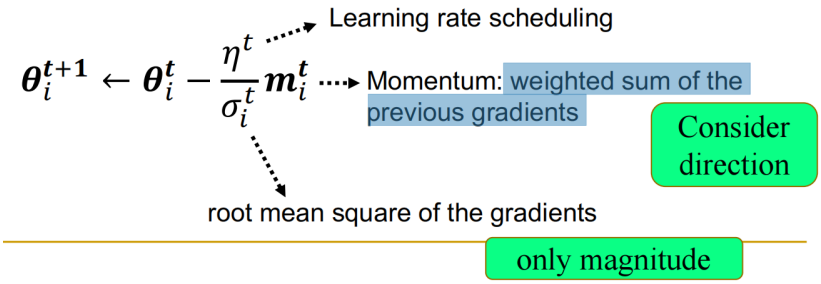
- lr：learning rate

其实不是local minimum导致模型不够好

小梯度

- mini-batch
- Momentum： weighted sum of the previous gradients 不会卡在里面，带方向

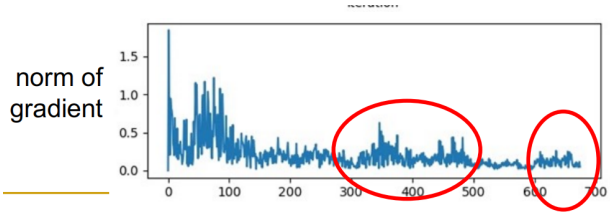


Adaptive Learning Rate

Training stuck

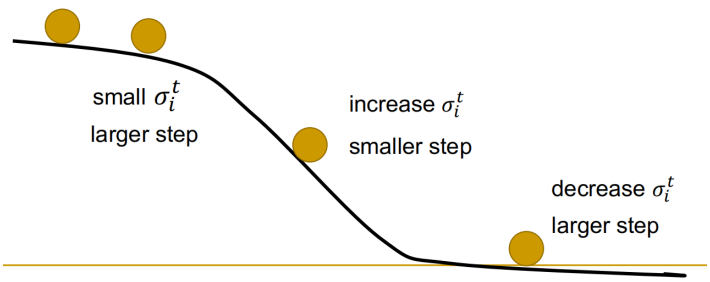
> 与小梯度不同

- 比较小，但难以继续降
- 原因：lr大，错过了小峡谷。但lr太小更新步长太短

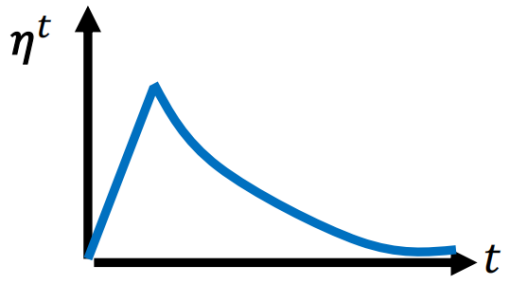


RMSProp

更新适中：梯度大时缩小lr，梯度小时扩大lr



- Adam: RMSProp + Momentum
- warm up：First Increase and then decrease



Leaky ReLU

它是一种专门设计用于解决Dead ReLU问题的激活函数：

$$\mathrm{LeakyReLU}\left(\mathrm{x}\right)=\begin{cases}\mathrm{x} & ,\mathrm{x}>0\\ \alpha\mathrm{x} & ,\mathrm{x}\leq 0\end{cases}$$

Leaky ReLU函数的特点：

- Leaky ReLU函数通过把x的非常小的线性分量给予负输入$0.01\mathrm{x}$来调整负值的零梯度问题。
- Leaky有助于扩大ReLU函数的范围，通常α的值为0.01左右。
- Leaky ReLU的函数范围是负无穷到正无穷。

- rounding error：计算机二进制存储数值与真实值的误差
-