

Deep Network Architecture

Advantage of Deep Network Architectures

The term **shallow neural network** typically refers networks with just a single hidden layer whereas the term **deep neural network** generally refers to networks with more than one hidden layer. Deep neural networks are also referred to as **deep-learning networks**.

In deep neural networks, each layer of nodes trains on a distinct set of features based on the previous layer's output. The deeper one advances into the neural net, the more complex the features the nodes can recognize, since they aggregate and recombine features from the previous layer. In essence, shallow networks can only detect the simplest features, whereas deep networks are able to extract more complex features. This hierarchy of increasingly complex and abstract features is known as a **feature hierarchy**. It makes deep-learning networks capable of handling very large, high-dimensional data sets with billions of parameters that pass through nonlinear functions.

As a hypothetical example, a network designed to label images might have the following:

- **Layer 1** → detects edges in pictures
- **Layer 2** → groups edges together and begins to detect different parts of the image
- **Layer 3** → groups parts recognized in previous layer into even larger parts
- **Layer 4** → ...

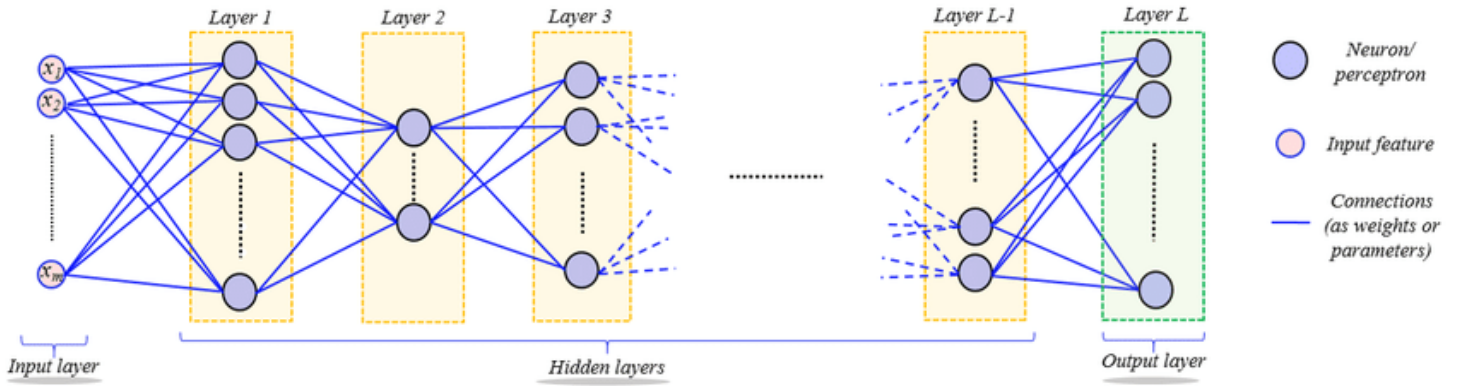
As another hypothetical example, a network designed to process audio files might have the following:

- **Layer 1** → detects low level wave features
- **Layer 2** → recognizes basic units of sounds (also known as "phonemes")
- **Layer 3** → groups phonemes from previous layer into words
- **Layer 4** → groups words from previous layer into sentences
- **Layer 5** → ...

Deep L -Layer Network Architecture

We can generalize the mathematical notation of neural networks to a arbitrarily deep network consisting of L layers.

The diagram below depicts the generalized form of a deep neural network:



In order to generalize the mathematical notation of neural networks to describe the architecture of a deep L -layer network, we must define the following:

- $n^{[l]} \equiv$ number of units in layer l
- $g^{[l]} \equiv$ activation function implemented by units in layer l

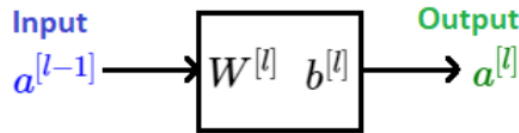
It is important to note that:

$$n^{[0]} = n_x = \text{number of features in dataset}$$

$$a^{[0]} = x = \text{input feature vector}$$

Forward Propagation

A forward pass through layer l of a deep L -layer network is shown in the function diagram below:



