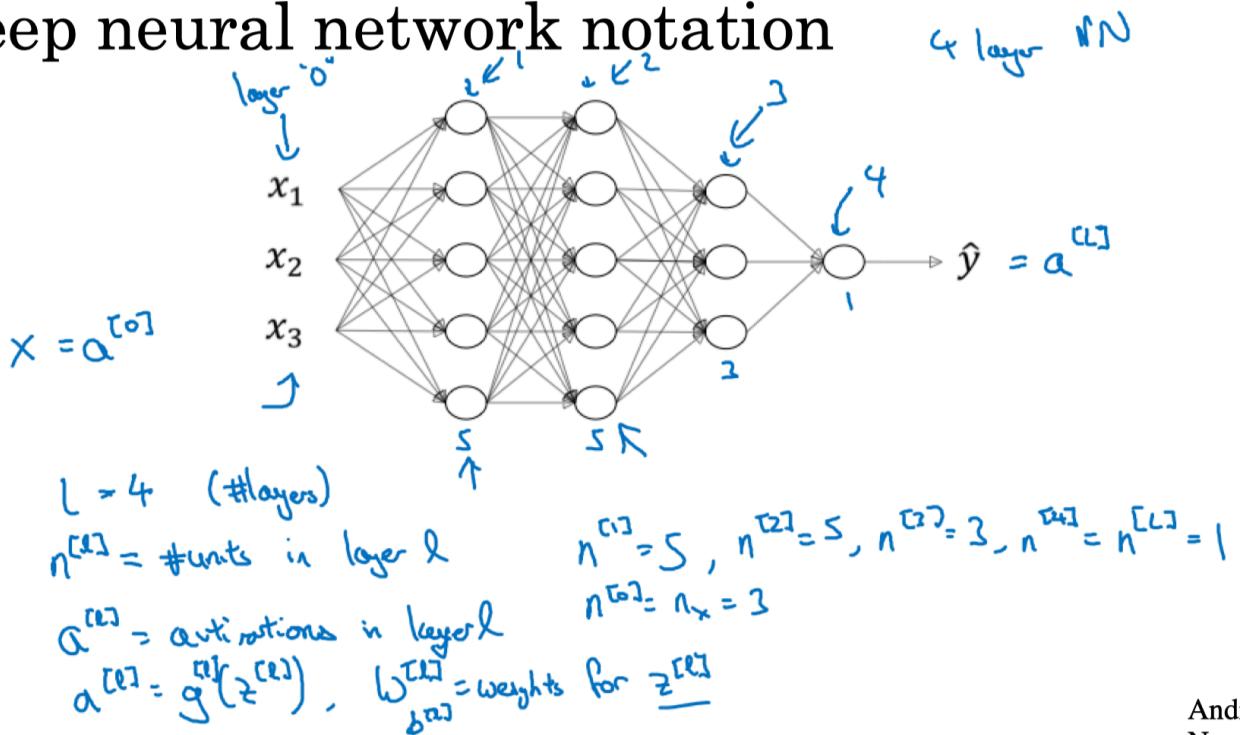


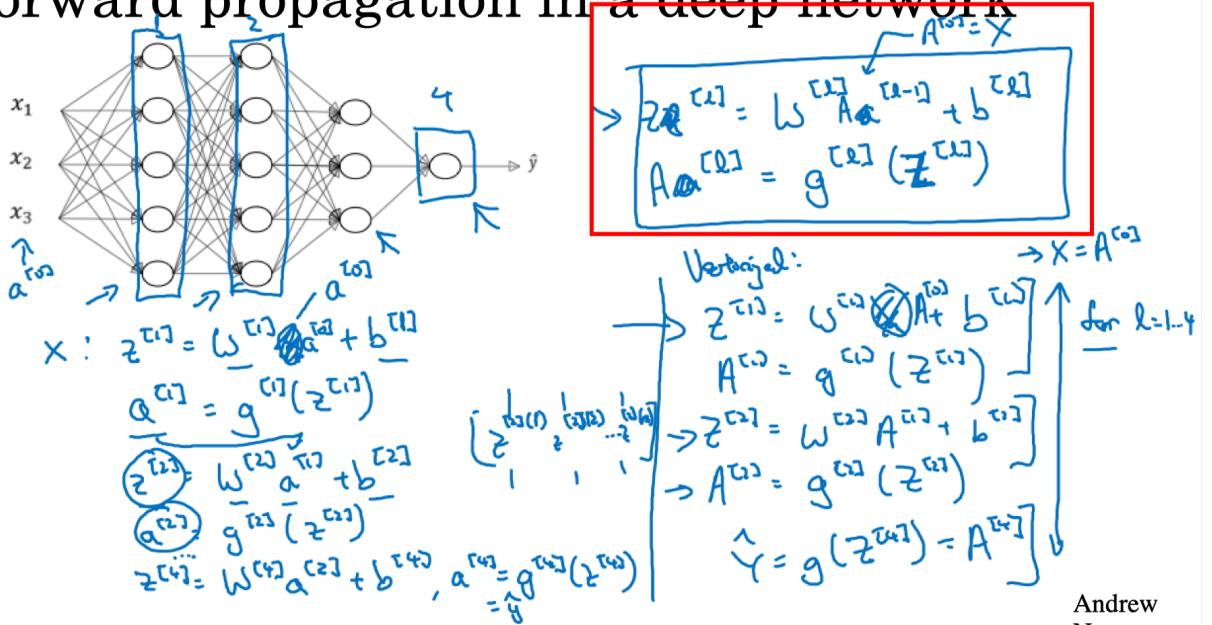
1. Deep neural network notation:

