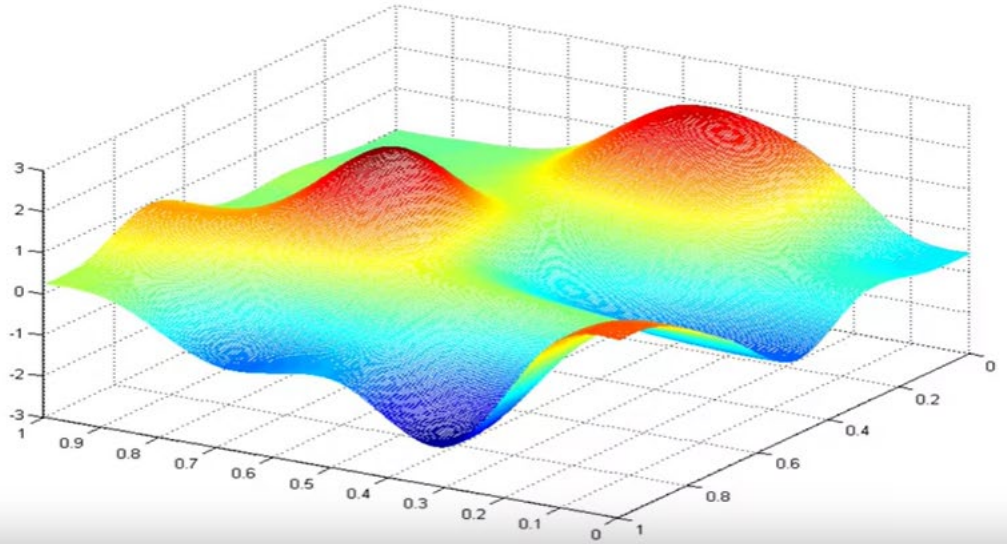
Training Linear Regression

Gradient Descent

Gradient Descent

Have some function $J(w,b)$

Want $\min_{w,b} J(w,b)$



Outline:

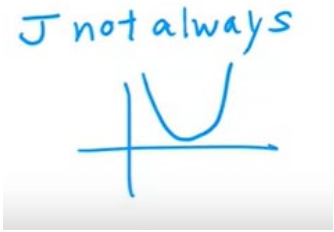
Start with some w,b (set $w=0$, $b=0$)

