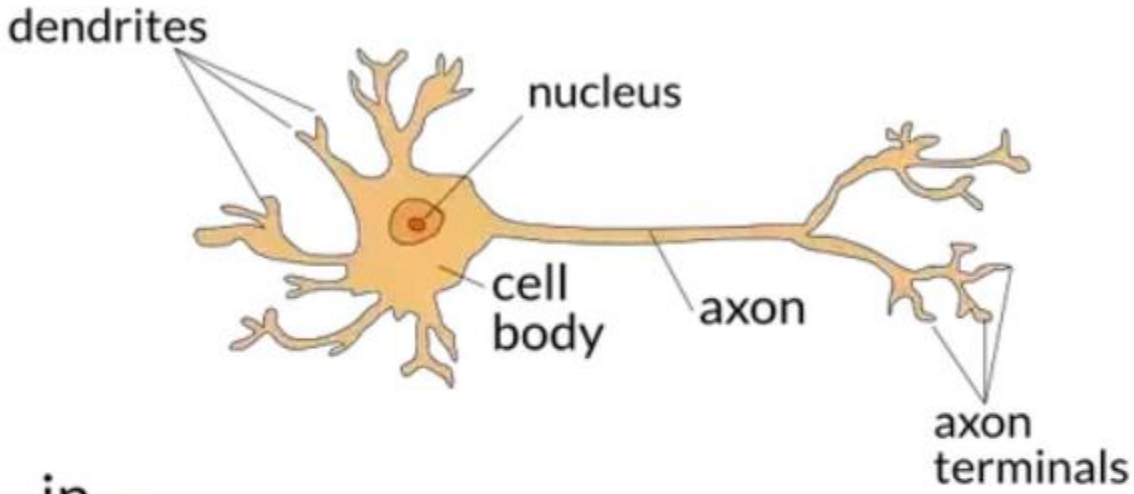
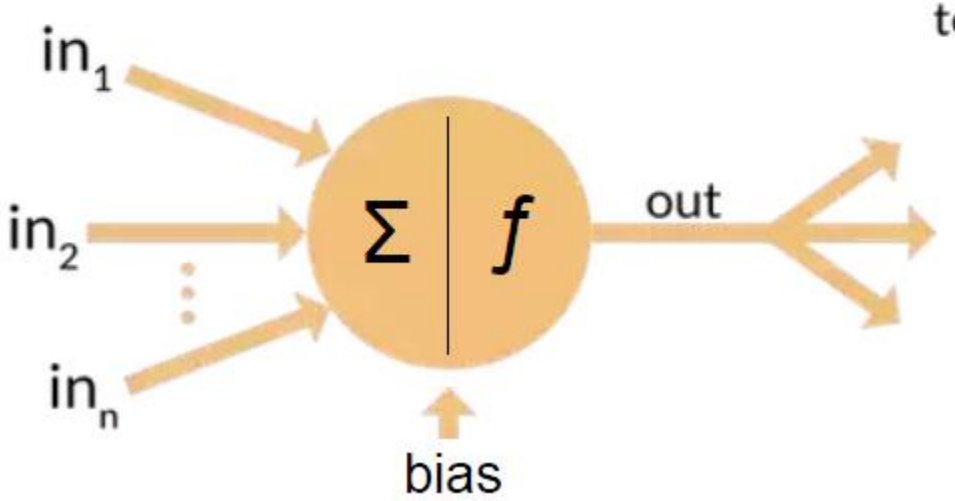
