

Training Options

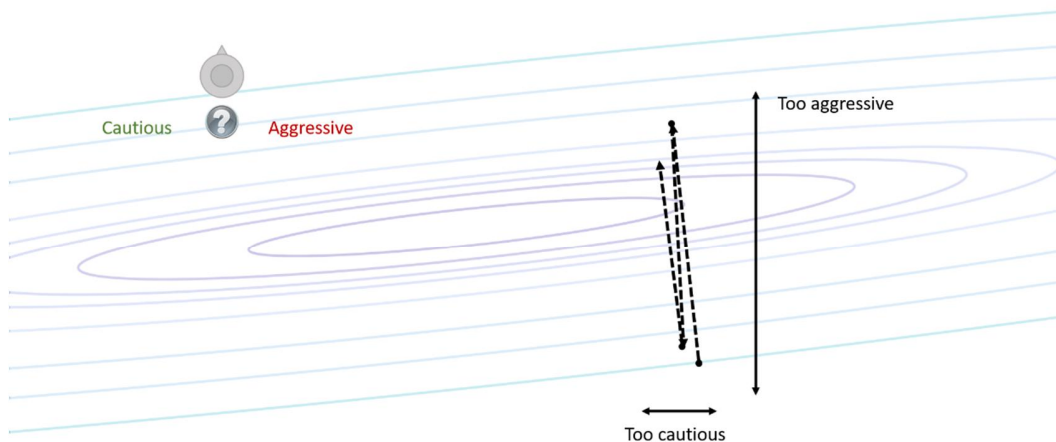
A basic understanding of how training works can help you sift through the many training options available. The next few pages will break down some of the important ones:

- Mini-batches
- Learning rates
- Gradient clipping
- Algorithms

Algorithms

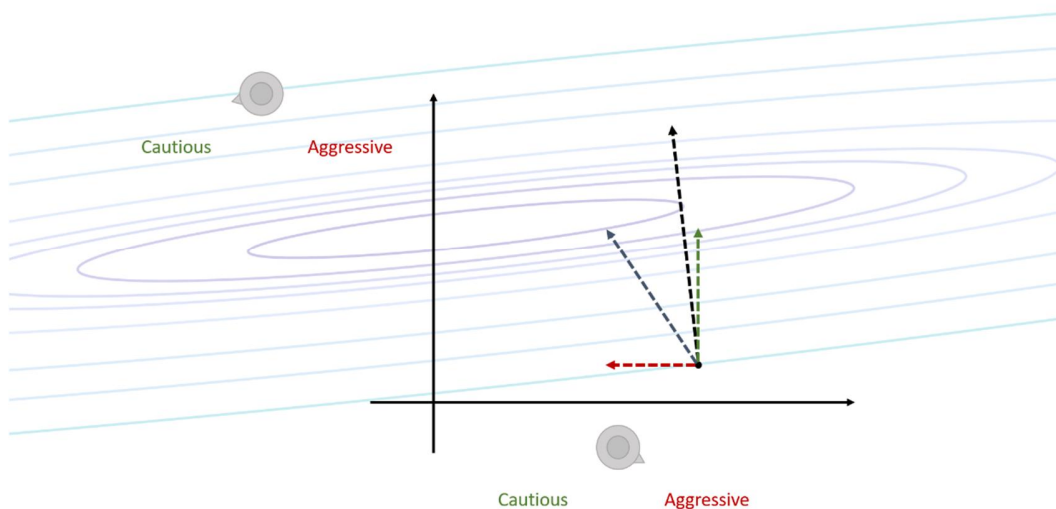
The first input to `trainingOptions` is the name of the training algorithm. There are '`sgdm`' (stochastic gradient descent with momentum) and other algorithms you can use.

1.



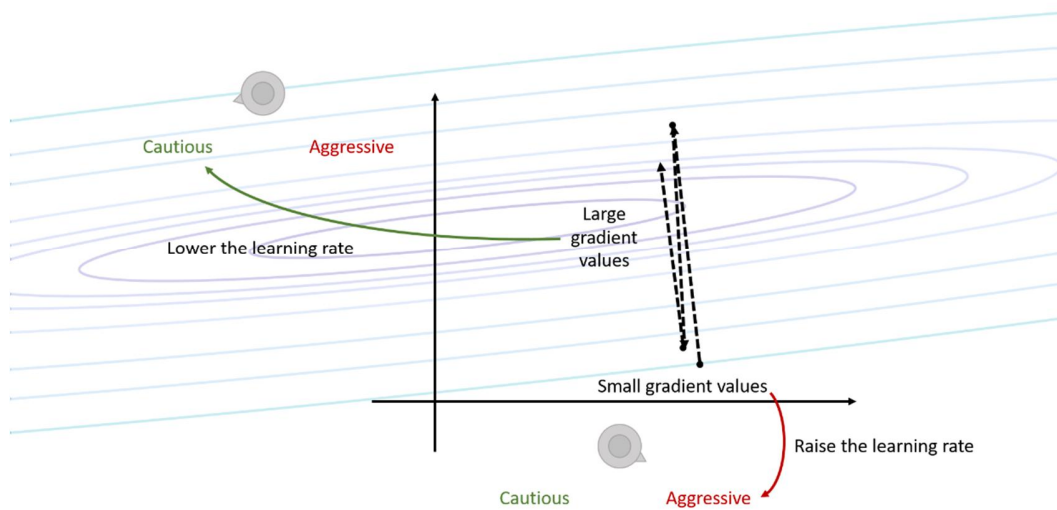
Even with an upper limit on the gradient values, if some values are at the maximum and others are near zero, it can be hard to choose just the right learning rate. You want smaller steps in the direction of rapid change, but larger steps in the direction of slow change.

