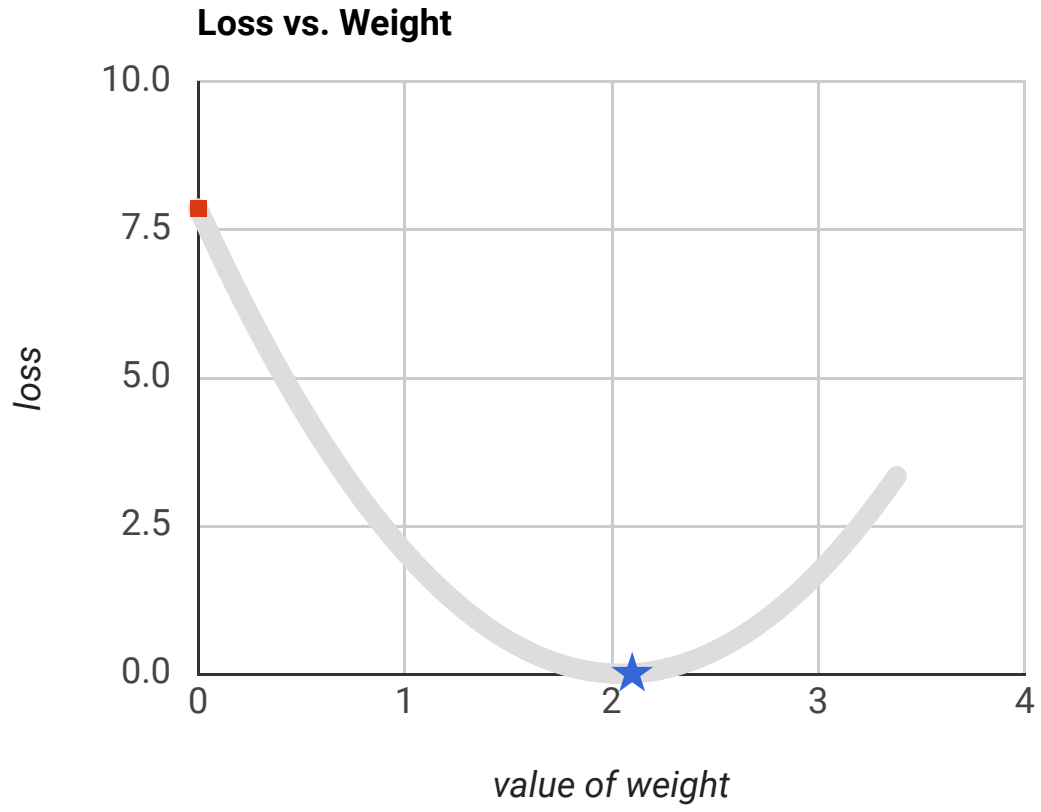
Optimizing Learning Rate

Estimated Time: 15 minutes

Experiment with different learning rates and see how they affect the number of steps require minimum of the loss curve. Try the exercises below the graph.

Set learning rate:		0.01
Execute single step:	STEP	0
Reset the graph:	RESET	



Exercise 1

Set a learning rate of 0.1 on the slider. Keep hitting the STEP button until the gradient descent algorithm reaches the minimum point of the loss curve. How many steps did it take?

Solution

Gradient descent reaches the minimum of the curve in 81 steps.

Exercise 2

Can you reach the minimum more quickly with a higher learning rate? Set a learning rate of 1, and keep hitting STEP until gradient descent reaches the minimum. How many steps did it take this time?

Solution

Gradient descent reaches the minimum of the curve in 6 steps.

Exercise 3

How about an even larger learning rate. Reset the graph, set a learning rate of 4, and try to reach the minimum of the loss curve. What happened this time?

Solution

Gradient descent **never reaches the minimum**. As a result, steps progressively increase in size. Each step jumps back and forth across the bowl, climbing the curve instead of descending to the bottom.

Optional Challenge

Can you find the Goldilocks (https://wikipedia.org/wiki/Goldilocks_principle) learning rate for this curve, where gradient descent reaches the minimum point in the fewest number of steps?

What is the fewest number of steps required to reach the minimum?

Solution

The Goldilocks learning rate for this data is 1.6, which reaches the minimum in 1 step.

NOTE: In practice, finding a "perfect" (or near-perfect) learning rate is not essential for successful model training. The goal is to find a learning rate large enough that gradient descent converges efficiently, but not so large that it never converges.

HELP CENTER (HTTPS://SUPPORT.GOOGLE.COM/MACHINELEARNINGEDUCATION?AUTHUSER=1)

Previous
← Learning Rate
(https://developers.google.com/machine-learning/crash-course/reducing-loss/learning-rate?authuser=1)

Next
Stochastic Gradient Descent →
(https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent?authuser=1)