Keep changing w,b to reduce $J(w,b)$

for linear regression or any function

$\min_{w_1,\ldots,w_n,b} J(w_1,w_2,\ldots,w_n,b)$

J not always

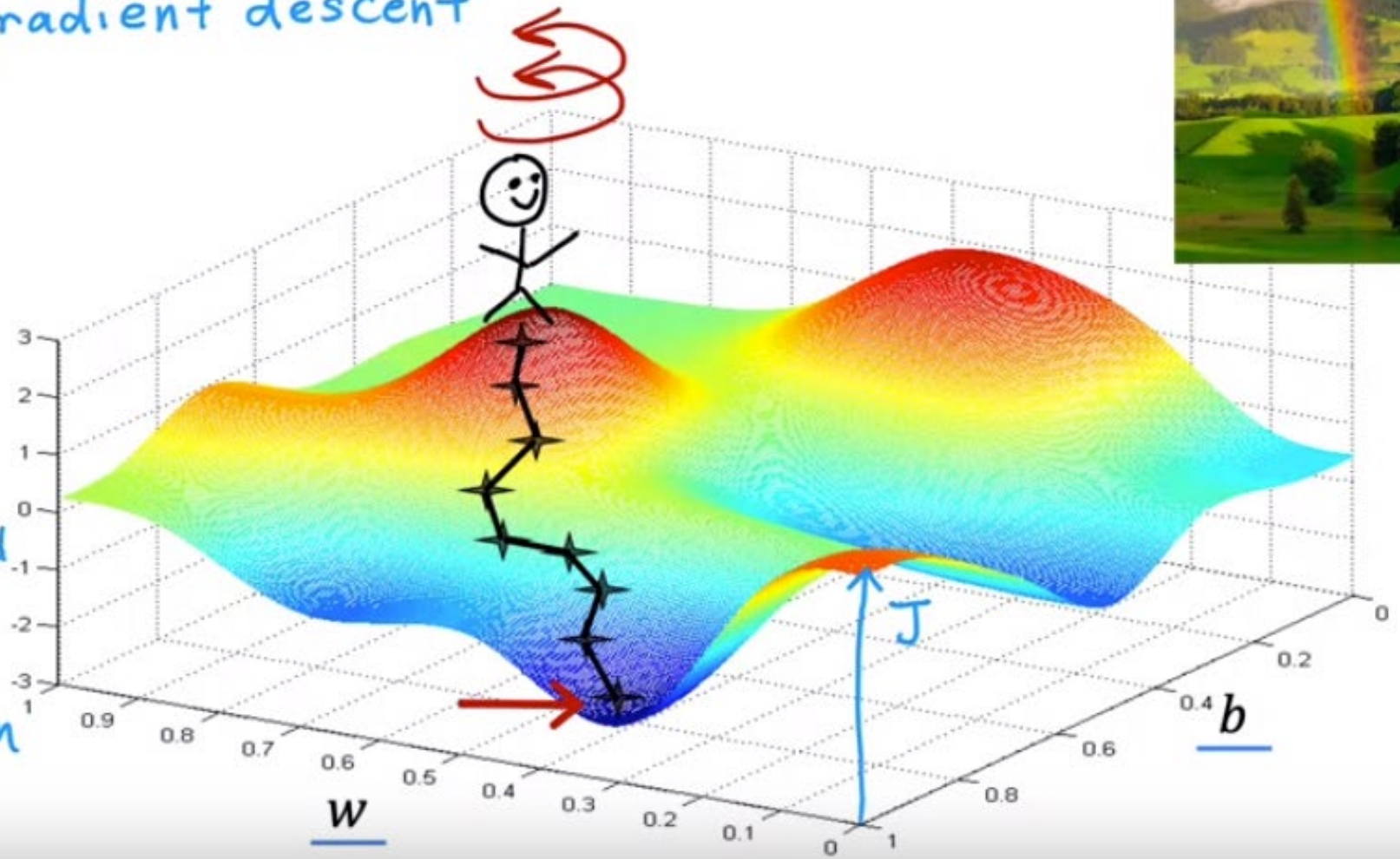


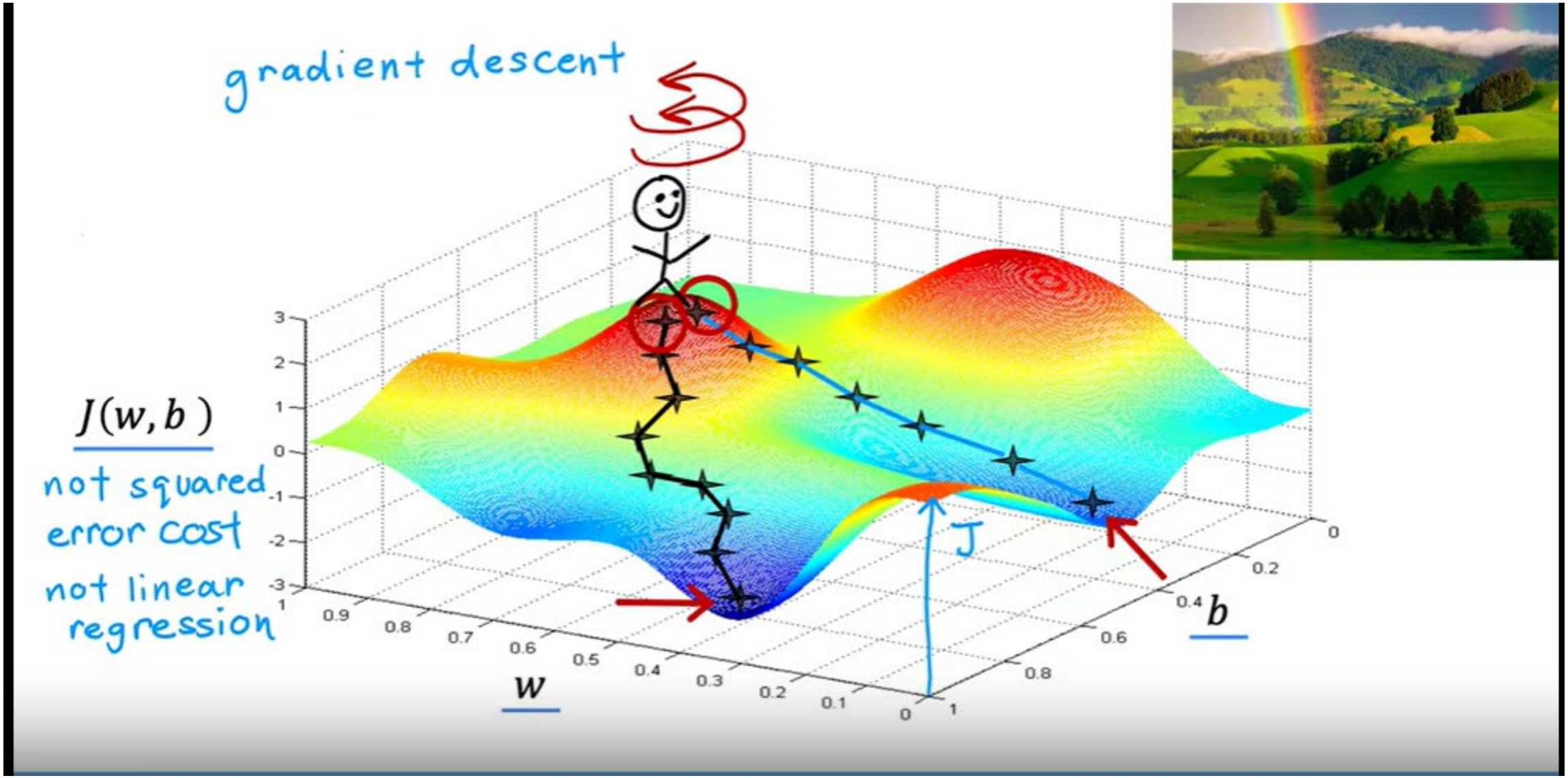
gradient descent

$J(w,b)$

not squared error cost

not linear regression




gradient descent
$J(w,b)$
not squared error cost
not linear regression
J
w
b

Training Linear Regression

Implementing Gradient Descent

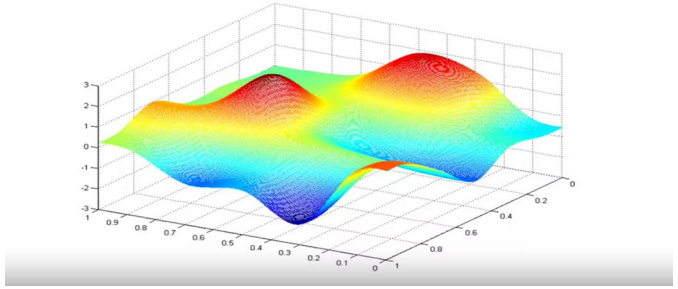
Gradient Descent Algorithm

$$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

Learning rate

Derivative



Repeat until convergence

Correct: Simultaneous update

$$tmp\_w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$tmp\_b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$w = tmp\_w$$

