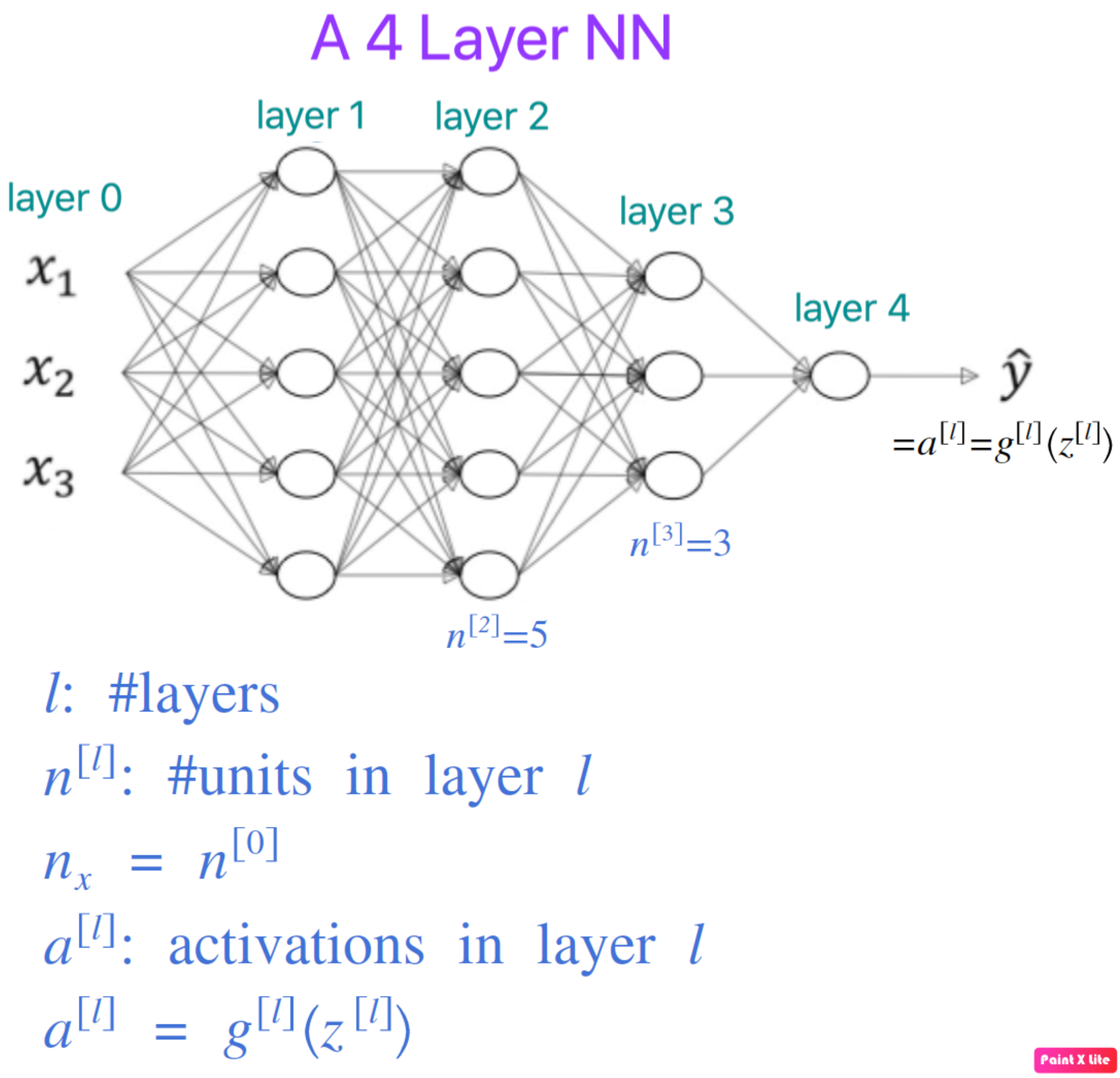
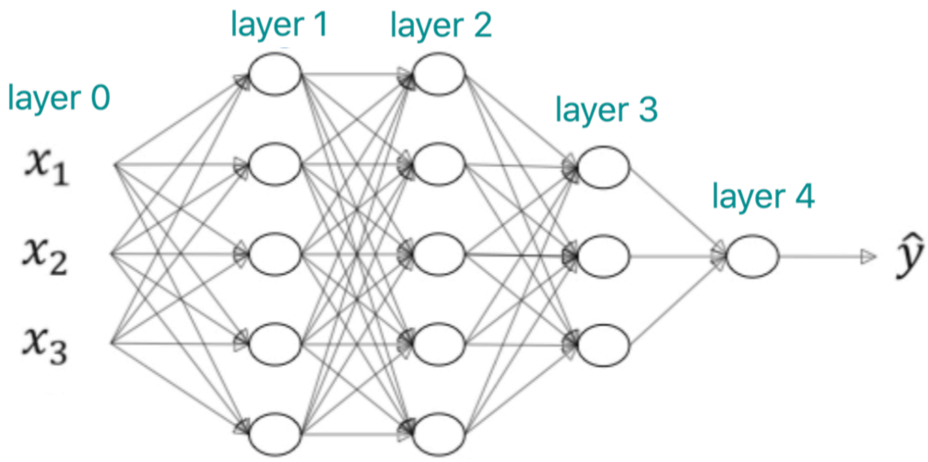
介绍了深度神经网络，DNN的Forward和Backward Propagation以及超参数。

