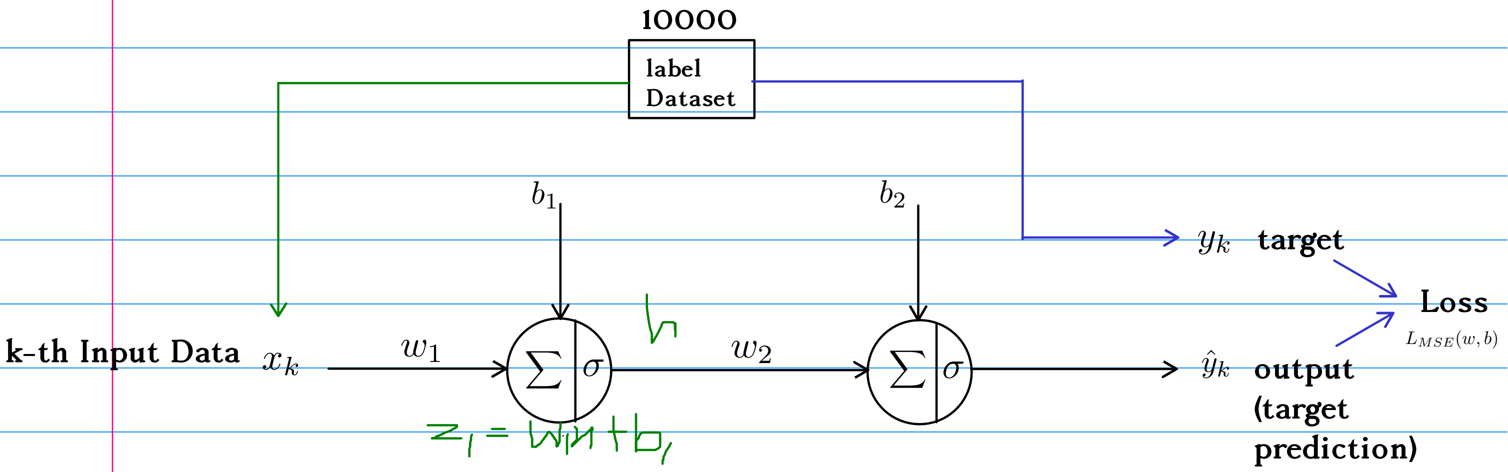


$$y_0 = \sigma_1(w_1x + b_1)$$

$$\hat{y} = \sigma_2(w_2y_0 + b_2)$$

$$\hat{y} = \sigma_2(w_2(\sigma_1(w_1x + b_1)) + b_2)$$

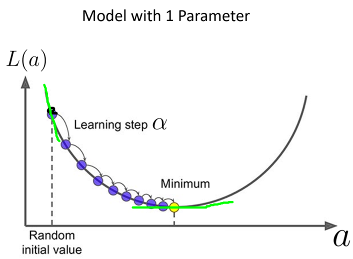
Backward Pass

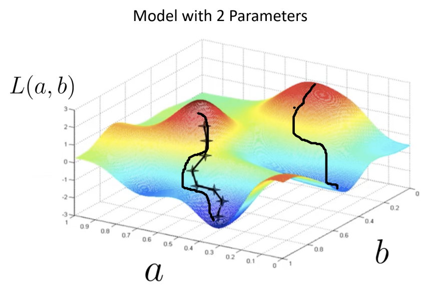


$$\hat{y} = \sigma_2(w_2(\sigma_1(w_1x + b_1)) + b_2)$$

Mean Squared Error (MSE) Loss

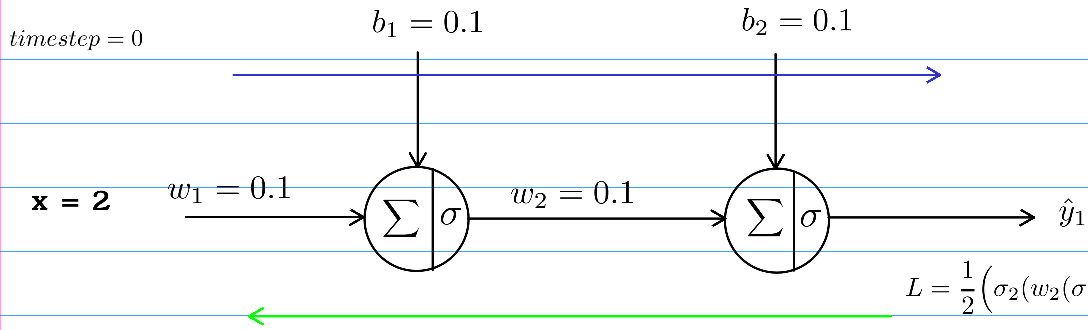
$$L_{MSE}(w,b) = \frac{1}{n}\Sigma_{i=1}^{n}(\hat{y}_i - y_i)^2$$





$$w_{t+1} = w_t - \alpha\frac{\partial L}{\partial w_t} \qquad \alpha = 0.001$$

Backpropagation



$$L = \frac{1}{2}(\hat{y} - t)^2$$

$$L = \frac{1}{2}\Big(\sigma_2(w_2(\sigma_1(w_1x + b_1)) + b) - t\Big)^2$$

$$L(w_1, b_1, w_2, b_2)$$

$$\frac{\partial L(w_1, b_1, w_2, b_2)}{\partial w_1} = \frac{\partial L}{\partial w_1}$$

$$w_{1,t+1} = w_{1,t} - \alpha\frac{\partial L}{\partial w_1} \qquad b_{1,t+1} = b_{1,t} - \alpha\frac{\partial L}{\partial b_1}$$

$$w_{2,t+1} = w_{2,t} - \alpha\frac{\partial L}{\partial w_2} \qquad b_{2,t+1} = b_{2,t} - \alpha\frac{\partial L}{\partial b_2}$$

Chain rule

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$\frac{dy}{dx}$ **derivative of y with respect to x**

$\frac{dy}{du}$ **derivative of y with respect to u**

$\frac{du}{dx}$ **derivative of u with respect to x**

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h_2}\frac{\partial h_2}{\partial h_1}\frac{\partial h_1}{\partial w_1}$$