$$b = tmp\_b$$

Incorrect

$$tmp\_w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$w = tmp\_w$$

$$tmp\_b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

$$b = tmp\_b$$

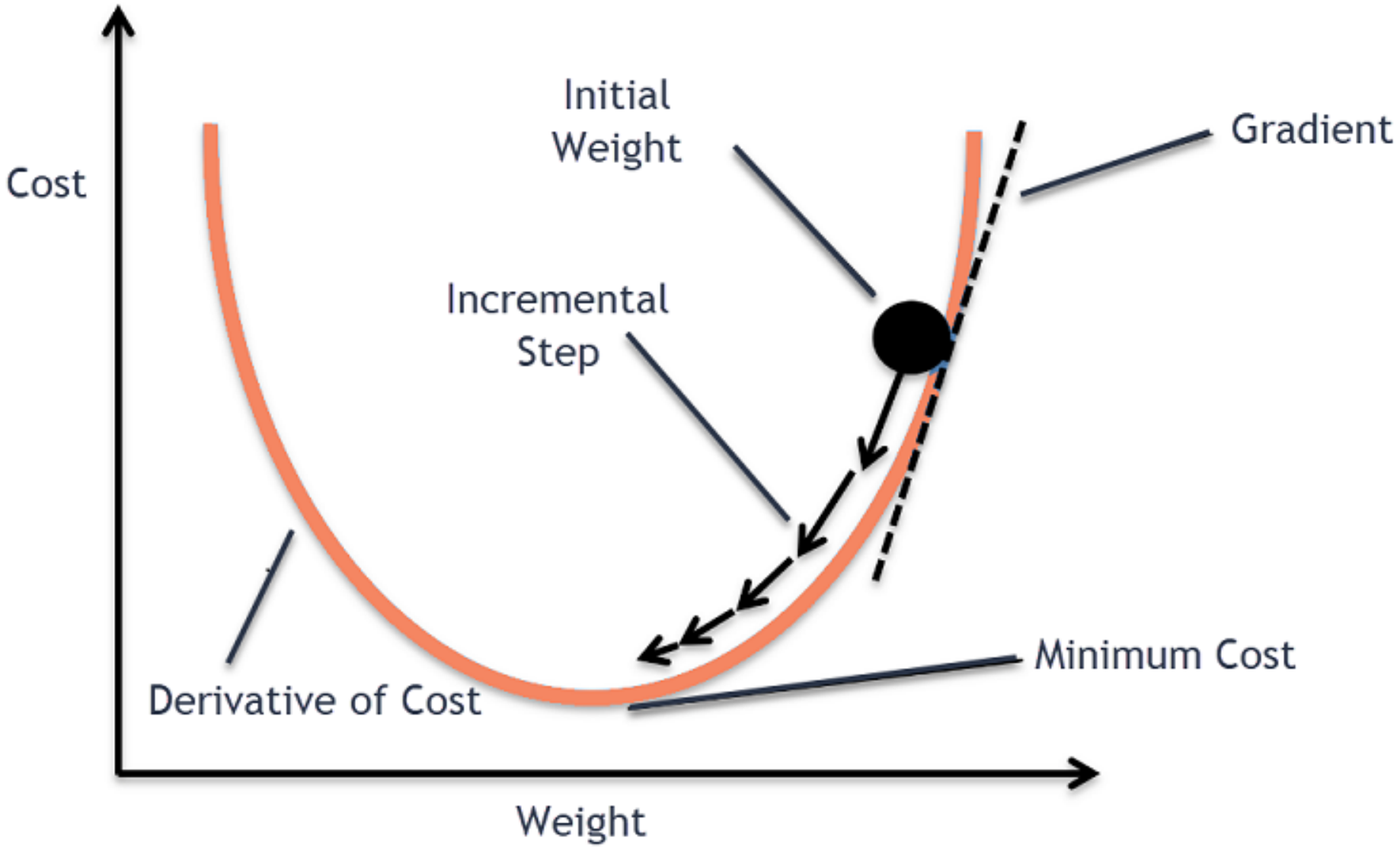
Quiz

Gradient descent is an algorithm for finding values of parameters w and b that minimize the cost function J. What does this update statement do? (Assume α is small.)

- ○ Checks whether w is equal to $w - \alpha \frac{\partial J(w,b)}{\partial w}$
- ○ Updates parameter w by a small amount

Training Linear Regression

Gradient Descent Intuition


Cost
Initial
Weight
Gradient
Incremental
Step
Derivative of Cost
Minimum Cost
Weight

Linear Algebra – bare minimum

- Negative slope
- Positive slope
- Tangent
- When do we get:

a) W = increments

b) W = decrements

In order to get J(w) minimum, the value of w increments or decrements?

Quiz

Assume the learning rate α is a small positive number. When $\frac{\partial J(w,b)}{\partial w}$ is a positive number (greater than zero) -- as in the example in the upper part of the slide shown above -- what happens to w after one update step?

- w decreases.
- w increases
- w stays the same
- It is not possible to tell if w will increase or decrease.

