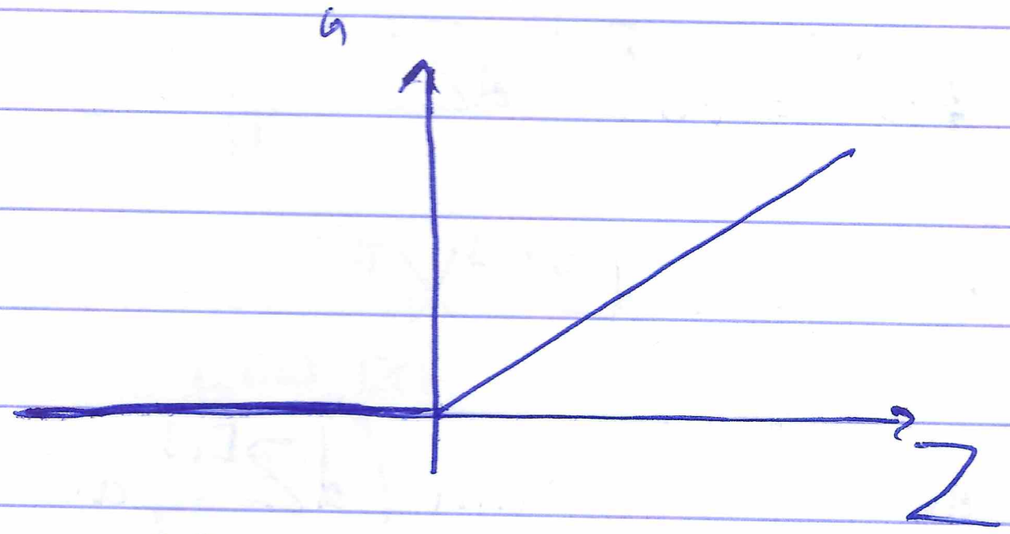
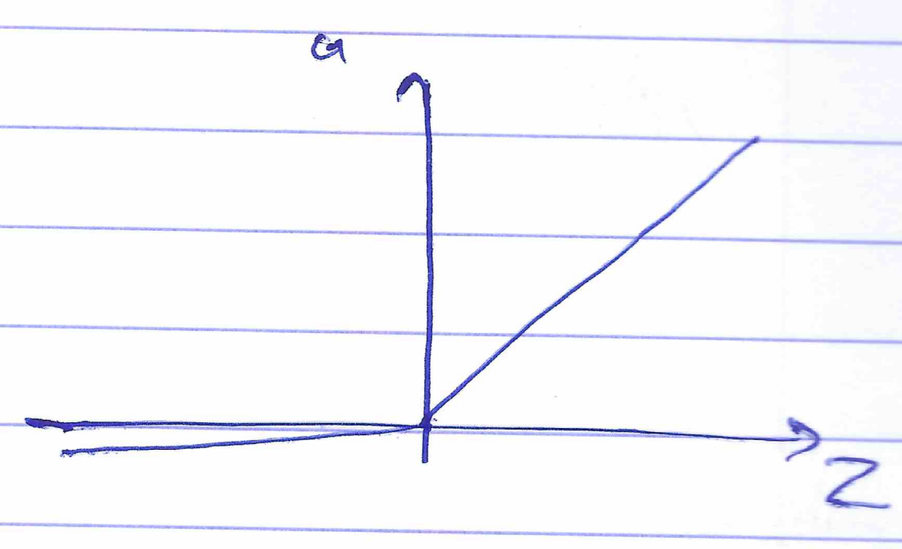
ReLU & Leaky ReLU



$g(z) = \max(0, z)$

$$g'(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \\ \text{undef. if } z = 0 \end{cases}$$

$z = 0.0000\ldots$



$g(z) = \max(0.01z, z)$

$$g'(z) = \begin{cases} 0.01 & \text{if } z < 0 \\ 1 & \text{if } z > 0 \end{cases}$$

$W^{[l]}$ is $(n^{[l]}, n^{[l-1]})$
$b^{[l]}$ is $(n^{[l]}, 1)$

GD for NNs

$n_x = n^{[0]}, n^{[1]}, n^{[2]} = 1$

Parameters: $W^{[1]}_{(n^{[1]}, n^{[0]})}, b^{[1]}_{(n^{[1]}, 1)}, W^{[2]}_{(n^{[2]}, n^{[1]})}, b^{[2]}_{(n^{[2]}, 1)}$

CF: $J(W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}, y)$ ($\hat{y} = a^{[2]}$)

GD: Repeat {

Compute predictions $(\hat{y}^{(i)}, i = 1:m)$

$dW^{[1]} = \frac{dJ}{dW^{[1]}}, db^{[1]} = \frac{dJ}{db^{[1]}}, \ldots$

$W^{[1]} := W^{[1]} - \alpha dW^{[1]}$

$b^{[1]} := b^{[1]} - \alpha db^{[1]}$

$b^{[2]} := \ldots$

Important to initialise parameters randomly

Key: know how to compute the derivatives

Formulas for computing derivatives:

- Forprop: $Z^{[1]} = W^{[1]}X + b^{[1]}$
  $A^{[1]} = g^{[1]}(Z^{[1]})$
  $Z^{[2]} = W^{[2]}A^{[1]} + b^{[2]}$
  $A^{[2]} = g^{[2]}(Z^{[2]}) = \sigma(Z^{[2]})$

- Backprop: $dZ^{[2]} = A^{[2]} - Y$ ($Y = [y^{(1)} y^{(2)} \ldots y^{(m)}]$)
  $dW^{[2]} = \frac{1}{m} dZ^{[2]} A^{[1]T}$
  $db^{[2]} = \frac{1}{m}$ np.sum($dZ^{[2]}$, axis=1, keepdims=True)
  (Ensures $(n, 1)$ Python output)