$$L = \frac{1}{2}(y-t)^2 = \frac{1}{2}\Big(\sigma_2(w_2(\sigma_1(w_1 x + b_1)) + b) - t\Big)^2 = L(w_1, b_1, w_2, b_2)$$

$$\frac{\partial L}{\partial y} = y - t$$

$$\hat{y} = \sigma_2(w_2(\sigma_1(w_1 x + b_1)) + b_2)$$

$$y = \sigma(z_2) = \frac{1}{1+e^{-z_2}}$$

$$z_2 = w_2 h + b_2$$

$$h = \sigma(z_1)$$

$$z_1 = w_1 x + b_1$$

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

$$\sigma'(x) = \frac{d}{dx}\sigma(x) = \sigma(x)(1-\sigma(x))$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial w_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial w_2} = (y-t)y(1-y)h$$

$$w_{2,t+1} = w_{2,t} - \alpha\frac{\partial L}{\partial w_2}$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial b_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z_2}\frac{\partial z_2}{\partial b_2} = (y-t)y(1-y)w_2$$

$$b_{2,t+1} = b_{2,t} - \alpha\frac{\partial L}{\partial b_2}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial w_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z_1}\frac{\partial z_1}{\partial w_1} = (y-t)y(1-y)w_2 h(1-h)x$$

$$w_{1,t+1} = w_{1,t} - \alpha\frac{\partial L}{\partial w_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial b_1} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial z_1}\frac{\partial z_1}{\partial b_1} = (y-t)y(1-y)w_2 h(1-h)$$

$$b_{1,t+1} = b_{1,t} - \alpha\frac{\partial L}{\partial b_1}$$