# Machine Learning

## Seminar 5 Solution

[Q1 Sample Solution]

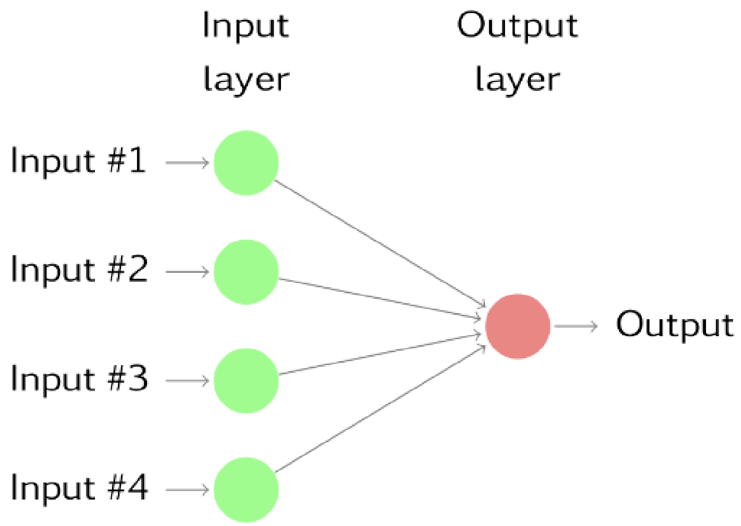
The input layer of a neural network is composed of artificial input neurons. It brings the initial data into the system for further processing by subsequent layers of artificial neurons. The input layer is the very beginning of the workflow for the artificial neural network.

The second type of layer is called the hidden layer. Hidden layers are either one or more in number for a neural network. Hidden layers are the ones that are actually responsible for the excellent performance and complexity of neural networks. They perform multiple functions at the same time such as data transformation, automatic feature creation, etc.

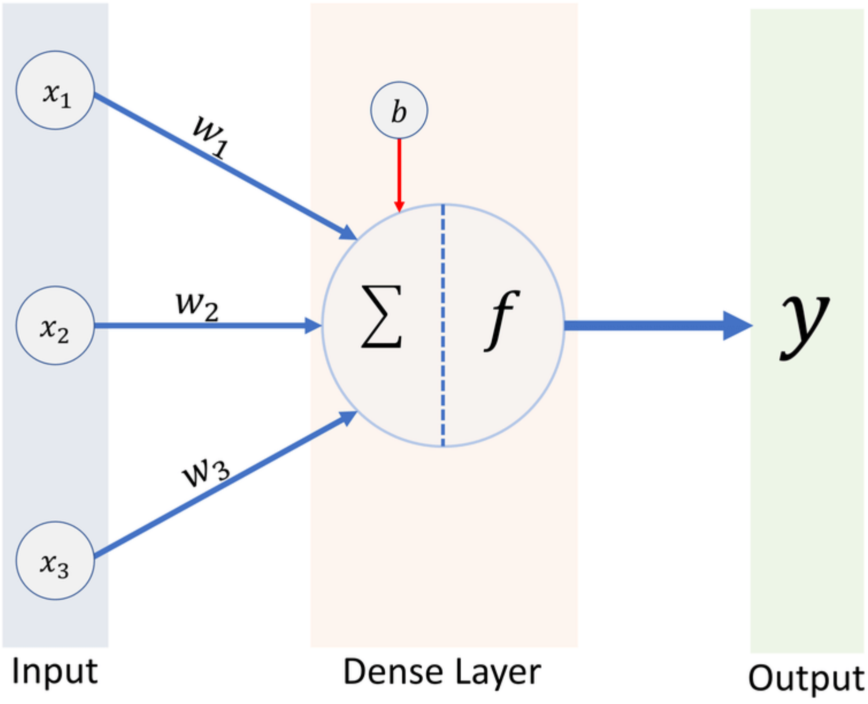
The output layer in an artificial neural network is the last layer of neurons that produces given outputs for the program. Though they are made much like other artificial neurons in the neural network, output layer neurons may be built or observed in a different way, given that they are the last “actor” nodes on the network.

A layer consists of small individual units called neurons.  A neuron in a neural network can be better understood with the help of biological neurons. An artificial neuron is similar to a biological neuron. It receives input from the other neurons, performs some processing, and produces an output.

[Q2 Sample Solution]



[Q3 Sample Solution]




$$z_j = b_j + \sum_{i=1}^{4} w_{i,j} x_i$$

$$s(z) = \frac{1}{1 + e^{-z}}$$

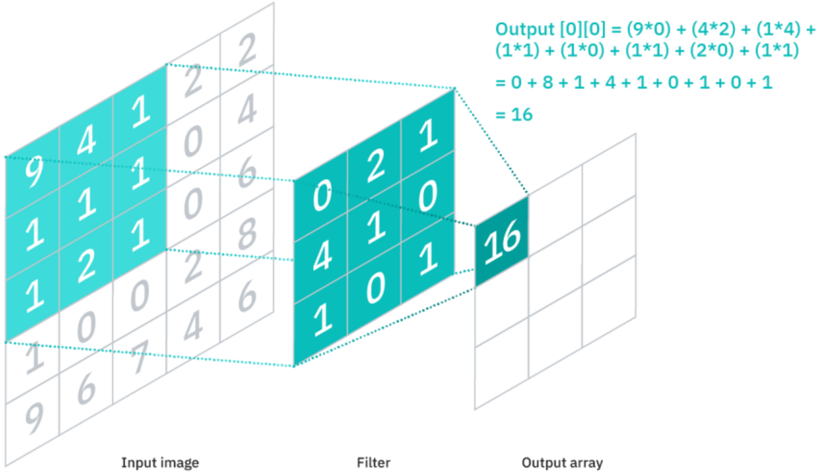
[Q4 Sample Solution]

A Convolution Layer is an important type of layer in a CNN. Its most common use is for detecting features in images, in which it uses a filter to scan an image, a few pixels at a time, and outputs a feature map that classifies each feature found.