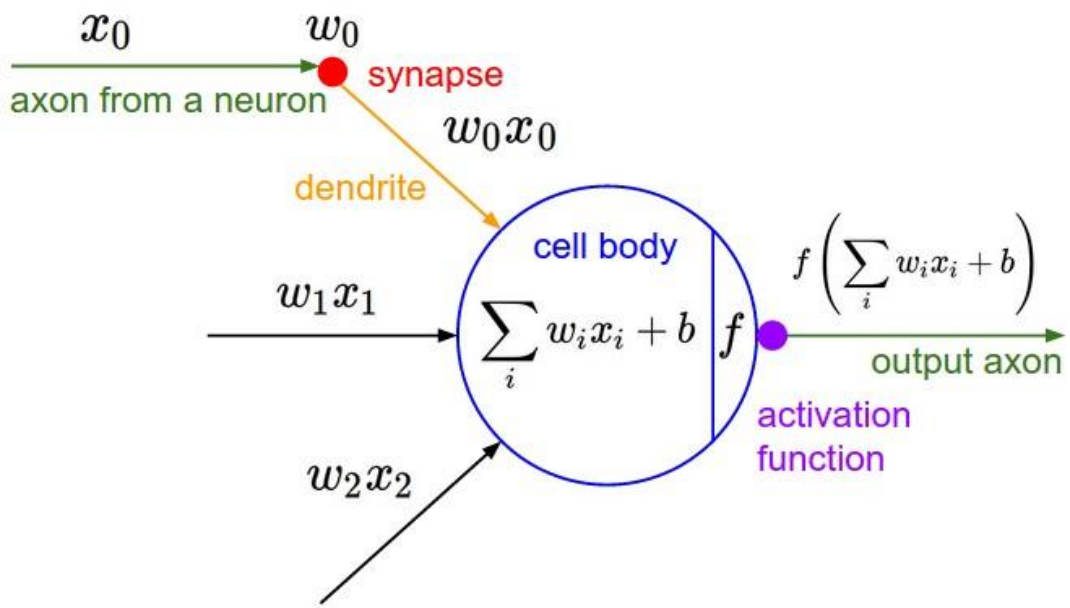
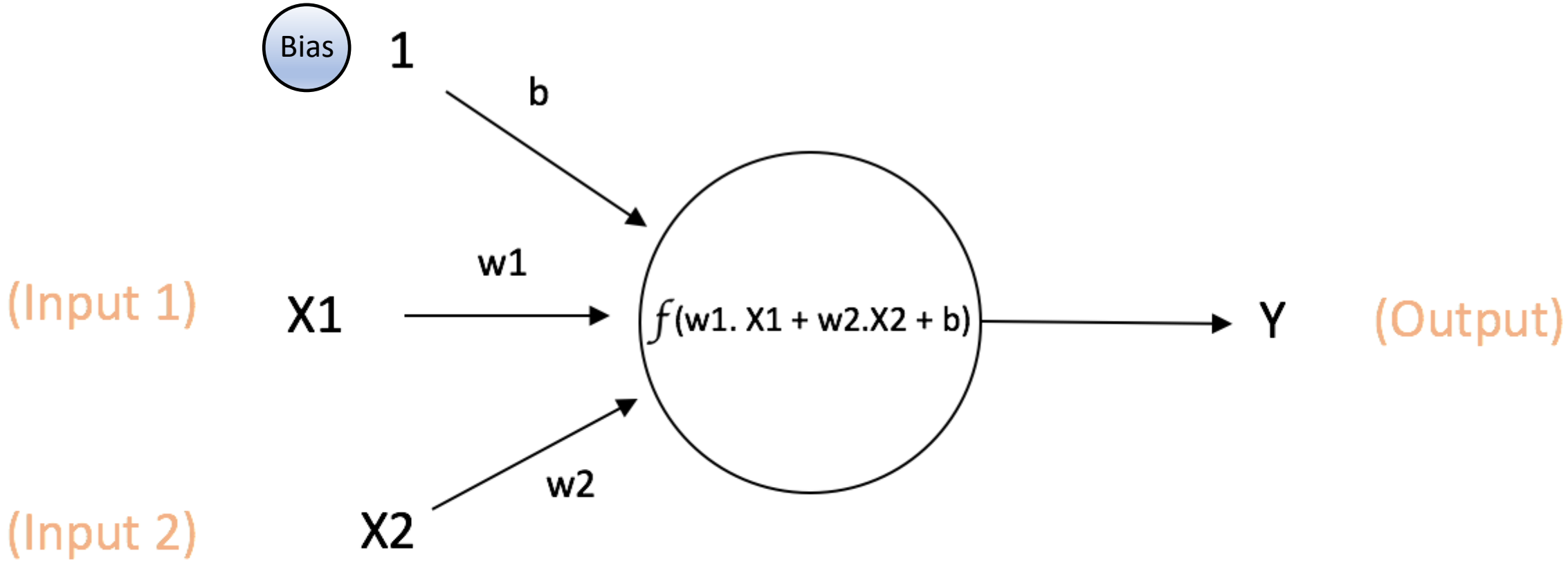

