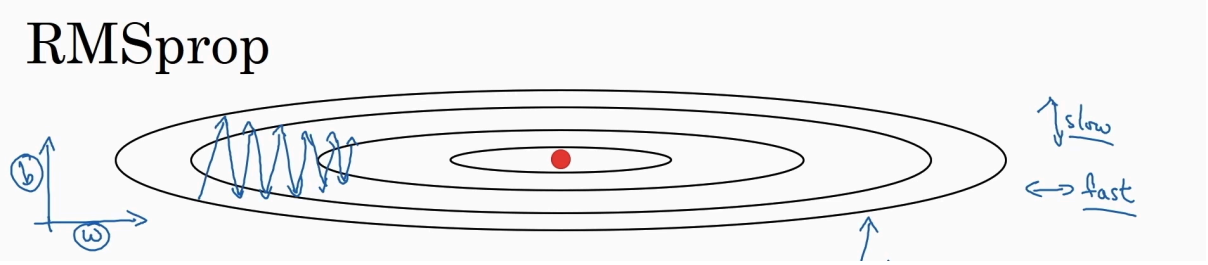
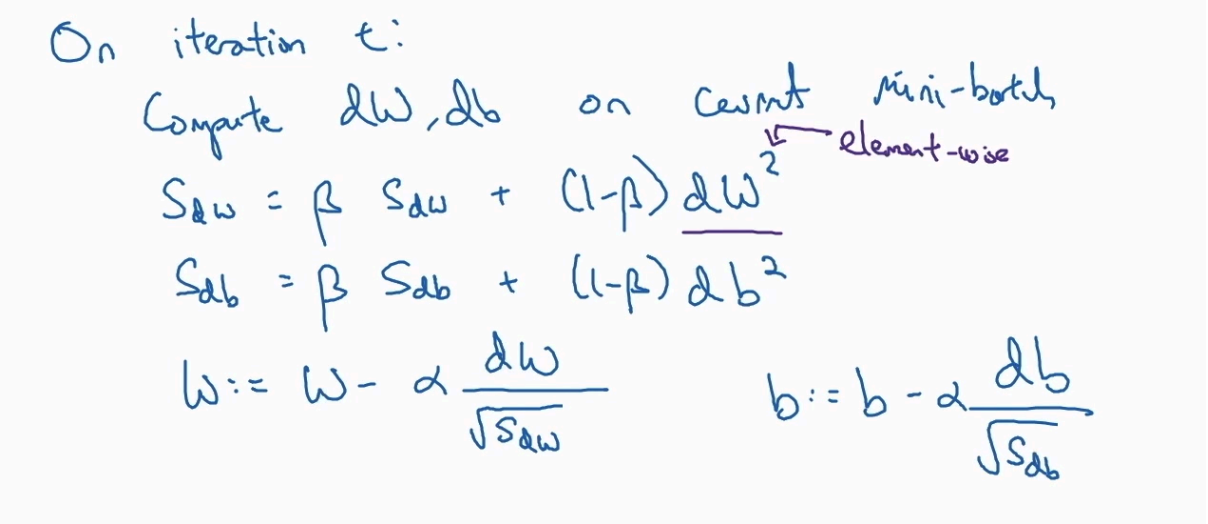
RMSProp

RMSPop → Root Mean Square Propagation



Want to:

Slow down learning in the b direction and increase learning in the W direction.



Do note that the squaring is done element wise!

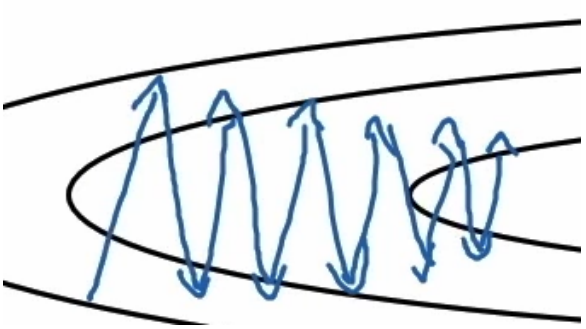
So what this is doing is keeping an EWA of the SQUARES of the derivatives.

Then look carefully at the update rule! We don't directly update by the EWA!

So, goal is to increase learning in the W direction and decrease/dampen learning in the b direction.

So what we hope is that S_dW is small so that when you divide by a smaller number, then the update to W is large. And we hope S_db is large so that when you divide by a larger number, then the update to b is small in order to slow down updates in the vertical direction.

And indeed, the dW's are much smaller than the db's!



Here, db is much larger compared to dW

Therefore oscillations in the vertical direction get damped out!

Think of this as adaptive learning rates!