The filter (sometimes called kernel) is a set of n-dimensional weights that are multiplied against the input. The filter's dimensions match that of the input (e.g., two dimensions when dealing with 2D images).

The filter describes the probabilities that a given pattern of pixels represents a feature. Thus the number of filter weights (i.e., size of the filter) are smaller than the input, and the multiplication performed by the layer's convolution process is performed on "patches" of the image that match the filter size.

## [Q5 Sample Solution]




## [Q6 Sample Solution]

**ANSWER (a):**
Assuming that bias is not shown in Figure and should not be included in calculation.

$$x=\begin{bmatrix}1\\5\\9\end{bmatrix}$$

$$\theta^{[1]}=\begin{bmatrix}0.01 & 0.05 & 0.09\\0.02 & 0.06 & 0.1\\0.03 & 0.07 & 0.11\\0.04 & 0.08 & 0.12\end{bmatrix}$$

$$\theta^{[2]}=\begin{bmatrix}0.17 & 0.19 & 0.21 & 0.23\\0.18 & 0.2 & 0.22 & 0.24\end{bmatrix}$$

Node values for second layer:

$$a^{[2]}=\theta^{[1]}x=\begin{bmatrix}0.01 & 0.05 & 0.09\\0.02 & 0.06 & 0.1\\0.03 & 0.07 & 0.11\\0.04 & 0.08 & 0.12\end{bmatrix}\begin{bmatrix}1\\5\\9\end{bmatrix}=\begin{bmatrix}1.07\\1.22\\1.37\\1.52\end{bmatrix}$$

where

$a_1^{(2)}$ = 0.01*1 + 0.05*5 + 0.09*9 = 0.01 + 0.25 + 0.81 = 1.07
$a_2^{(2)}$ = 0.02*1 + 0.06*5 + 0.1*9 = 0.02 + 0.3 + 0.9 = 1.22
$a_3^{(2)}$ = 0.03*1 + 0.07*5 + 0.11*9 = 0.03 + 0.35 + 0.99 = 1.37
$a_4^{(2)}$ = 0.04*1 + 0.08*5 + 0.12*9 = 0.04 + 0.4 + 1.08 = 1.52

Node values for third layer:

$$a^{[3]}=\theta^{[2]}a^{[2]}=\begin{bmatrix}0.17 & 0.19 & 0.21 & 0.23\\0.18 & 0.2 & 0.22 & 0.24\end{bmatrix}\begin{bmatrix}1.07\\1.22\\1.37\\1.52\end{bmatrix}=\begin{bmatrix}1.051\\1.1028\end{bmatrix}$$

where:

$a_1^{(3)} = 0.17*1.07 + 0.19*1.22 + 0.21*1.37 + 0.23*1.52 = 1.051$
$a_2^{(3)} = 0.18*1.07 + 0.2*1.22 + 0.22*1.37 + 0.24*1.52 = 1.1028$

**ANSWER (b):**
Assuming that no activation function is applied to node outputs, i.e. g(z) = z, and g'(z)=1

Errors for third layer:

$$\delta^{(3)} = a^{(3)} - y = \begin{bmatrix} 1.051 - 0.4886 \\ 1.1028 - 0.5114 \end{bmatrix} = \begin{bmatrix} 0.5624 \\ 0.5914 \end{bmatrix}$$

where:

$\delta_1^{(3)} = a_1^{(3)} - y_1 = 1.051 - 0.4886 = 0.5624$
$\delta_2^{(3)} = a_2^{(3)} - y_2 = 1.1028 - 0.5114 = 0.5914$

Errors for second layer:

$$\delta^{(2)} = \left[\theta^{(2)}\right]^T \delta^{(3)} = \begin{bmatrix} 0.17 & 0.18 \\ 0.19 & 0.2 \\ 0.21 & 0.22 \\ 0.23 & 0.24 \end{bmatrix} \begin{bmatrix} 0.5624 \\ 0.5914 \end{bmatrix} = \begin{bmatrix} 0.20206 \\ 0.225136 \\ 0.248212 \\ 0.271288 \end{bmatrix}$$

where:

$\delta_1^{(2)} = 0.17*0.5624 + 0.18*0.5914 = 0.20206$
$\delta_2^{(2)} = 0.19* 0.5624 + 0.2* 0.5914 = 0.225136$
$\delta_3^{(2)} = 0.21* 0.5624 + 0.22* 0.5914 = 0.248212$
$\delta_4^{(2)} = 0.23* 0.5624 + 0.24* 0.5914 = 0.271288$