dendrites
nucleus
cell body
axon
axon terminals

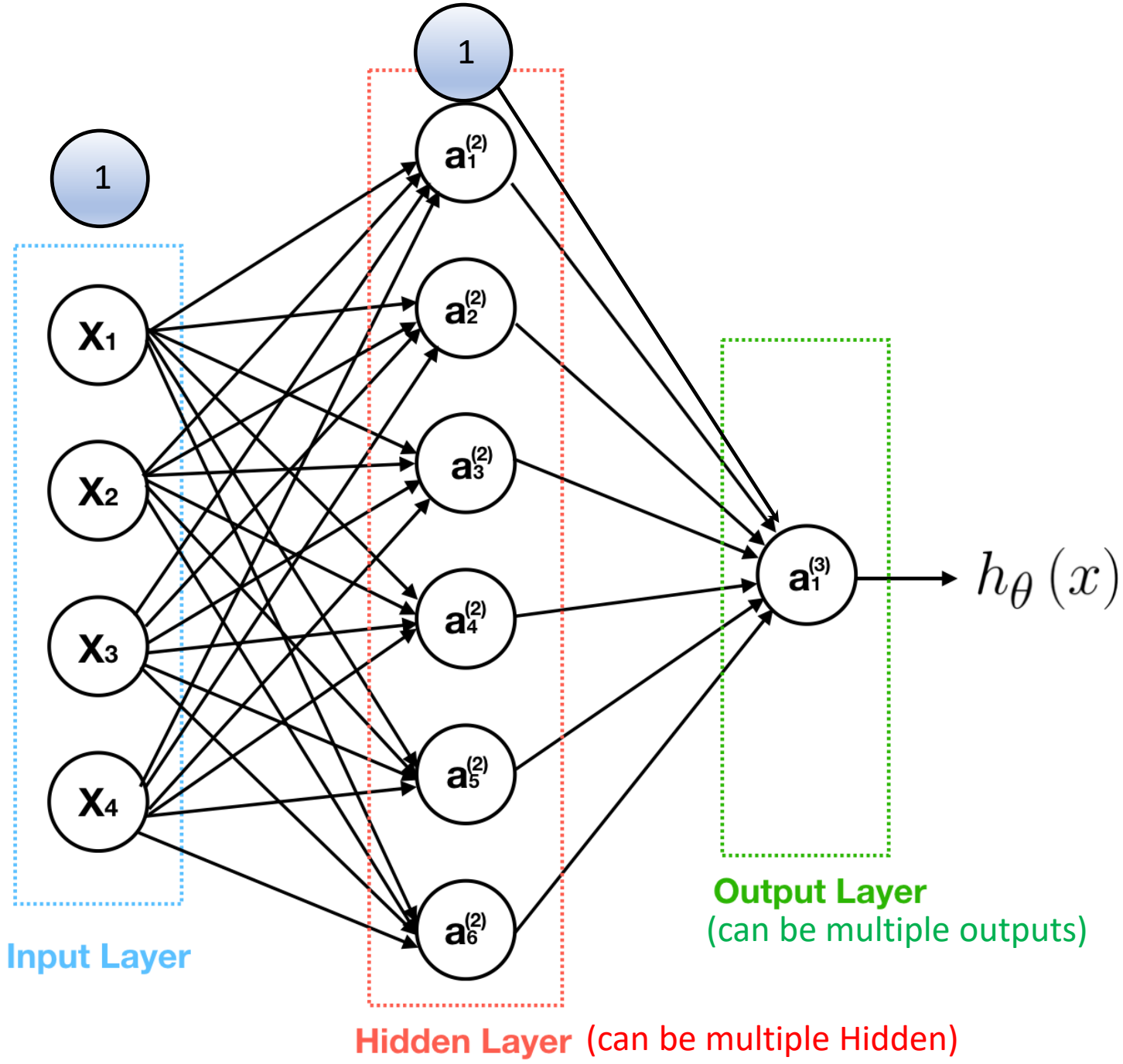

in1
in2
inn
Σ
f
out
bias


x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$w_1 x_1$
$\sum_i w_i x_i + b$
f
$f\left(\sum_i w_i x_i + b\right)$
output axon
activation function
$w_2 x_2$

