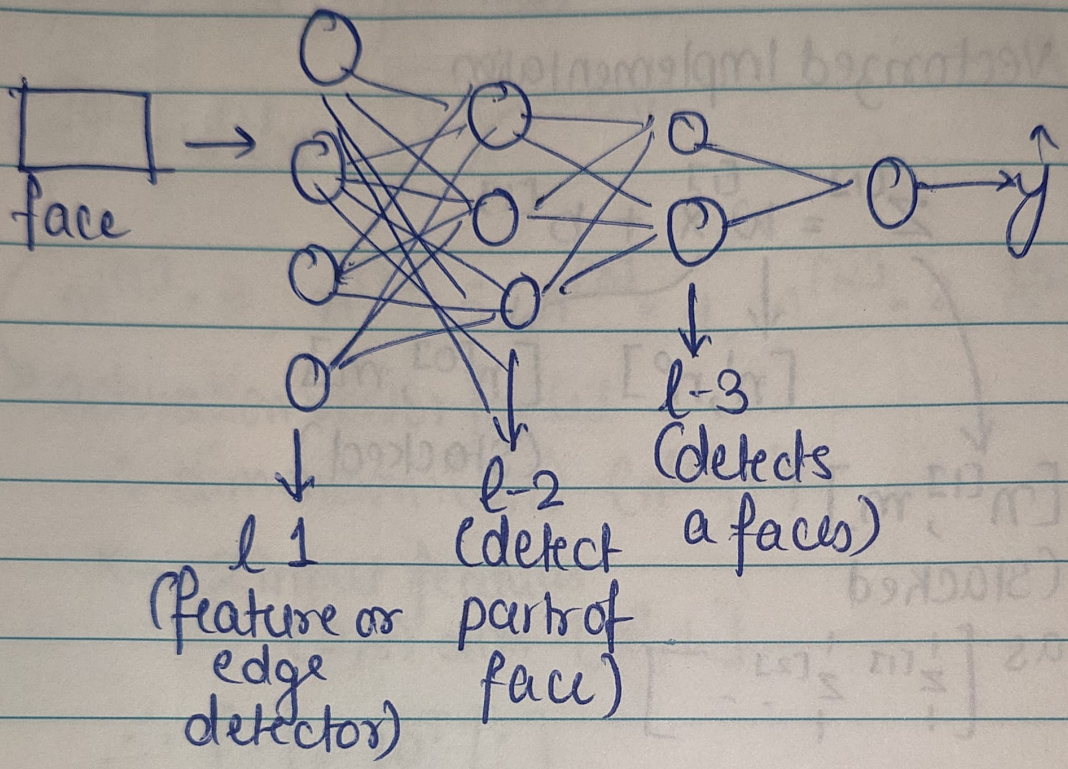
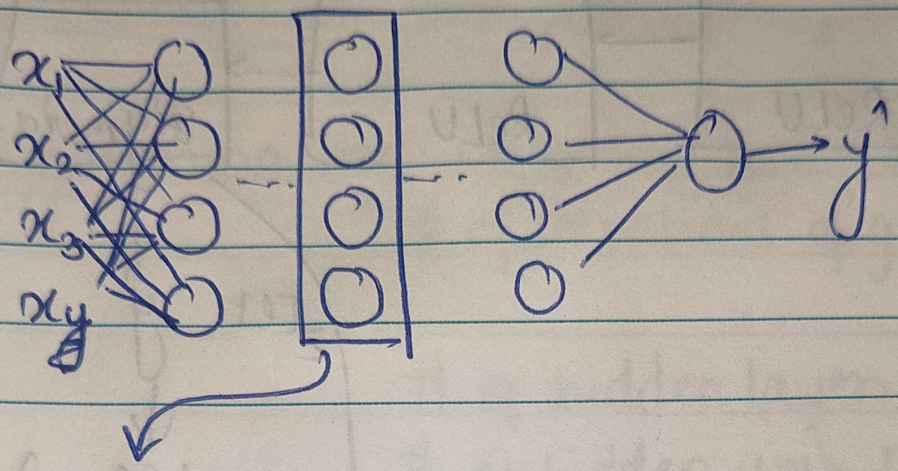
Why deep network?



Circuit theory : there r functions you can compute w/ a small L-layer deep neural network that shallower networks require exponentially more hidden units to compute.

Building Blocks of deep Net.



layer l: $w^{[l]}, b^{[l]}$

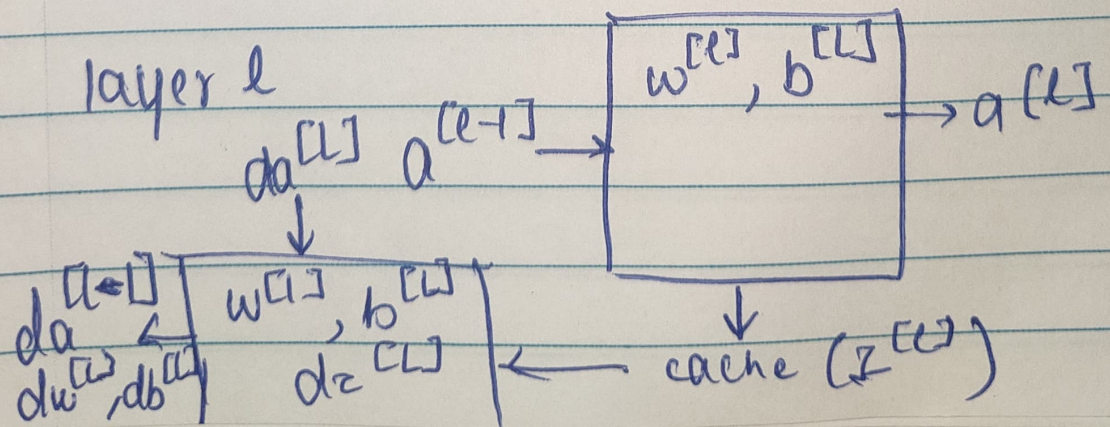
forward ⇒ I/P: $a^{[l-1]}$ O/P: $a^{[l]}$