1. DNN的表示



上图是一个4层神经网络的示意图，在计算层数是，输入层不计入。

2. Forward Propagation



Vectorization：

$$
\begin{aligned}
z^{[1]} &= w^{[1]}a^{[0]} + b^{[1]} \\
a^{[1]} &= g^{[1]}(z^{[1]}) \\
&\dots \\
z^{[4]} &= w^{[4]}a^{[3]} + b^{[4]} \\
a^{[4]} &= g^{[4]}(z^{[4]}) = \hat{y}
\end{aligned}
$$

$$
\begin{aligned}
Z^{[1]} &= W^{[1]}A^{[0]} + b^{[1]} \\
A^{[1]} &= g^{[1]}(Z^{[1]}) \\
&\dots \\
Z^{[4]} &= W^{[4]}A^{[3]} + b^{[4]} \\
A^{[4]} &= g^{[4]}(Z^{[4]}) = \hat{Y}
\end{aligned}
$$

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

通项

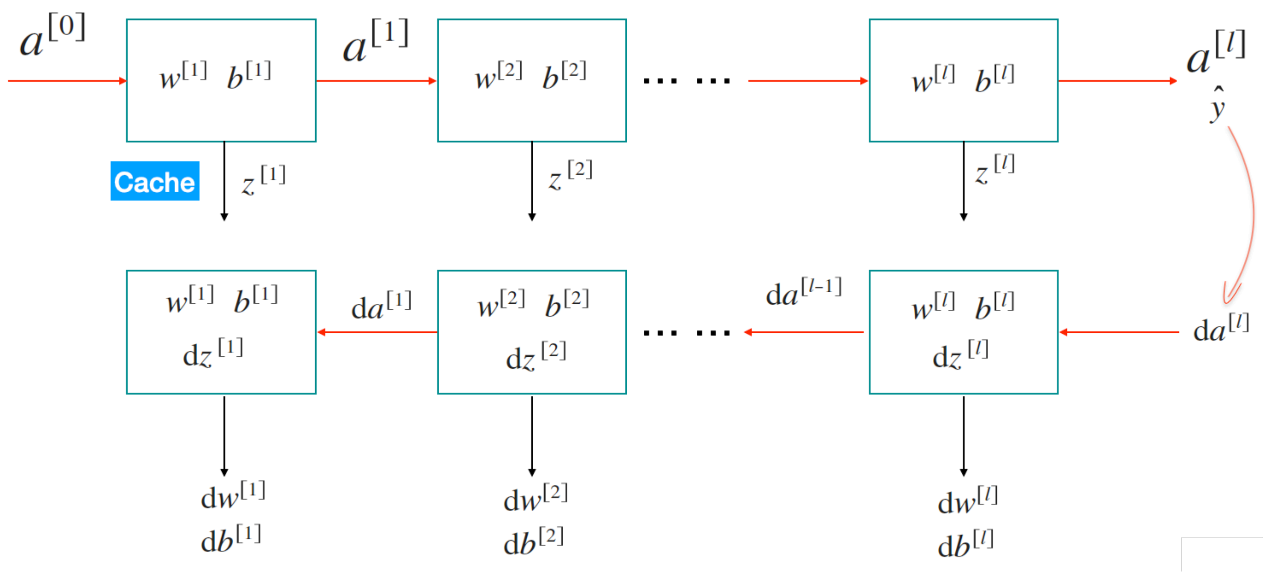
$$
\begin{aligned}
A^{[0]} &= X \\
Z^{[l]} &= W^{[l]}A^{[l-1]} + b^{[1]} \\
A^{[l]} &= g^{[l]}(Z^{[l]})
\end{aligned}
$$