2.



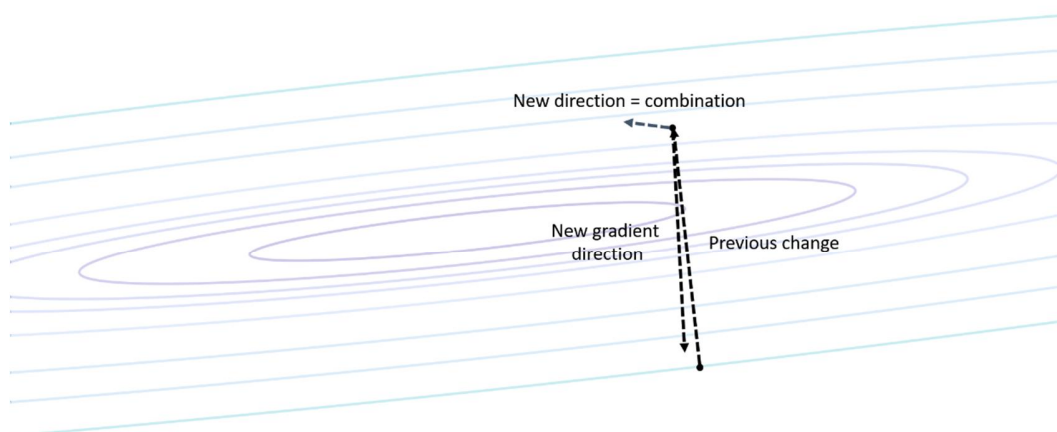
A simple fix is to do just that: use a different learning rate for each parameter.

3.



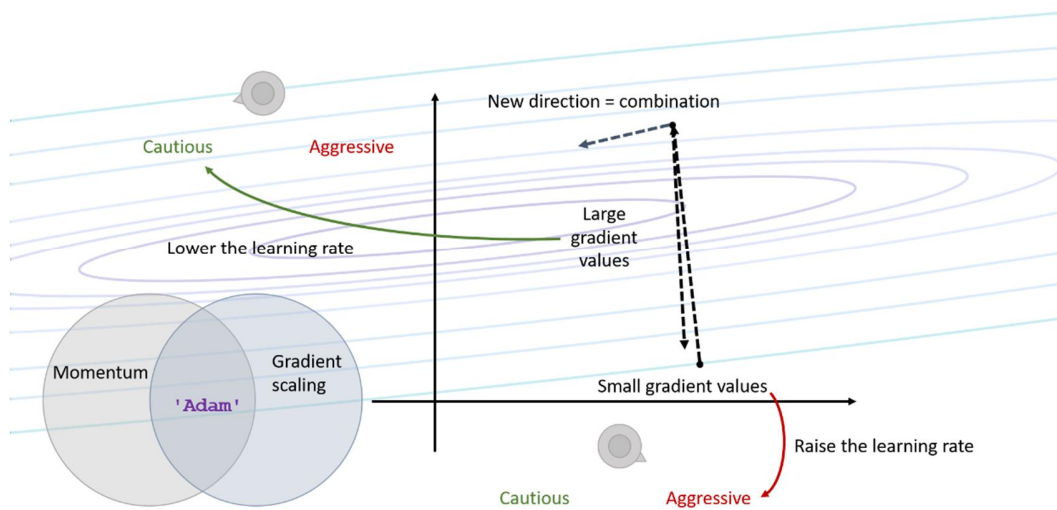
The **RMSProp** algorithm keeps a history of the size of the gradient, and uses this to scale the learning rate for each parameter.

4.



You've seen that including momentum can prevent the updates jumping wildly back and forth.

5.



The **adaptive moment estimation**, or “**Adam**”, algorithm scales the learning rate for each parameter, like RMSProp, but also uses momentum to smooth out the updates.