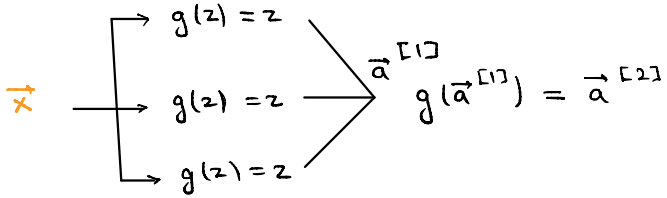
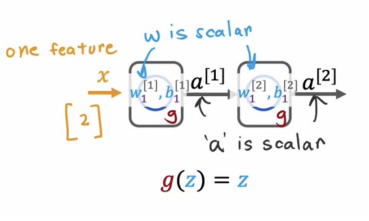
We need activation functions?

why not linear activation function in all layers?



(because everything is outputted to z)

⇒ if all $g(z)$ are linear, it will be no different than linear regression.



$$a^{[1]} = w_1^{[1]} x + b_1^{[1]}$$

$$a^{[2]} = w_1^{[2]} a^{[1]} + b_1^{[2]}$$

basically,

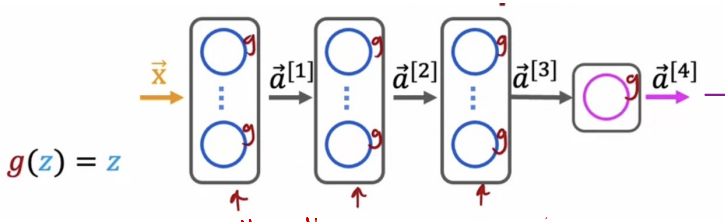
$$a^{[2]} = w_1^{[2]}(w_1^{[1]} x + b_1^{[1]}) + b_1^{[2]}$$

$$\Rightarrow a^{[2]} = \underbrace{w_1^{[2]} w_1^{[1]}}_{w} \; x \; + \underbrace{w_1^{[2]} b_1^{[1]} + b_1^{[2]}}_{b}$$

⇒ a linear regression model!

no change

Even if we change the output layer activation function and keep linear function for all layers, we'll get whatever function we have for the output layer.



⇒ hidden layers will have no influence and we basically get a logistics regression model.

In Short,

Don't use linear activation function in hidden layer.