Embrace Pessimism

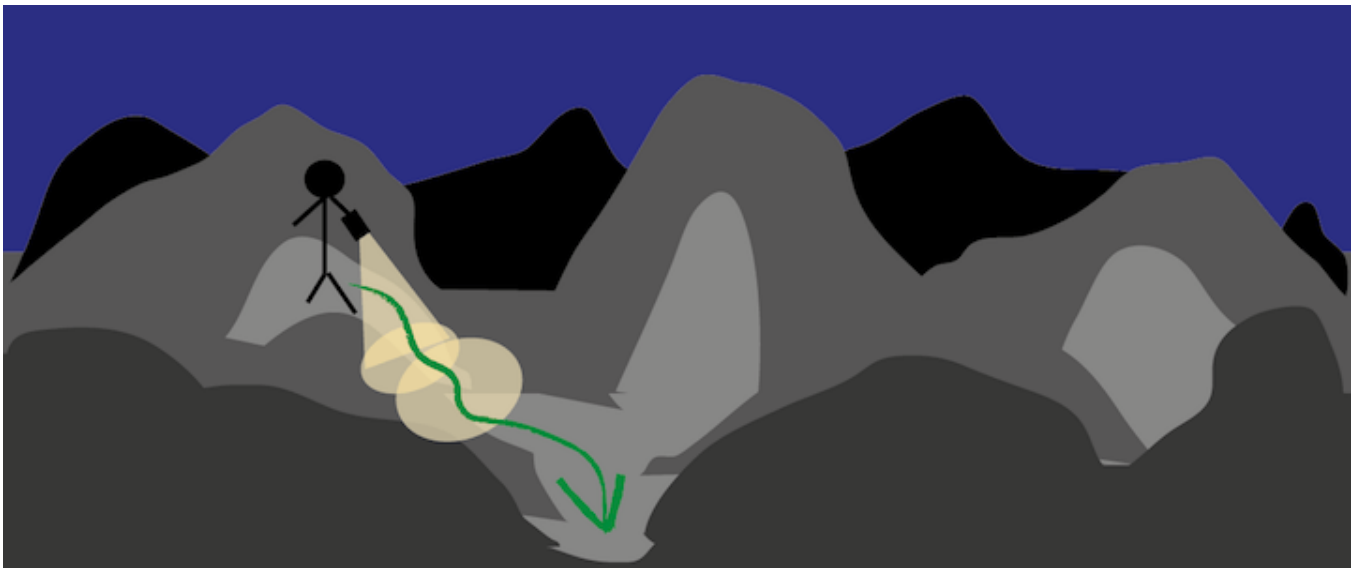
A little background on Gradient Descent.

You can see that the brute force approach isn't practical at all. In fact, it gets worse very quickly as we add network layers, nodes or possibilities for weight values. This puzzle resisted mathematicians for years and was only really solved practically as late as the $1960s - 70$s. There are different views on who did it first or made the key breakthrough, but the important point is that this late discovery led to the explosion of modern neural networks which can carry out some very impressive tasks. So how do we solve such an apparently hard problem? Believe it or not, you've already got the tools to do it yourself. We've covered all of them earlier. So let's get on with it. The first thing we must do is embrace *pessimism*.

The mathematical expressions showing how all the weights result in a neural network's output are too complex to untangle easily. The weight combinations are too many to test one by one to find the best. There are even more reasons to be pessimistic. The training data might not be sufficient to properly teach a network. The training data might have errors so our assumption that it is a perfect truth, something to learn from, is then flawed. The network itself might not have enough layers or nodes to model the right solution to the problem. What this means is we must take an approach that is realistic and recognizes these limitations. If we do that, we might find an approach which isn't mathematically perfect but does actually give us better results because it doesn't make false idealistic assumptions.

Let's illustrate what we mean by this. Imagine a very complex landscape with peaks and troughs, and hills with treacherous bumps and gaps. It's dark, and you can't see anything. You know you're on the side of a hill, and you need to get to the bottom. You don't have an accurate map of the entire landscape. You do have a torch. What do you do? You'll probably use the torch to look at the area close to your feet. You can't use it to see much further anyway, and certainly not the entire landscape. You can see which bit of earth seems to be

going downhill and take small steps in that direction. In this way, you slowly work your way down the hill, step by step, without having a full map and without having worked out a journey beforehand.



The mathematical version of this approach is called *gradient descent*, and you can see why. After you've taken a step, you look again at the surrounding area to see which direction takes you closer to your objective, and then you step again in that direction. You keep doing this until you're happy you've arrived at the bottom. The gradient refers to the slope of the ground. You step in the direction where the slope is steepest downwards.