参数说明

参数更新公式

*V**t*+1=*μ**V**t*−*α*∇*L*(*W**t*)

*W**t*+1=*W**t*+*V**t*+1

*base\_lr:*基础学习速率，对应*α*

*gamma:*调节学习速率的因子 *α =α \*gamma^(floor(iter/stepsize))*

*stepsize:*每隔*stepsize*降低学习速率，每隔*stepsize*，基础学习速率*\*gamma*

*power:*调节学习速率需要用到的参数

*lr\_policy:*

*"fixed":*固定学习速率，始终等于*base\_lr*

*"step":*步进衰减，*base\_lr\*gamma^(floor(iter/stepsize))*

*"exp":*指数衰减，*base\_lr\*gamma^(iter)*

*"inv":*倒数衰减，*base\_lr\*(1+gamma\*iter)^(-power)*

*"multistep":*多步衰减，与步进衰减类似，允许非均匀步进值*(stepvalue)*

*"ploy":*多项式衰减，在*max\_iter*时达到*0*，*base\_lr\*(1-iter/max\_iter)^(power)*

*"sigmoid":S*形衰减，*base\_lr\*(1/(1+exp^(-gamma\*(iter-stepsize))))*

*momentum:*冲量，灵感来自于牛顿第一定律，基本思路是为寻优加入了“惯性”的影响，这样一来，当误差曲面中存在平坦区的时候，*SGD*可以更快的速度学习。对应*μ*