3. DNN中矩阵的维度

$$
\begin{aligned}
\underset{(n^{[l]},1)}{z^{[l]}} &= \underset{(n^{[l]},n^{[l-1]})}{w^{[l]}}\underset{(n^{[l-1]},1)}{a^{[l-1]}} + b^{[1]} \\
\underset{(n^{[l]},1)}{a^{[l]}} &= g^{[l]}(z^{[l]})
\end{aligned}
$$

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

4. Building Blocks of DNN



5. Forward & Backward Propagation

Forward

$$Z^{[1]} = W^{[1]}A^{[1]} + b^{[1]}$$
$$A^{[1]} = g^{[1]}(Z^{[1]})$$
$$\cdots$$
$$Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$$
$$A^{[l]} = g^{[l]}(Z^{[l]})$$
$$J = -\frac{1}{m} np.sum(Y * \log(A^{[l]}) + (1-Y) * \log(1-A^{[l]}))$$

Backward

$$dZ^{[l]} = dA^{[l]} * g^{[l]\prime}(Z^{[l]})$$
$$= A^{[l]} - Y \quad (\text{if } g^{[l]} \text{ is sigmoid})$$
$$dW^{[l]} = \frac{1}{m} dZ^{[l]} A^{[l-1]T}$$
$$db^{[l]} = \frac{1}{m} np.sum(dZ^{[l]}, axis=1, keepdims=True)$$
$$dA^{[l-1]} = W^{[l]T} dZ^{[l]}$$
$$\cdots$$

Q1: $\mathcal{J}$用矩阵怎么表示?

二分类问题：

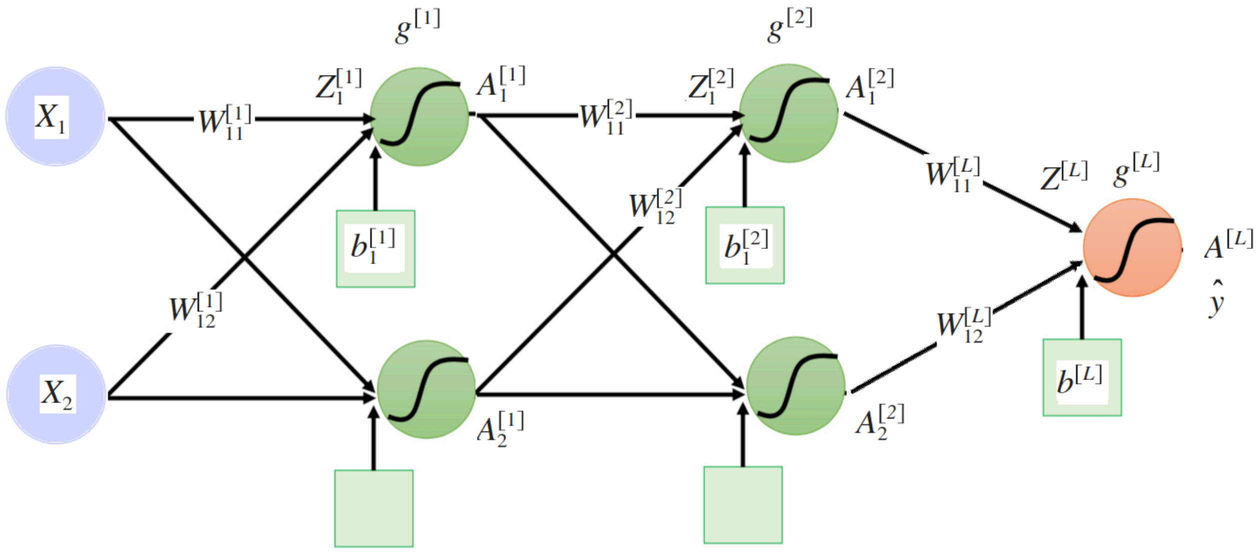
$$\mathcal{J} = -\frac{1}{m}\left(Y^T \log A^{[L]} + (1 - Y^T)\log(1 - A^{[L]})\right)$$

多分类问题：

$$\mathcal{J} = -\frac{1}{m}\left(\mathbf{1}_{n^{[L]}}^T \left(Y^T \odot \log A^{[L]}\right)\right)$$

