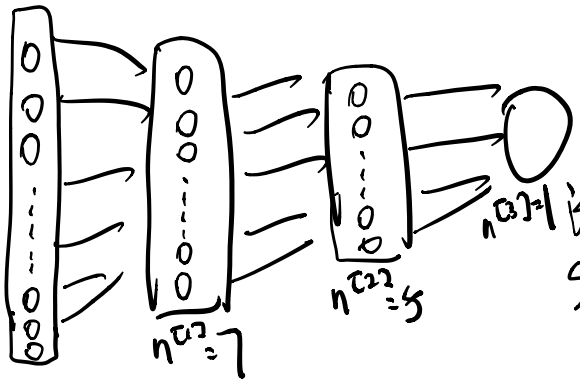
给定数据集(如大小为64x64x3的图片，m张)
写为向量形式：

$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots\cdots & x^{(m)} \\ | & | & & | \end{bmatrix}$$

X的shape：(12288, m)，其中12288=64x64x3



一般最后一层输出激活函数用sigmoid(多分类用Softmax)，其它层用RELU.

正向传播流程如下：

初始时，$A^{[0]}=X$，$W^{[l]}$.shape: $(n^{[l]}, n^{[l-1]})$

$b^{[l]}$.shape: $(n^{[l]}, 1)$

$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$ $(7\times 12288)\times(12288, m)+(7\times 1)=(7\times m)$

$A^{[1]} = g(Z^{[1]})$ $(7\times m)$

$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$ $(5\times 7)\times(7\times m)+(5\times 1)=(5\times m)$

$A^{[2]} = g(Z^{[2]})$ $(5\times m)$

$Z^{[3]} = W^{[3]} A^{[2]} + b^{[3]}$ $(1\times 5)\times(5\times m)+(1\times 1)=(1\times m)$

$A^{[3]} = g(Z^{[3]})$ $(1\times m)$

参数初始化：

$W^{[1]} = np.random.rand(n^{[1]}, n^{[l-1]}) * np.sqrt\left(\frac{2}{n^{[l-1]}}\right)$

$b^{[1]} = np.zeros((n^{[1]}, 1))$

np.random.rand 产生一个或一组服从正态分布的随机取样值.

正向模板：

$Z^{[l]} = W^{[l]T} A^{[l-1]} + b^{[l]}$ $(A^{[0]} = X)$

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

反向传播流程如下：

$$L=\frac{1}{m}\sum_{i=1}^{m}L(A^{[3](i)},y^{(i)})$$

$$=-\frac{1}{m}\left(\sum_{i=1}^{m}y^{(i)}\ln A^{[3](i)}+(1-y^{(i)})\ln(1-A^{[3](i)})\right)$$

$$dA^{[3]}=-\frac{1}{m}\left(\frac{Y}{A^{[3]}}-\frac{1-Y}{1-A^{[3]}}\right) \quad (1\times m)$$

$$dz^{[3]}=dA^{[3]}\cdot g'(z^{[3]}) \quad np.multiply(1\times m,1\times m)=(1\times m)$$

$$dw^{[3]}=dz^{[3]}\cdot A^{[2]} \quad (1\times m)\times(5\times m)^T=(1\times 5)$$

$$db^{[3]}=dz^{[3]} \quad np.sum(1\times m,axis=1)=(1\times 1)$$

$$dA^{[2]}=dz^{[3]}\cdot w^{[3]} \quad (1\times 5)^T\times(1\times m)=(5\times m)$$

$$dz^{[2]}=dA^{[2]}\cdot g'(z^{[2]}) \quad np.multiply(5\times m,5\times m)=(5\times m)$$

$$dw^{[2]}=dz^{[2]}\cdot A^{[1]} \quad (5\times m)\times(7\times m)^T=(5\times 7)$$

$$db^{[2]}=dz^{[2]} \quad np.sum(5\times m,axis=1)=(5\times 1)$$

$$dA^{[1]}=dz^{[2]}\cdot w^{[2]} \quad (5\times 7)^T\times(5\times m)=(7\times m)$$

$$dz^{[1]}=dA^{[1]}\cdot g'(z^{[1]}) \quad np.multiply(7\times m,7\times m)=(7\times m)$$

$$dw^{[1]}=dz^{[1]}\cdot A^{[0]} \quad (7\times m)\times(12288,m)^T=(7\times 12288)$$

$$db^{[1]}=dz^{[1]} \quad np.sum(7\times m,axis=1)=(7\times 1)$$

因为反向传播时会用到$A^{[1]}$，$z^{[1]}$，所以应在正向传播中将它们存起来。

反向模板：

$$dZ^{[l]} = np.multiply(dA^{[l]}, g'Z^{[l]})$$

$$dw^{[l]} = dZ^{[l]} \cdot A^{[l-1]T}$$

$$db^{[l]} = np.sum(dZ^{[l]}, axis=1)$$

$$dA^{[l]} = w^{[l+1]T} \cdot dZ^{[l+1]} \text{（第一层除外）}$$