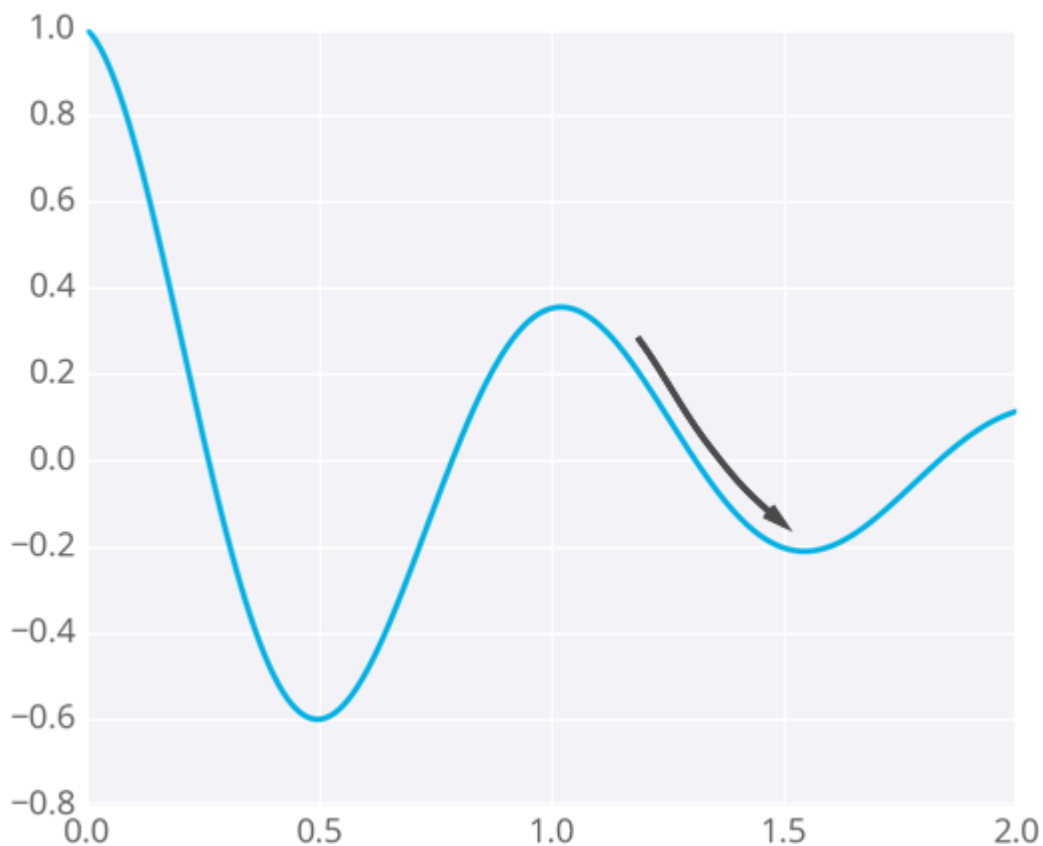


Gradient descent steps to the lowest error

## Caveats

Since the weights will just go wherever the gradient takes them, they can end up where the error is low, but not the lowest. These spots are called local minima. If the weights are initialized with the wrong values, gradient descent could lead the weights into a local minimum, illustrated below.



Gradient descent leading into a local minimum

There are methods to avoid this such as using [momentum](#)