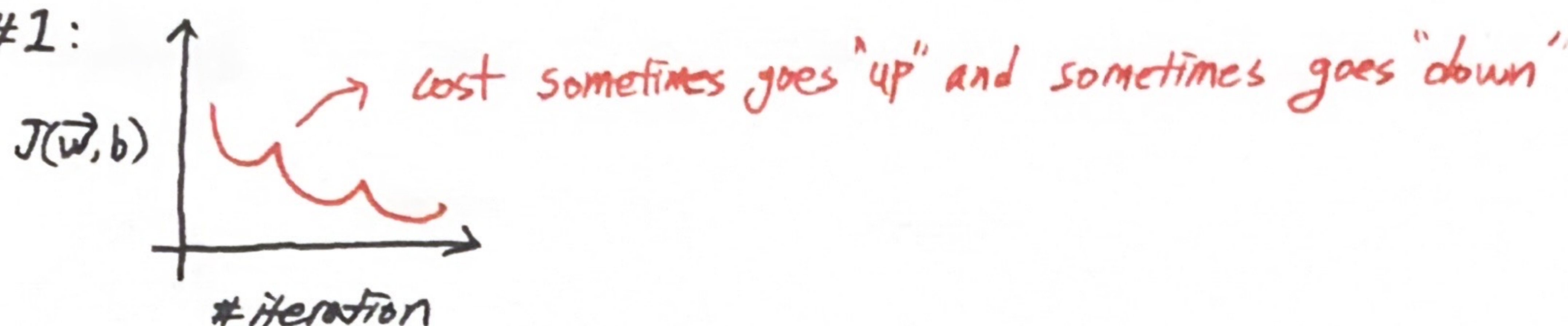


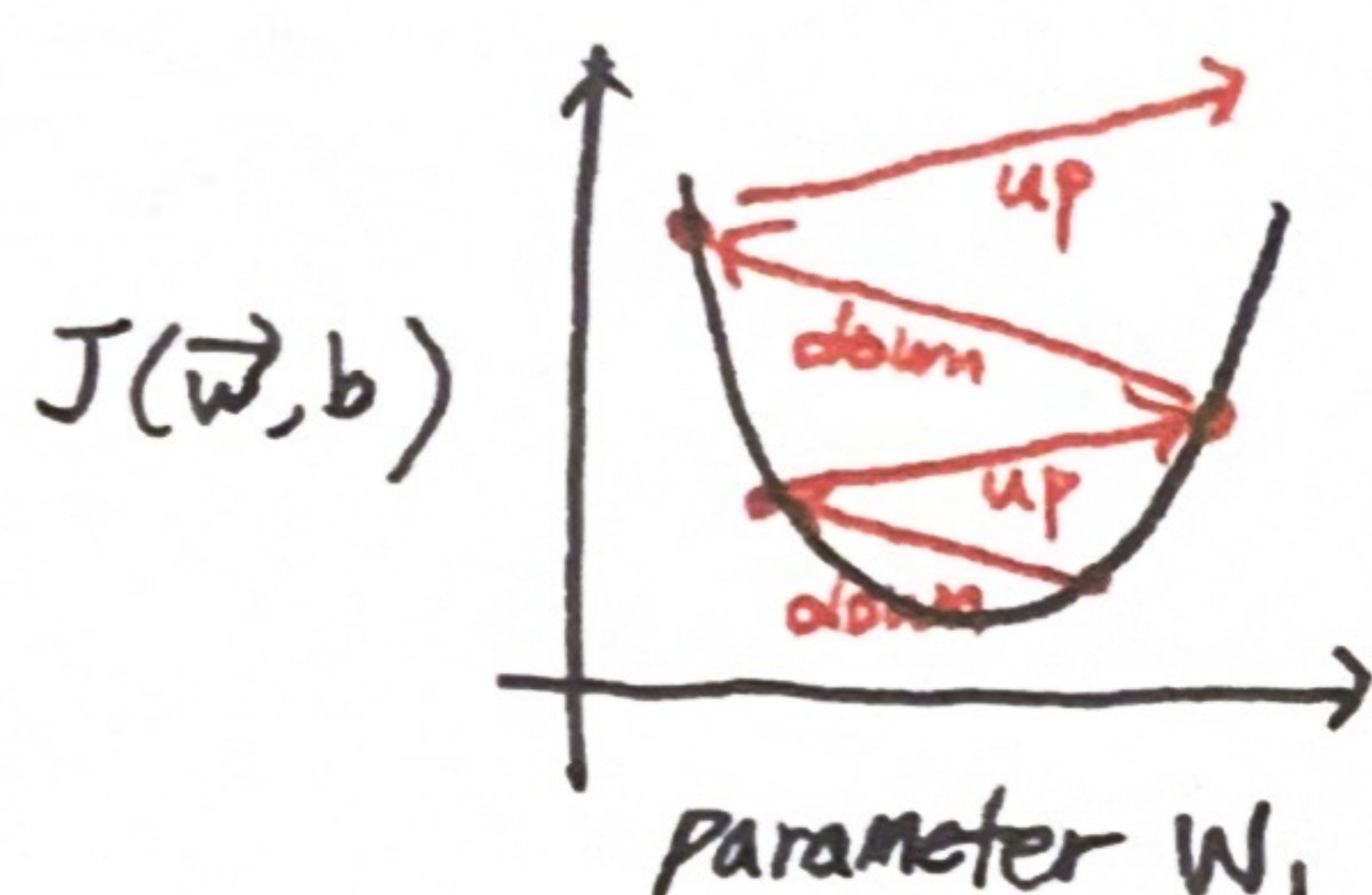
## < Choosing learning rate >

- Identify problem with gradient descent

