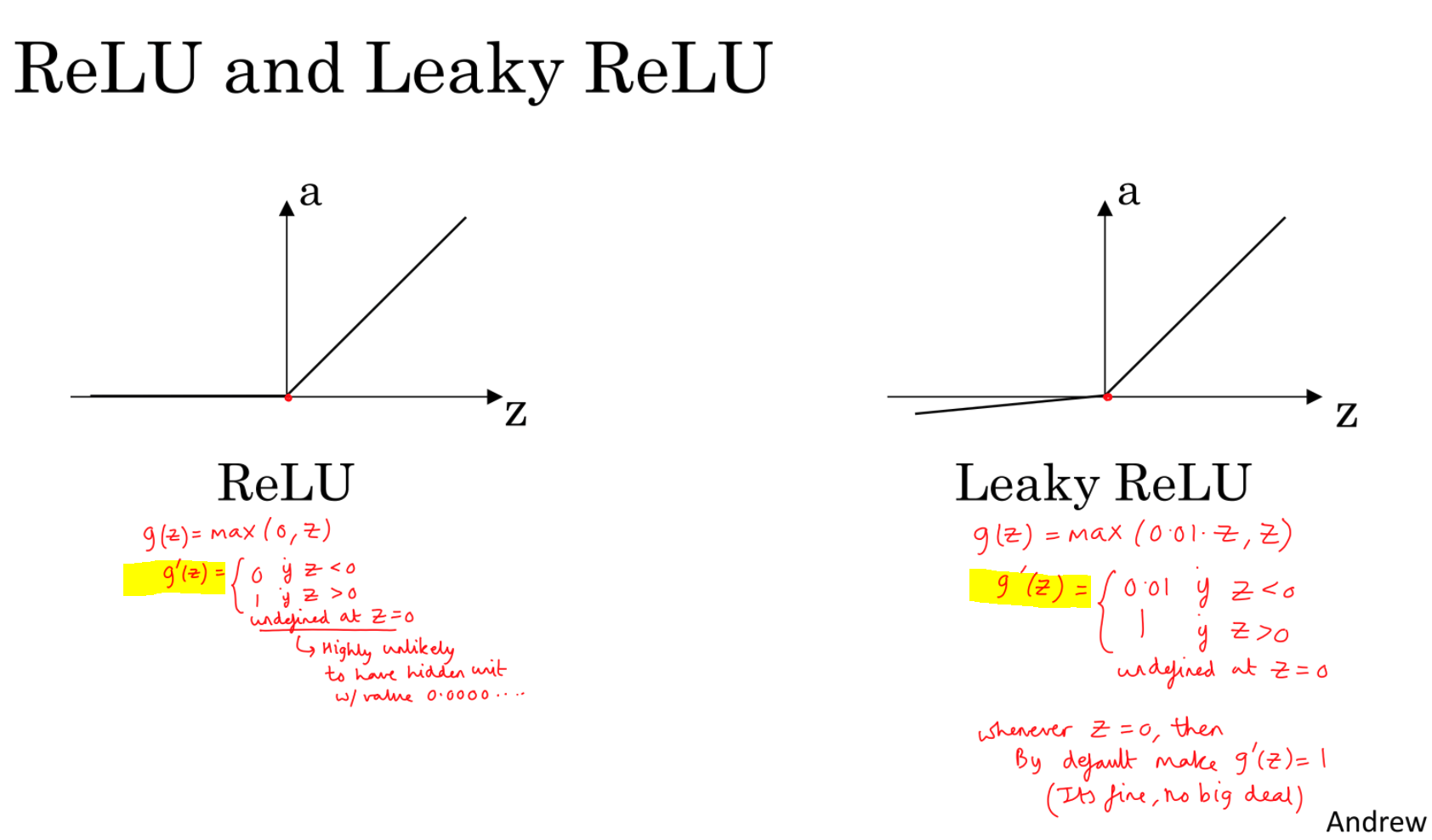
Notice the graphs in lec 6 - **last slide**

They are between ‘a’ and ‘z’, ie, y axis is ‘a’ and x axis is ‘z’

It is showing the activation values corresponding to z values



The important thing to note down is the stuff in yellow.

Consider ReLu graph, g(z) = max(0,z)

=> derivative of g(z) will be 0 when z is < 0

=> derivative of g(z) will be 1 when z is > 0

=> the derivative is a constant value 1 when z > 0, which means that dz (and consequently dw) will have some concrete direction during back prop as its derivatives are not 0.0001

Ie - when we do updates on weights, w = w - alpha \* dw, we will move in a tangible direction, compared to dw being .0001, in which case we would **move very slowly**!!

That **slow movement** is essentially what happens in the case of sigmoid/tanh!

Derivative of sigmoid = a \* (1 - a), where a = 1 / (1 + exp(-z))

When z > 0 (say z = 5), then dz = some small value (0.006), now if dz is small, dw is small => change in weights of w are negligible, which means it takes longer to converge!!