The **non-vectorized** equations describing forward propagation through layer l of a deep L -layer network are:

$$z^{[l]} = \mathbf{W}^{[l]} a^{[l-1]} + b^{[l]}$$

$$a^{[l]} = g^{[l]}(z^{[l]})$$

The **vectorized** equations describing forward propagation through layer l of a deep L -layer network are:

$$\mathbf{Z}^{[l]} = \mathbf{W}^{[l]} \mathbf{A}^{[l-1]} + b^{[l]}$$

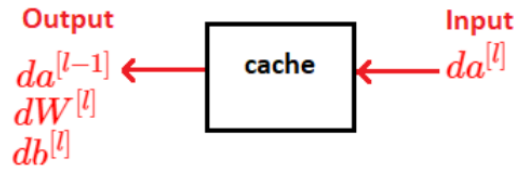
$$\mathbf{A}^{[l]} = g^{[l]}(\mathbf{Z}^{[l]})$$

The dimensions of all parameters involved in forward propagation are:

$$\begin{aligned}\mathbf{W}^{[l]} &: (n^{[l]}, n^{[l-1]}) \\ b^{[l]} &: (n^{[l]}, 1) \\ z^{[l]}, a^{[l]} &: (n^{[l]}, 1) \\ \mathbf{Z}^{[l]}, \mathbf{A}^{[l]} &: (n^{[l]}, m)\end{aligned}$$

Backpropagation

A backward pass through layer l of a deep L -layer network is shown in the function diagram below:



The **non-vectorized** equations describing forward propagation through layer l of a deep L -layer network are:

$$\begin{aligned}z^{[l]} &= \mathbf{W}^{[l]} a^{[l-1]} + b^{[l]} \\ a^{[l]} &= g^{[l]}(z^{[l]})\end{aligned}$$

The **vectorized** equations describing forward propagation through layer l of a deep L -layer network are:

$$\begin{aligned}\mathbf{Z}^{[l]} &= \mathbf{W}^{[l]} \mathbf{A}^{[l-1]} + b^{[l]} \\ \mathbf{A}^{[l]} &= g^{[l]}(\mathbf{Z}^{[l]})\end{aligned}$$

The shapes of all parameters involved in forward propagation are:

$$\begin{aligned}\mathbf{W}^{[l]} &: (n^{[l]}, n^{[l-1]}) \\ b^{[l]} &: (n^{[l]}, 1) \\ z^{[l]}, a^{[l]} &: (n^{[l]}, 1) \\ \mathbf{Z}^{[l]}, \mathbf{A}^{[l]} &: (n^{[l]}, m)\end{aligned}$$

The **non-vectorized** equations describing forward propagation through layer l of a deep L -layer network are:

$$\begin{aligned}
dz^{[l]} &= da^{[l]} * g'^{[l]}(z^{[l]}) \\
d\mathbf{W}^{[l]} &= \frac{1}{m} dz^{[l]} a^{[l-1]T} \\
db^{[l]} &= dz^{[l]} \\
da^{[l-1]} &= dz^{[l]} \mathbf{W}^{[l]T}
\end{aligned}$$

The **vectorized** equations describing forward propogation through layer l of a deep L -layer network are:

$$\begin{aligned}
d\mathbf{Z}^{[l]} &= d\mathbf{A}^{[l]} * g'^{[l]}(\mathbf{Z}^{[l]}) \\
d\mathbf{W}^{[l]} &= \frac{1}{m} d\mathbf{Z}^{[l]} \mathbf{A}^{[l-1]T} \\
db^{[l]} &= \frac{1}{m} \sum_{j=1}^m dz_{ij}^{[l]} \\
d\mathbf{A}^{[l-1]} &= d\mathbf{Z}^{[l]} \mathbf{W}^{[l]T}
\end{aligned}$$

The dimensions of all parameters involved in forward propogation are:

$$\begin{aligned}
d\mathbf{W}^{[l]} &: (n^{[l]}, n^{[l-1]}) \\
db^{[l]} &: (n^{[l]}, 1) \\
dz^{[l]}, da^{[l]} &: (n^{[l]}, 1) \\
d\mathbf{Z}^{[l]}, d\mathbf{A}^{[l]} &: (n^{[l]}, m)
\end{aligned}$$