1 General notation

1.1 Sizes and common indexes:

\( (i) \) Superscript refers to the \( i \)th training example.

\( [l] \) Superscript refers to the \( l \)th layer in the neural network.

\( \{i\} \) Superscript refers to the \( i \)th minibatch.

\( m \) Number of examples in the dataset, or minibatch, depending on context.

\( n_x \) Input size.

\( n_y \) Output size.

\( n_h[l] \) Number of hidden units of the \( l \)th layer.

\( L \) Total number of layers in the network.

1.2 Objects:

\( X \in \mathbb{R}^{n_x \times m} \) Input matrix.

\( Y \in \mathbb{R}^{n_y \times m} \) Label matrix.

\( x^{(i)} \) \( i \)th training example.

\( y^{(i)} \) \( i \)th label.
$W^{[l]}$ Weight matrix of layer $l$.

$b^{[l]}$ Bias vector of the $l$th layer.

$\hat{y} \in \mathbb{R}^{n_y}$ Predicted output vector.

### 1.3 Common forward propagation equations:

- $z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]}$ Linear combination at layer $l$.
- $a^{[l]} = g(z^{[l]})$ Output after activation function.
- $g(\cdot)$ Activation function: ReLu, tanh, sigmoid, etc.
- $\hat{y} = a^{[L]}$ Predicted output vector.
- $x = z^{[0]} = a^{[0]}$ Input vector.

### 1.4 Example of neural network

![Diagram of a neural network with 3 hidden layers and 4 input nodes.](image)

Figure 1: Example of neural network of depth $L = 3$. Weight and offset parameters have been omitted for clarity.
1.5 Convolutional networks

\[ n_W^{[l]} \] Width size of the output at the \( l \)th layer.

\[ n_H^{[l]} \] Height size of the output at the \( l \)th layer.

\[ n_C^{[l]} \] Number of filters (channels) at the \( l \)th layer.

\( f \) Filter size of the convolutional layer.

\( p \) Padding value of the convolutional layer.

\( s \) Stride size of the convolutional layer.

**Acknowledgements**

The notation used has been heavily inspired from the deeplearning.ai course.