Deep neural network notation



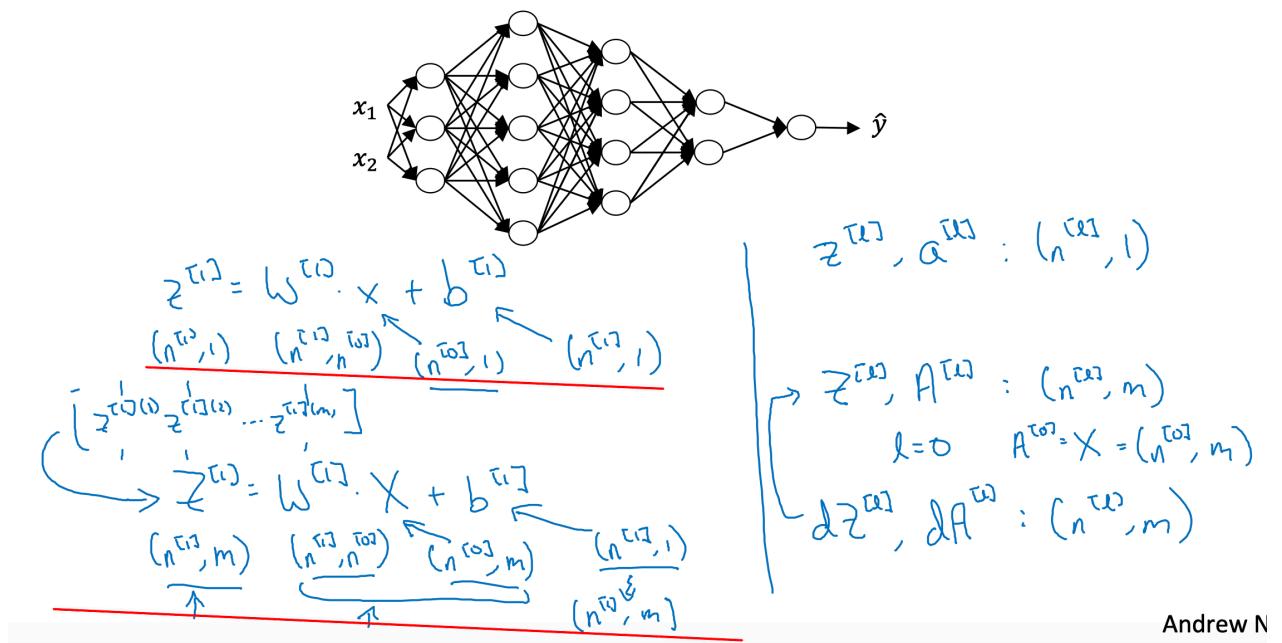
2. Forward propagation in a deep network:

Forward propagation in a deep network



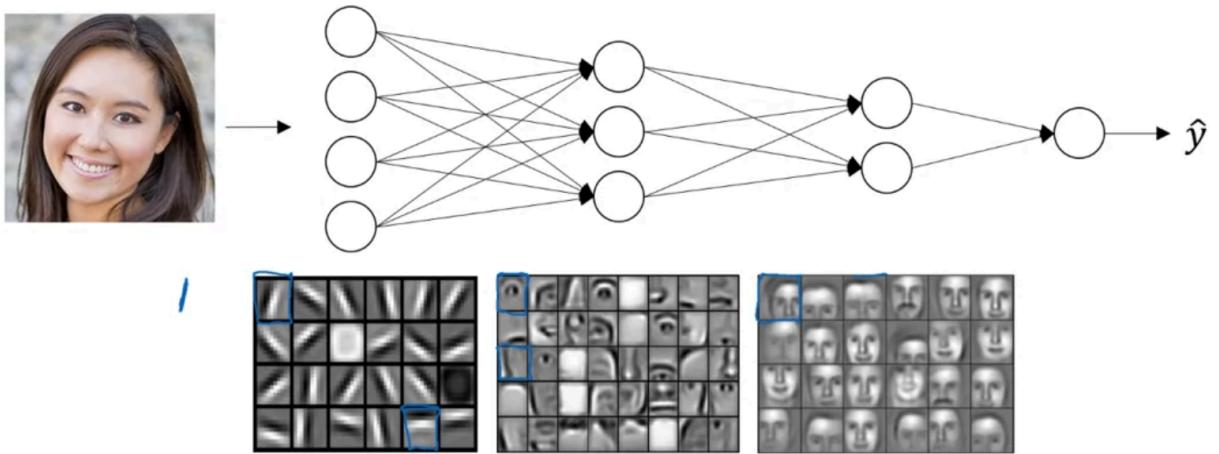
3. Vectorized implementation: dimension

Vectorized implementation



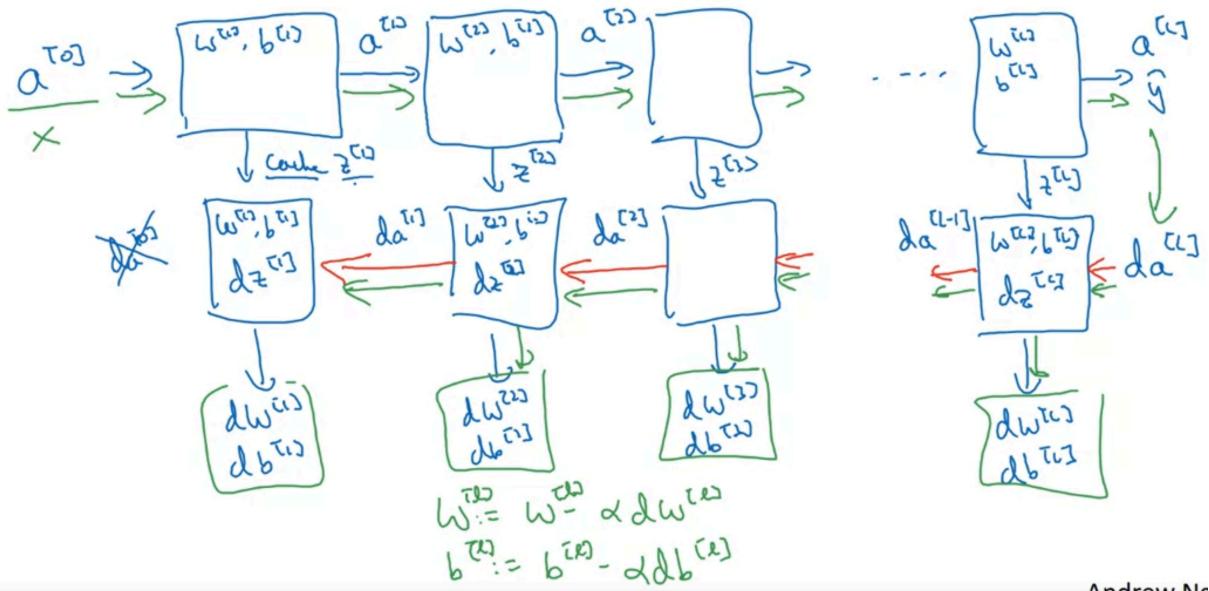
4. Why Deep Representations?