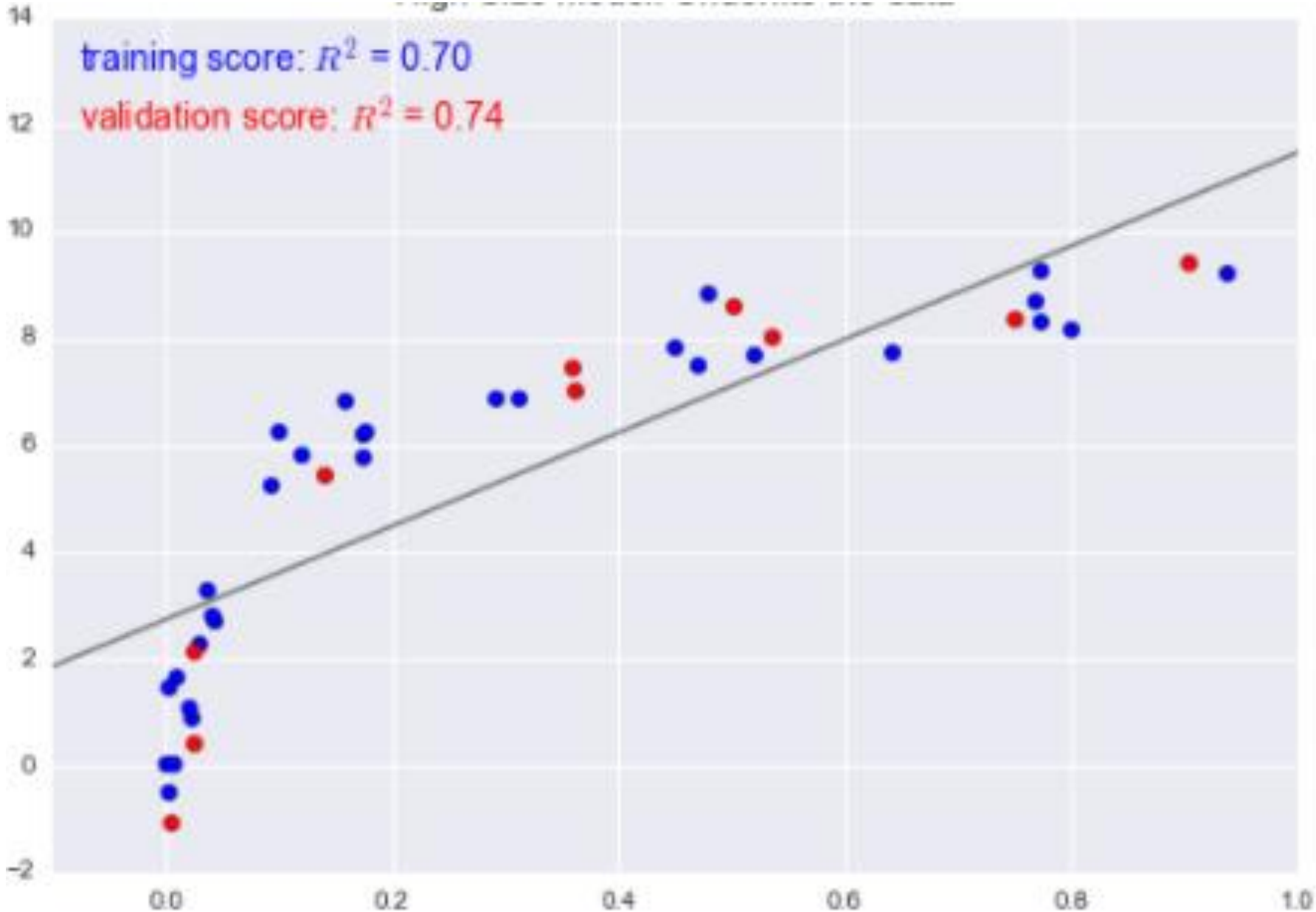


Output of neuron = Y= f(w1. X1 + w2.X2 + b)

Neural Network Components



Still remember?



