由于V是在变化的，当前对W有效的数据不一定对之后新的V有效，所以刚开始在W和V不稳定的时候A的浮动很大。有些数据可能对提升V完全没帮助就weight很小

最后W和V趋于稳定的时候，在最优点附近震动，那么一个数据的weight对震动而言不是很重，导致最后相同。

两层FC参数量不够，学不明白，效果不好。

先让W学一个epoch导致，w能很确定的怎么根据数据教好这个不稳定的V，导致A发散了，效果不好

算loss是用logit算的是否有影响