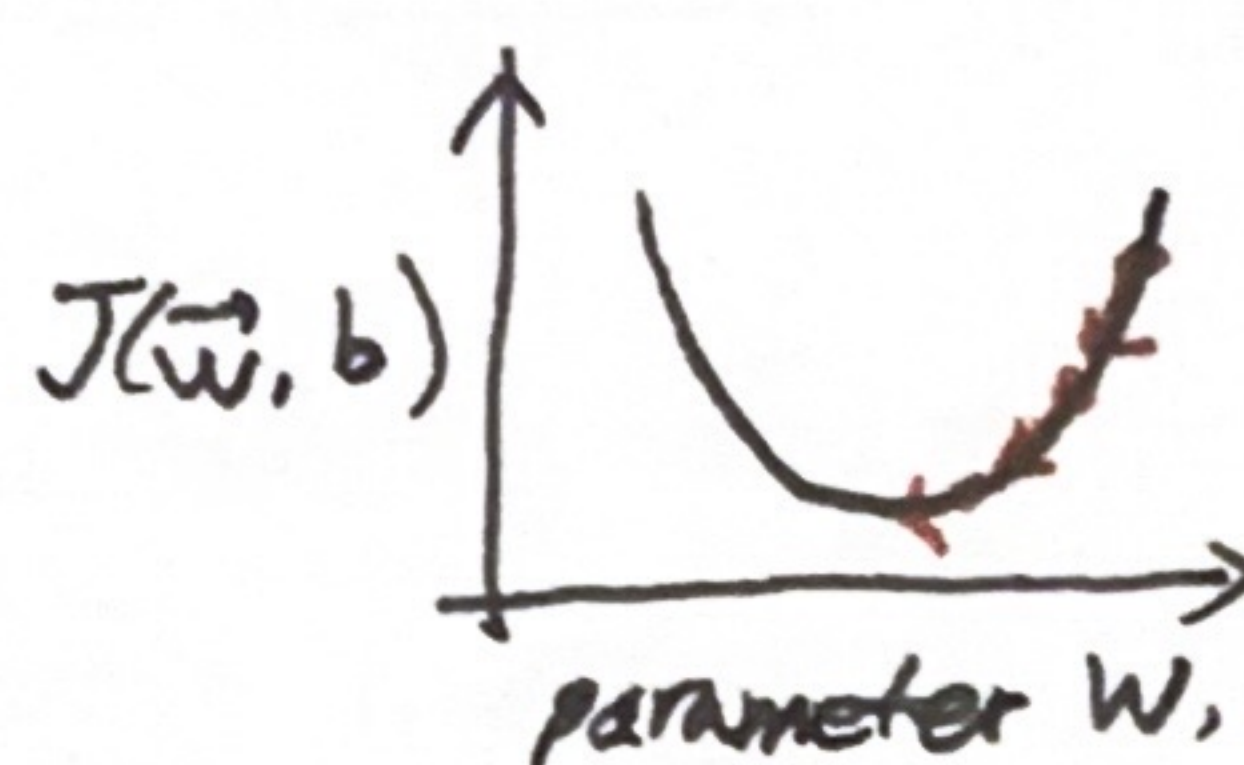
case #1:



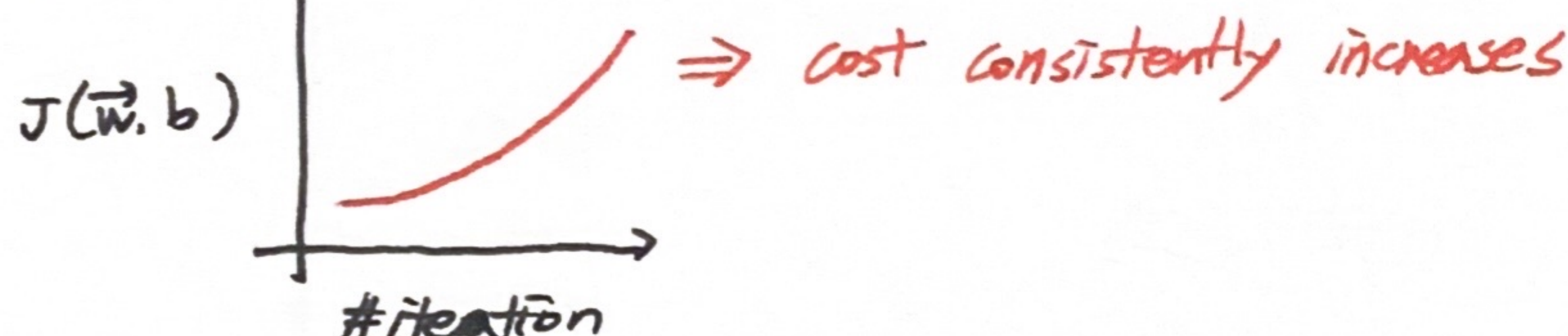
reason #1: Too large value of learning rate