Q2: 用矩阵求导的方式推导Backward

Forward Propagation：



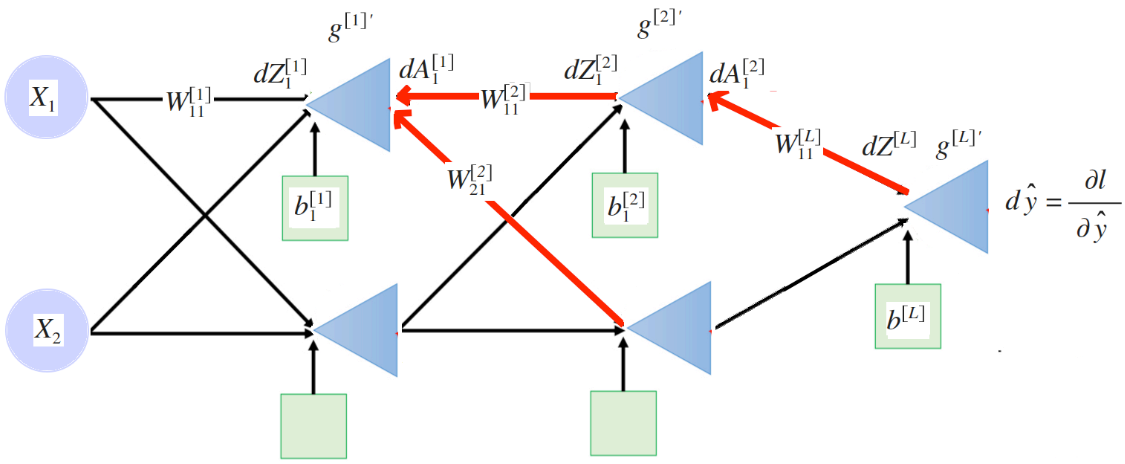
标量形式（示例）

$$
\begin{aligned}
Z_1^{[1]} &= W_{11}^{[1]} X_1 + W_{12}^{[1]} X_2 + b_1^{[1]} \\
A_1^{[1]} &= g^{[1]}\left(Z_1^{[1]}\right) \\
Z_1^{[2]} &= W_{11}^{[2]} A_1^{[1]} + W_{12}^{[2]} A_2^{[1]} + b_1^{[2]} \\
A_1^{[2]} &= g^{[2]}\left(Z_1^{[2]}\right) \\
Z^{[L]} &= W_{11}^{[L]} A_1^{[2]} + W_{12}^{[L]} A_2^{[2]} + b^{[L]} \\
\hat{y} &= g^{[L]}\left(Z^{[L]}\right)
\end{aligned}
$$

矩阵形式

$$
\begin{aligned}
Z^{[1]} &= W^{[1]} X + b^{[1]} \\
A^{[1]} &= g^{[1]}\left(Z^{[1]}\right) \\
Z^{[2]} &= W^{[2]} A^{[1]} + b^{[2]} \\
A^{[2]} &= g^{[2]}\left(Z^{[2]}\right) \\
Z^{[L]} &= W^{[L]} A^{[2]} + b^{[L]} \\
\hat{y} &= g^{[L]}\left(Z^{[L]}\right)
\end{aligned}
$$

Backward Propagation：



标量形式（示例）

$$dZ^{[L]} = g^{[L]'}\left(Z^{[L]}\right) d\hat{y}$$
$$dW_{11}^{[L]} = A_1^{[2]} dZ^{[L]}$$
$$db^{[L]} = dZ^{[L]}$$
$$dZ_1^{[2]} = g^{[2]'}\left(Z_1^{[2]}\right)\left(W_{11}^{[L]} dZ^{[L]}\right)$$
$$dW_{11}^{[2]} = A_1^{[1]} dZ_1^{[2]}$$
$$db_1^{[2]} = dZ_1^{[2]}$$
$$dZ_1^{[1]} = g^{[1]'}\left(X_1\right)\left(W_{11}^{[2]} dZ_1^{[2]} + W_{21}^{[2]} dZ_2^{[2]}\right)$$
$$dW_{11}^{[1]} = X_1 dZ_1^{[1]}$$
$$db_1^{[1]} = dZ_1^{[1]}$$

矩阵形式

$$\mathrm{d}Z^{[L]} = g^{[L]\prime}\left(Z^{[L]}\right) \odot \mathrm{d}A^{[L]}$$
$$\mathrm{d}W^{[L]} = \frac{1}{m}\mathrm{d}Z^{[L]} A^{[L-1]T}$$
$$\mathrm{d}b^{[L]} = \frac{1}{m} np.sum\left(\mathrm{d}Z^{[L]}, axis = 1, keepdim = True\right)$$
$$\mathrm{d}Z^{[L-1]} = \boxed{g^{[L-1]\prime}\left(Z^{[L-1]}\right)} \odot \boxed{\left(W^{[L]T}\mathrm{d}Z^{[L]}\right)}$$
$$\cdots$$

Forward Pass   Backward Pass

6. Hyper Parameters

1. Parameter

   $W^{[1]}, b^{[1]}, \cdots, W^{[l]}, b^{[l]}$

2. Hyperparameters

   - Learning rate: α
   - #iterations
   - #Hidden layer L
   - #Hidden units
   - activation function
   - … …