Intuition about deep representation



Better Feature Abstraction: 第一层提取出检测边的方向，第二层拼接这些边来检测五官，第三层拼接五官来检测整张脸

5. Forward and backward functions and block:



6. Forward:

Forward propagation for layer l

→ Input $a^{[l-1]}$

→ Output $a^{[l]}$, cache ($\underline{z}^{[l]}$)

$$z^{[l]} = w^{[l]} \cdot a^{[l-1]} + b^{[l]}$$

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

Vertwigd:

$$\underline{z}^{[l]} = w^{[l]} \cdot A^{[l-1]} + b^{[l]}$$

$$A^{[l]} = g^{[l]}(\underline{z}^{[l]})$$

7. Backward:

Backward propagation for layer l

→ Input $\underline{da}^{[l]}$

→ Output $\underline{da}^{[l-1]}, dW^{[l]}, db^{[l]}$

$$d\underline{z}^{[l]} = \underline{da}^{[l]} * g'^{[l]}(z^{[l]})$$

$$dW^{[l]} = d\underline{z}^{[l]} \cdot a^{[l-1]T}$$

$$db^{[l]} = d\underline{z}^{[l]}$$

$$\underline{da}^{[l-1]} = W^{[l]T} \cdot d\underline{z}^{[l]}$$

$$d\underline{z}^{[l]} = W^{[l+1]T} \cdot d\underline{z}^{[l+1]} * g'^{[l+1]}(z^{[l]})$$

$$d\underline{z}^{[l]} = \underline{da}^{[l]} * g'^{[l]}(z^{[l]})$$

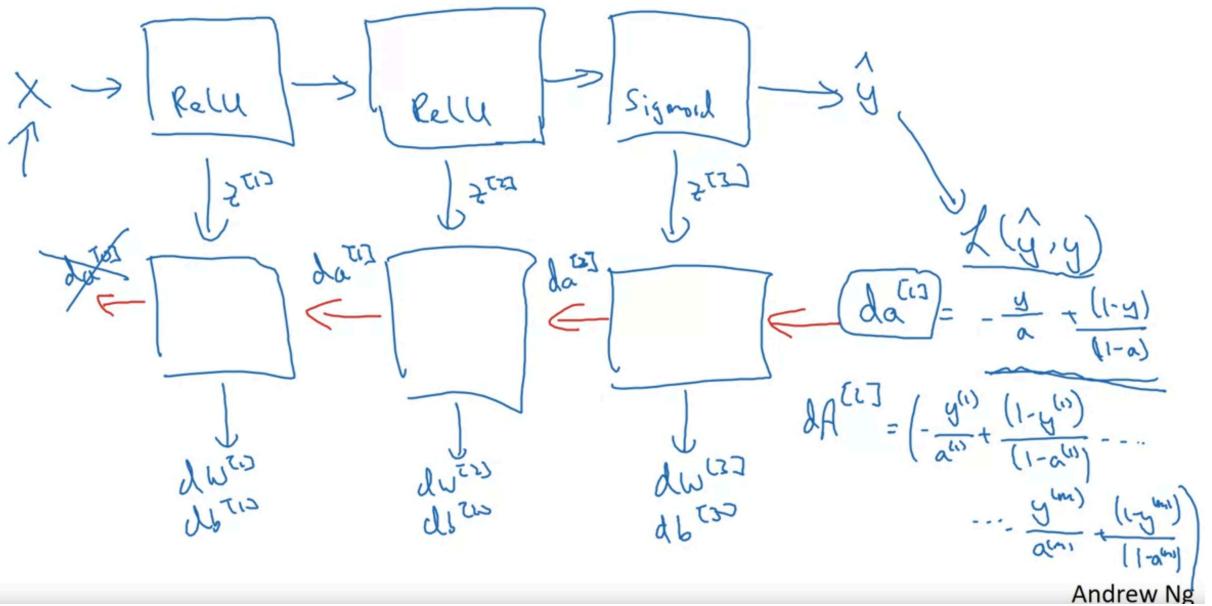
$$dW^{[l]} = \frac{1}{m} d\underline{z}^{[l]} \cdot A^{[l-1]T}$$

$$db^{[l]} = \frac{1}{m} \text{np.sum}(d\underline{z}^{[l]}, \text{axis}=1, \text{keepdims=True})$$

$$\underline{da}^{[l-1]} = W^{[l]T} \cdot d\underline{z}^{[l]}$$

Andrew Ng

8. Summary:



9. Hyperparameters: parameters that control ultimate parameters of model W and B

What are hyperparameters?

Parameters: $W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]} \dots$

Hyperparameters: learning rate α

#iterations

#hidden layers L

#hidden units $n^{[1]}, n^{[2]}, \dots$

choice of activation function

Later: Momentum, minibatch size, regularizations, ...