To fix this problem, Use smaller learning rate



case #2:



reason #2: Bug on a code

ex) gradient descent code:  $w := w + \alpha \underset{\substack{\text{derivate} \\ \text{term}}}{d}$   $\xrightarrow{\text{fix}}$   $w := w - \alpha d$

\* Debugging tip: set very small  $\alpha$  and see if that causes the cost to decrease on every iteration

Trade-off

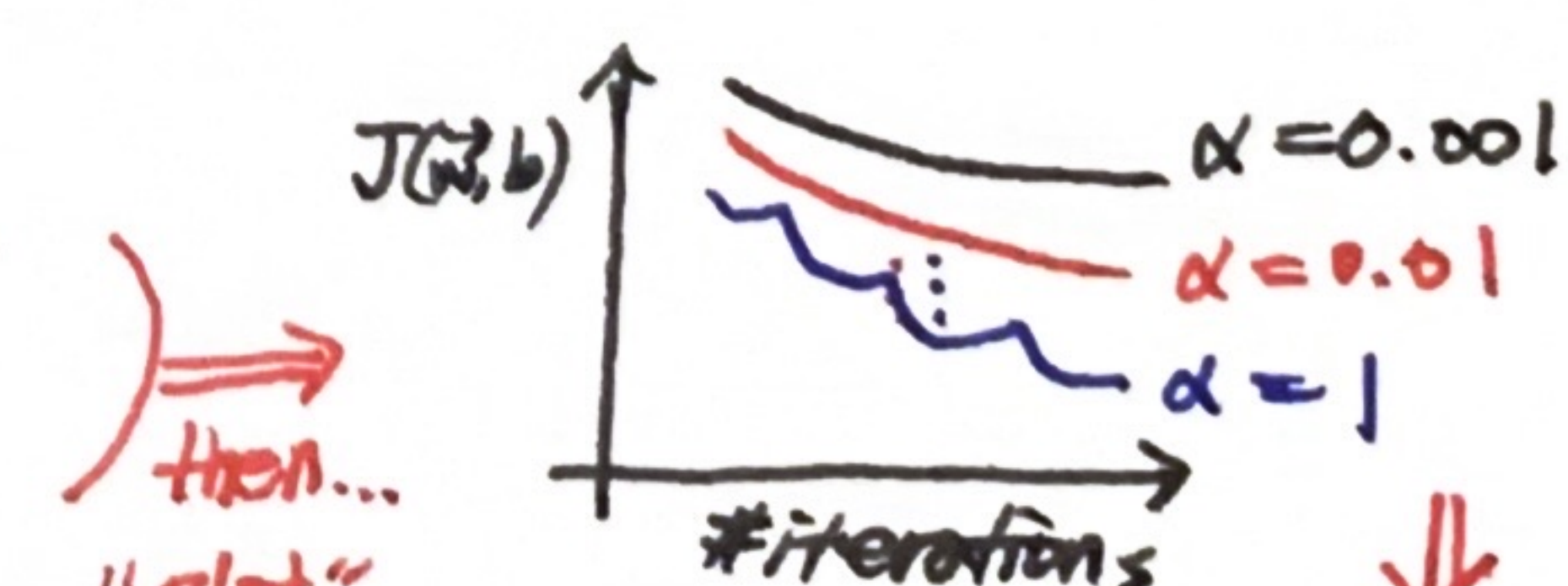
$\Rightarrow$  if, even with very small number of alpha,  $J(\text{cost})$  doesn't decrease on every iteration that usually means there's a bug somewhere in my code.

gradient descent can take a lot of iterations to converge

\* Setting value of  $\alpha$  tip: set a range of values for  $\alpha$

ex) ... 0.001 0.01 0.1 1 ...

or 0.001 0.03 0.01 0.03 0.1 0.3 1 ...  
 $\times 3$   $\times 3 \approx$   $\times 3$   $\times 3 \approx$   $\times 3$   $\times 3 \approx$



then...  
"plot" learning curve

\* "pick value of  $\alpha$  that makes cost to decrease rapidly and consistently"