Training Linear Regression

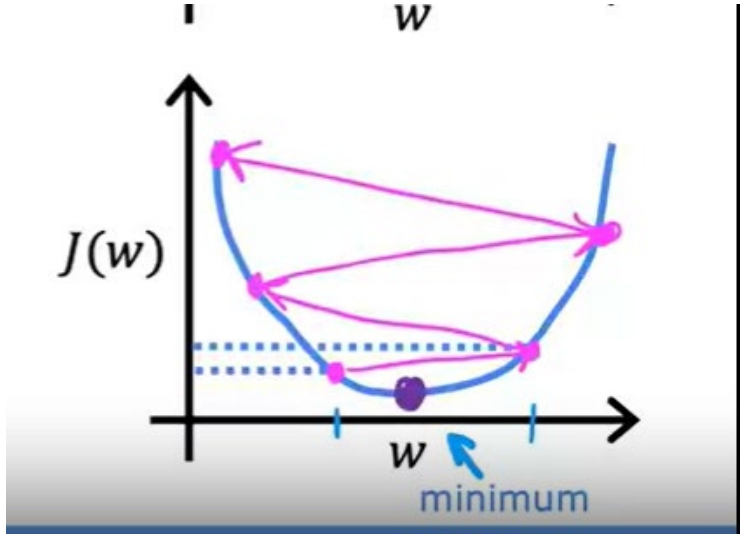
Learning Rate

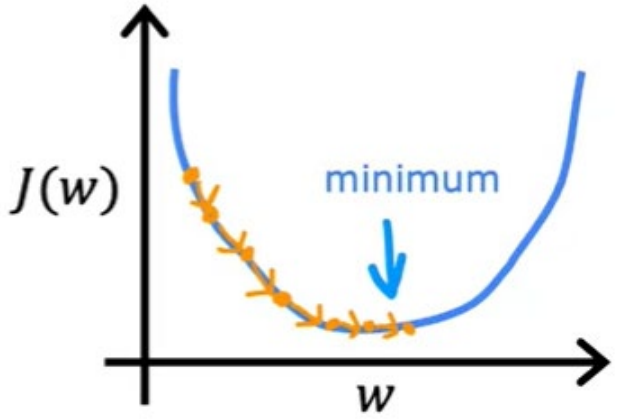
Answer the following questions

- What do u think happens when learning rate is too small?
- Too large?

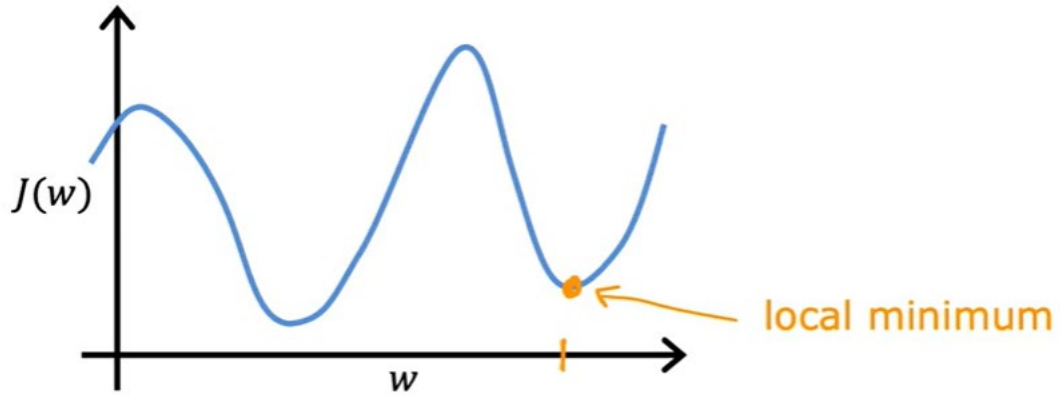
Answer the following questions

- What do u think happens when learning rate is too small?
- Too large?



