

< Choosing Activation function >

- depending on what the target label or ground truth true variable Y is
- I can choose different activation functions for different neurons

* Natural choice of activation function at the output layer

- ① Binary classification : "Sigmoid" at the output layer
- ② Regression problem : "Linear activation function" at the output layer
(negative/positive)
- ③ Regression problem (target value never be negative) : "ReLU" at the output layer
ex) house price

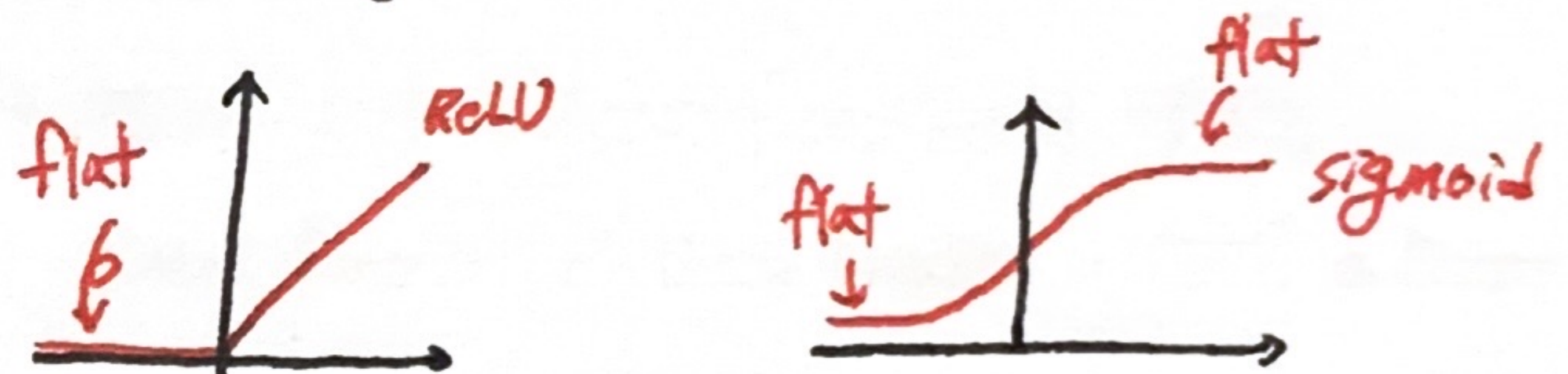
* Natural choice of activation function at the hidden layer

- ① Most common choice : 'ReLU'
- ② Binary classification : 'sigmoid'

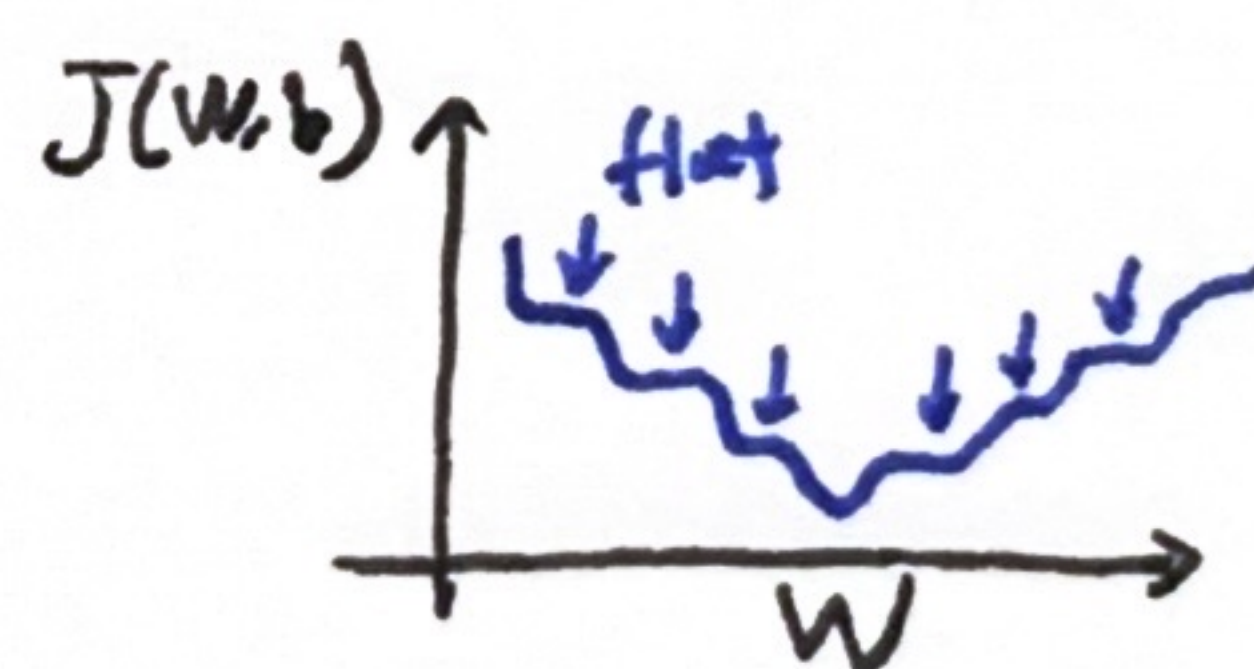
Why ReLU is the most common choice?

① ReLU is faster than sigmoid to compute the function
 $\max(0, z)$ vs $\frac{1}{1+e^{-z}}$

② ReLU function goes flat only in one part of graph whereas sigmoid goes flat on two places.



if there are lots of flat place on the activation function it results in more flat places in the cost function $J(w,b)$



↓

"This results gradient descents work slowly"

∴ Faster learning : $\text{ReLU} > \text{sigmoid}$