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

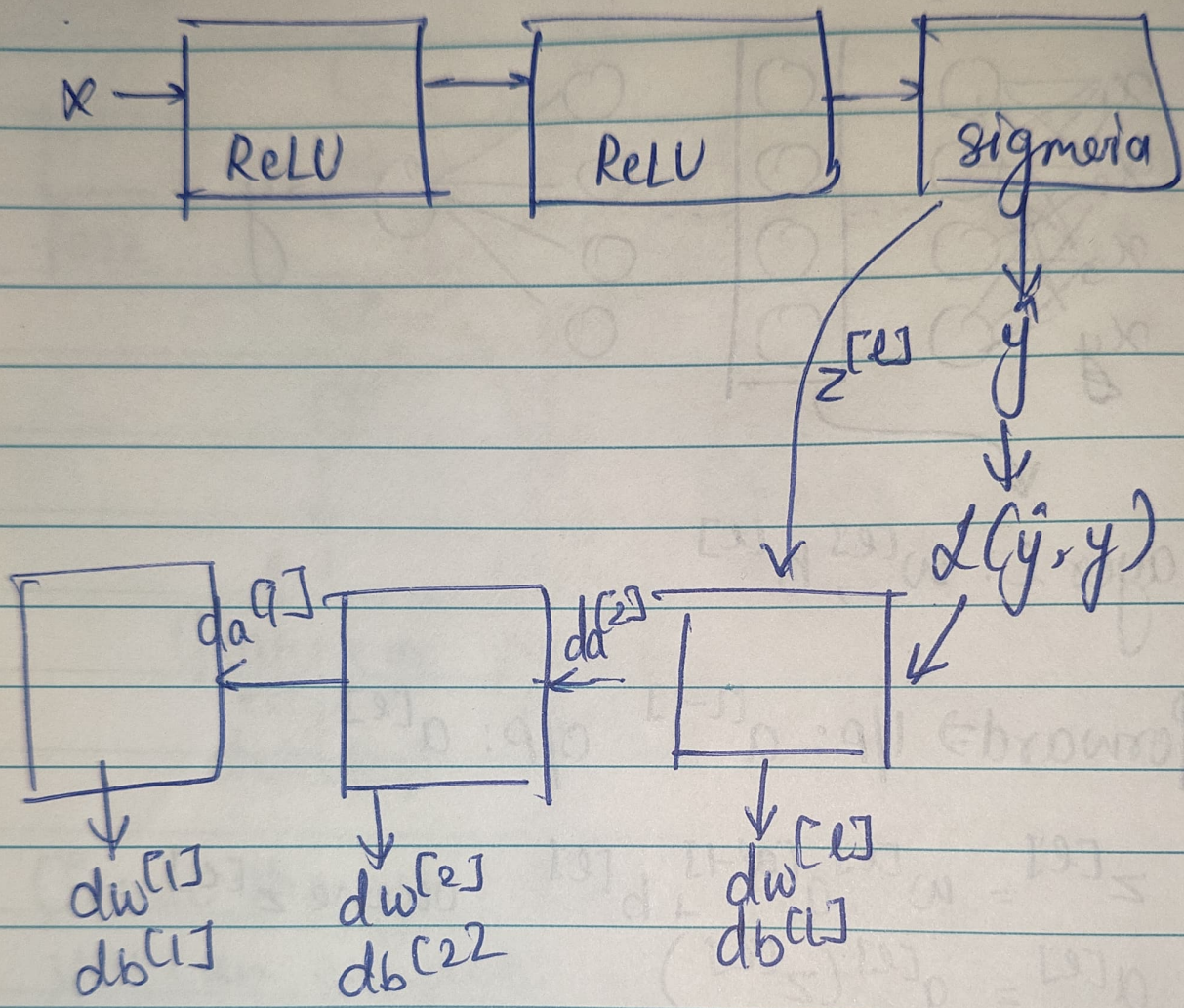
$a^{[l]} = g^{[l]}(z^{[l]})$

Backward ⇒ I/P: $da^{[l]}$ O/P: $da^{[l-1]}$

cache ($dz^{[l]}$) $dw^{[l]}$ $db^{[l]}$



Forward + Backward Propagation



Parameters & Hyperparameters

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

Hyperparameters: Learning rate (α)
# of iterations of gradient descent

# of hidden layer
# of hidden unit $n^{[1]} \ldots n^{[L]}$
choice of activation fn

trial & error