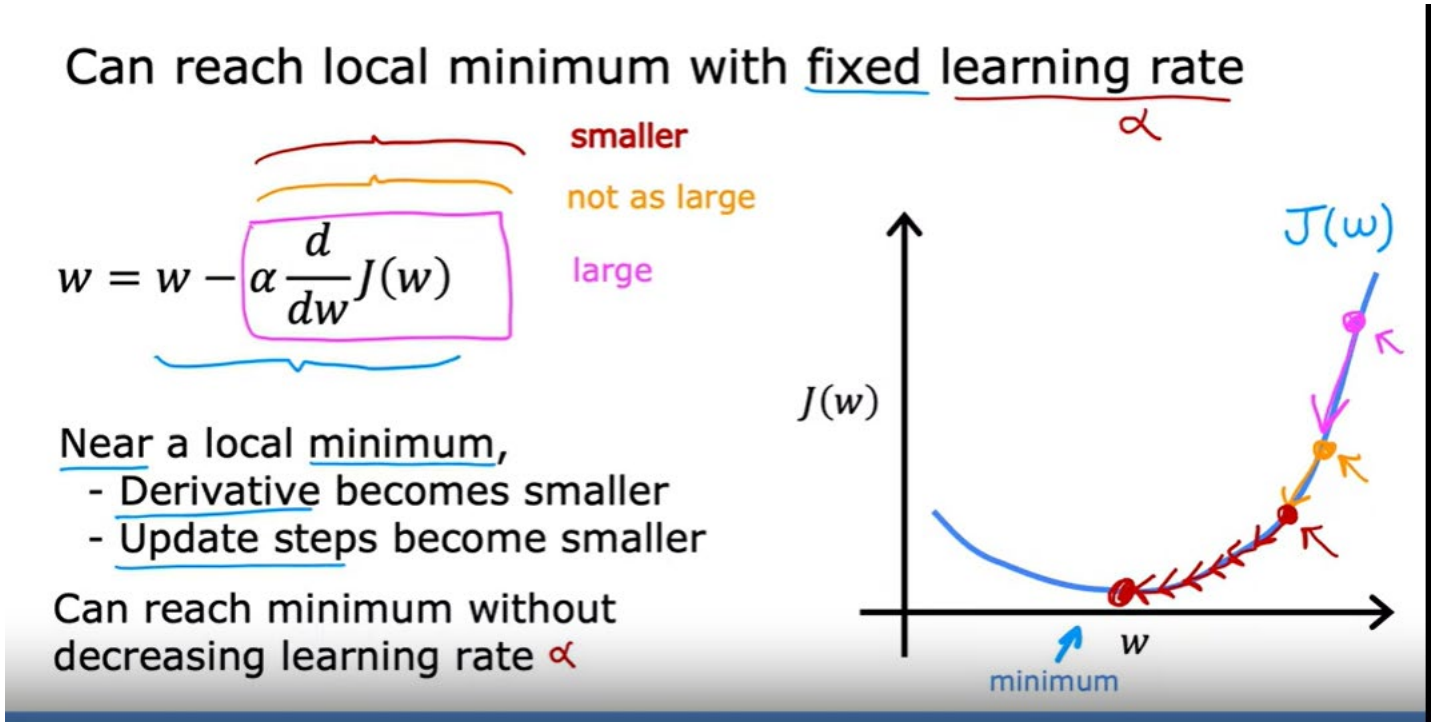


Can also reach local minima with fixed α



$$w = w - \alpha \frac{d}{dw} J(w)$$

Can also reach local minima with fixed α



Gradient Descent Summary

- Gradient descent is an optimization algorithm used to minimize the cost function of a model
- The cost function measures how well the model fits the training data and is defined based on the difference between the predicted and actual values.
- The gradient of the cost function is the derivative with respect to the model's parameters and points in the direction of the steepest descent.
- The algorithm starts with an initial set of parameters and updates them in small steps to minimize the cost function.

Gradient Descent Summary

- In each iteration of the algorithm, the gradient of the cost function with respect to each parameter is computed.
- The size of the step is controlled by the learning rate, which determines how quickly the algorithm moves towards the minimum.
- The process is repeated until the cost function converges to a minimum, indicating that the model has reached the optimal set of parameters.

Gradient Descent Summary

- There are different variations of gradient descent, including batch gradient descent, stochastic gradient descent, and mini-batch gradient descent, each with its own advantages and limitations.
- Efficient implementation of gradient descent is essential for achieving good performance in machine learning tasks. The choice of the learning rate and the number of iterations can significantly impact the performance of the algorithm.