Linear Regression

Model definition

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad \epsilon \sim N(0, \sigma_\epsilon^2)$$

Estimated (fitted) model

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Residual errors

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i$$

$$Y_i = \hat{Y}_i + \hat{\epsilon}_i$$

Logistic regression equation :

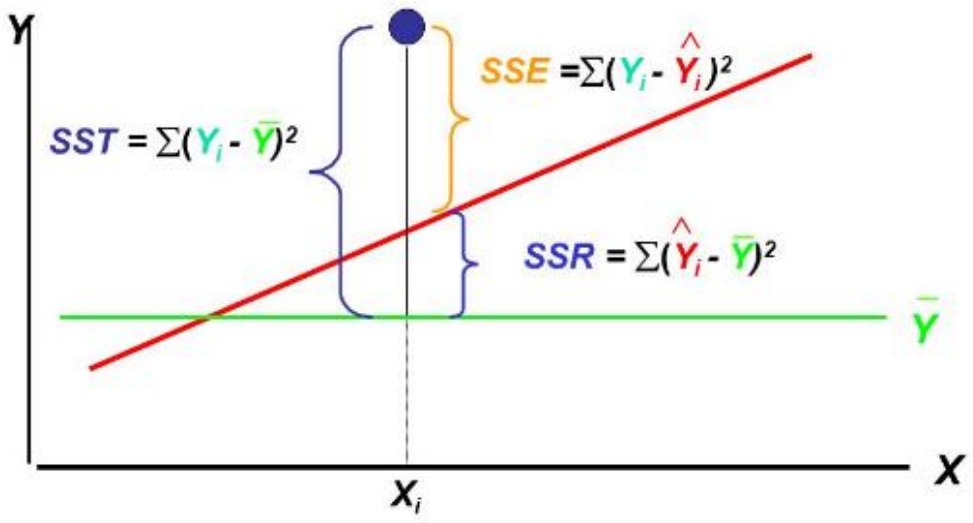
Linear regression $Y = b_0 + b_1 \times X_1 + b_2 \times X_2 \ldots\ldots + b_K \times X_K$

Sigmoid Function $P = \frac{1}{1+e^{-Y}}$

By putting Y in Sigmoid function, we get the following result.

$$\ln\left(\frac{P}{1-P}\right) = b_0 + b_1 \times X_1 + b_2 \times X_2 \ldots\ldots + b_K \times X_K$$

