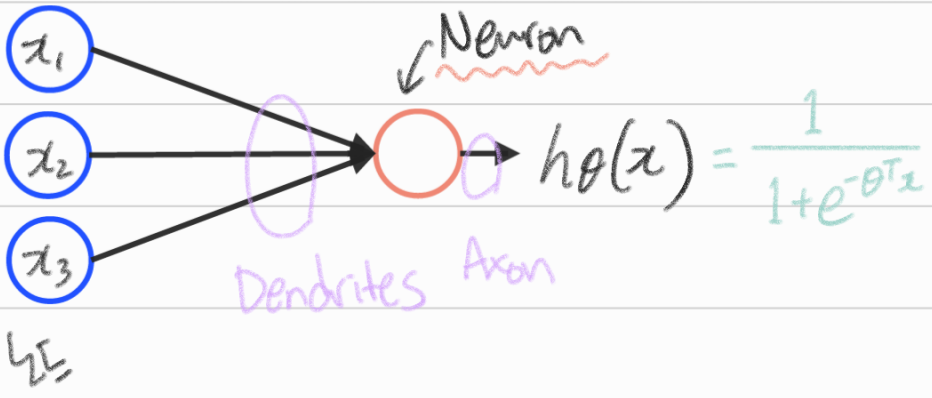
Model Representation I

∴ Neuron : Computational unit

Neuron model : Logistic unit

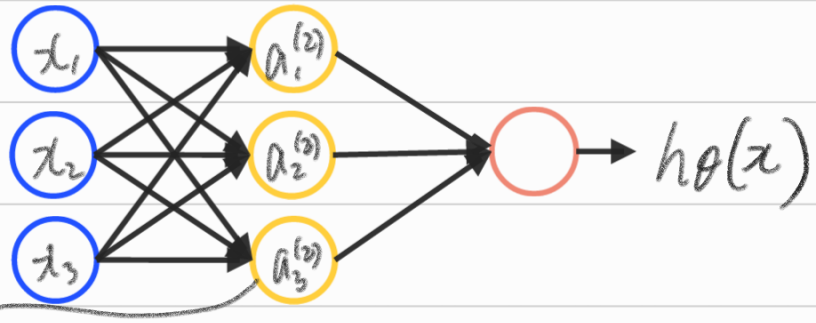


뉴런

Sigmoid (logistic) activation function

θ theta = Weights 가중치

Neural Network



Layer1 Layer2 Layer3

Input Layer Hidden Layer Output Layer

x y

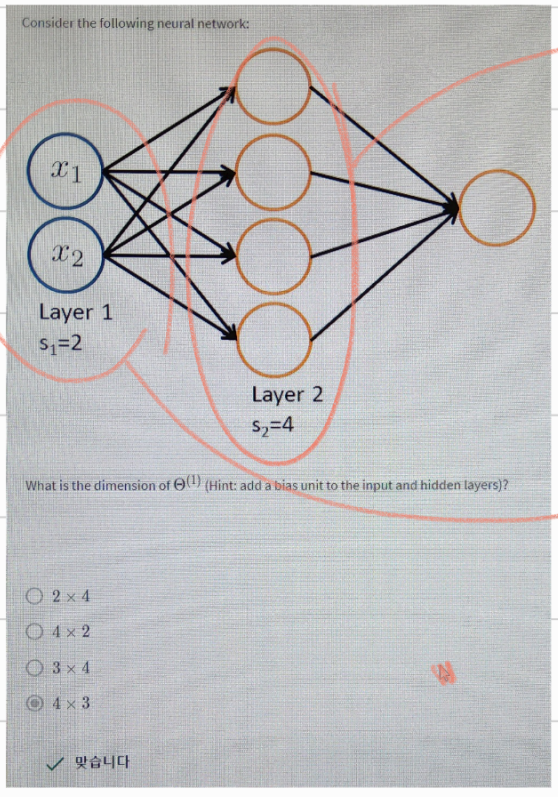
$a_i^{(j)}$: "activation" of unit i in layer j

output value computed by unit

$\Theta^{(j)}$: matrix of weights controlling function mapping from layer j to layer $j+1$

$$a_1^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3)$$
$$a_2^{(2)} = g(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3)$$
$$a_3^{(2)} = g(\Theta_{30}^{(1)} x_0 + \Theta_{31}^{(1)} x_1 + \Theta_{32}^{(1)} x_2 + \Theta_{33}^{(1)} x_3)$$
$$h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)})$$

If network has s_j (3, 1) units in layer j (1),
s_{j+1} (3, 2) units in layer $j+1$ (2),
then $\Theta^{(j)}$ (1) will be of dimension $s_{j+1} \times (s_j + 1)$ (3, 2 × 4)



$4 \times (2+1)$

Model Representation Ⅱ

Forward propagation : Vectorized implementation

(layer 2)

$$a_1^{(2)} = g(z_1^{(2)})$$
$$a_2^{(2)} = g(z_2^{(2)}) \quad \} z^{(2)}$$
$$a_3^{(3)} = g(z_3^{(2)})$$

선형결합

z value is a linear combination of input values x_0, x_1, x_2, x_3

$$\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3$$
$$\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3$$
$$\theta_{30}^{(1)} x_0 + \theta_{31}^{(1)} x_1 + \theta_{32}^{(1)} x_2 + \theta_3^{(1)} x_3$$

$\therefore \Theta^{(1)} \cdot x$

(able to vectorize this computation)

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix}$$ 3차원 vector

$z^{(2)} = \Theta^{(1)} \cdot x$

$a^{(2)} = g(z^{(2)})$ Sigmoid f.

$a^{(2)}$: 3차원 vector, $g(z^{(2)})$: 3차원 vector

Add $a_0^{(2)} = 1$

⇓

4차원 vector

$z^{(3)} = \Theta^{(2)} a^{(2)}$

$h_\theta(x) = a_3 = g(z^{(3)})$

$h_\theta(x) = a_1^{(3)} = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})$

Layer가 진행되면 it learned from its own features

신경망 형태를

Network Architectures라 명명