Model Fit, R-square

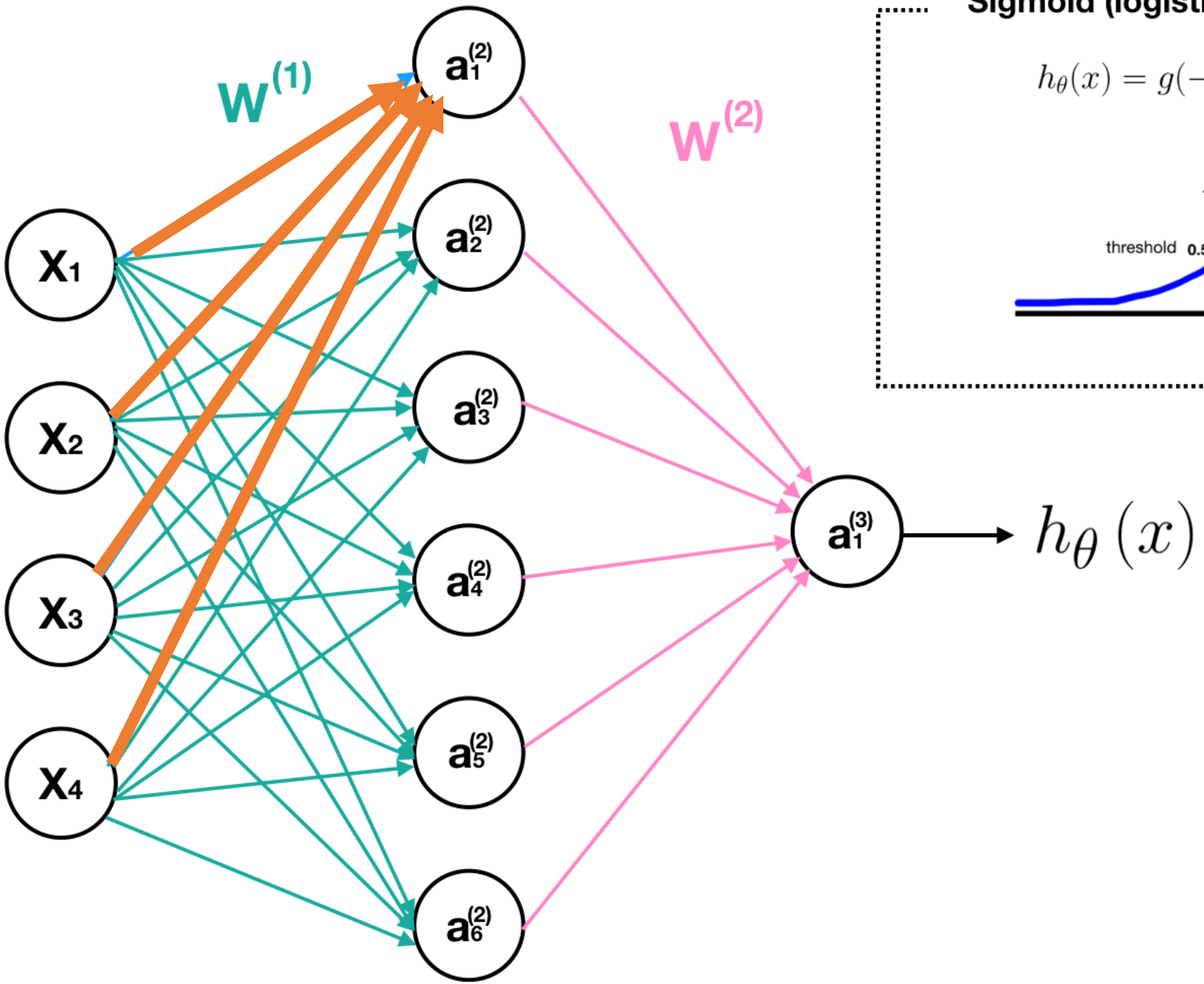


P :: a(1) $h_\theta(x)$

b_0 :: Bias

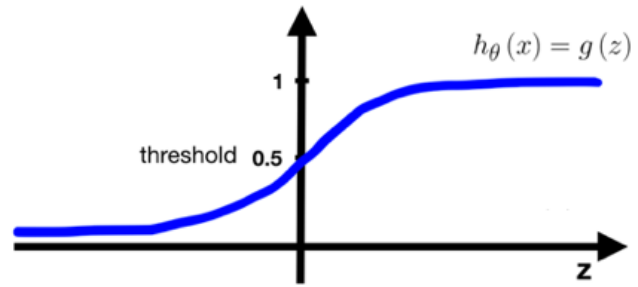
b_k :: $w^{(1)}$ $w^{(2)}$ θ

SSE :: J(b) Loss Functions

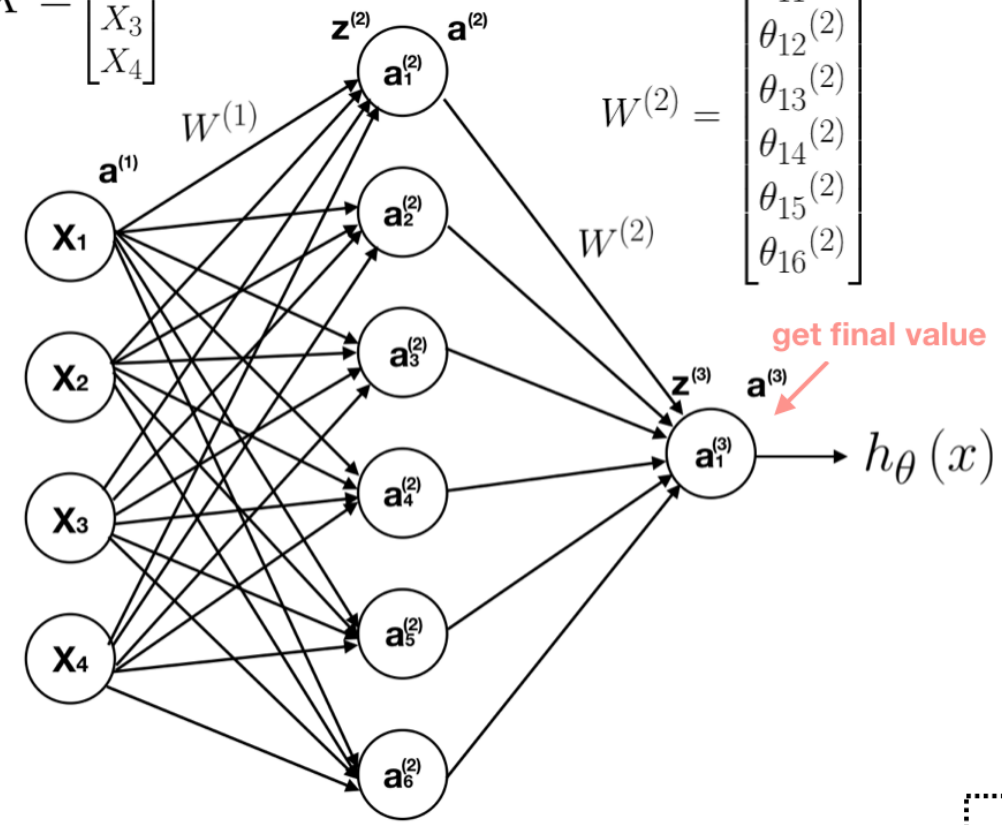


Sigmoid (logistic) Activation Function

$$h_\theta(x) = g(-\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$



$$a^{(1)} = X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$



$$W^{(2)} = \begin{bmatrix} \theta_{11}^{(2)} \\ \theta_{12}^{(2)} \\ \theta_{13}^{(2)} \\ \theta_{14}^{(2)} \\ \theta_{15}^{(2)} \\ \theta_{16}^{(2)} \end{bmatrix}$$

$$W^{(1)} = \begin{bmatrix} \theta_{11}^{(1)} & \theta_{21}^{(1)} & \theta_{31}^{(1)} & \theta_{41}^{(1)} & \theta_{51}^{(1)} & \theta_{61}^{(1)} \\ \theta_{12}^{(1)} & \theta_{22}^{(1)} & \theta_{32}^{(1)} & \theta_{42}^{(1)} & \theta_{52}^{(1)} & \theta_{62}^{(1)} \\ \theta_{13}^{(1)} & \theta_{23}^{(1)} & \theta_{33}^{(1)} & \theta_{43}^{(1)} & \theta_{53}^{(1)} & \theta_{63}^{(1)} \\ \theta_{14}^{(1)} & \theta_{24}^{(1)} & \theta_{34}^{(1)} & \theta_{44}^{(1)} & \theta_{54}^{(1)} & \theta_{64}^{(1)} \end{bmatrix}$$

Forward Propagation

$$W^{(1)^T} X = z^{(2)}$$

$$a^{(2)} = g(z^{(2)})$$

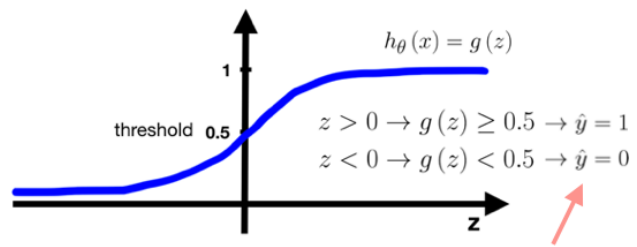
$$W^{(2)^T} a^{(2)} = z^{(3)}$$

$$a^{(3)} = g(z^{(3)})$$

Final value, for prediction

Sigmoid (logistic) Activation Function

$$h_\theta(x) = g(-\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$



Prediction

Table 4.1 Choosing the right last-layer activation and loss function for your model

Problem type	Last-layer activation	Loss function
Binary classification	`sigmoid`	`binary_crossentropy`
Multiclass, single-label classification	`softmax`	`categorical_crossentropy`
Multiclass, multilabel classification	`sigmoid`	`binary_crossentropy`
Regression to arbitrary values	None	`mse`
Regression to values between 0 and 1	`sigmoid`	`mse` or `binary_crossentropy`

Source: Francois Chollet