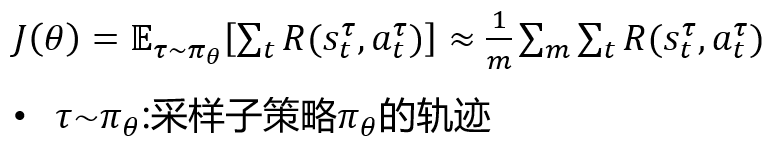
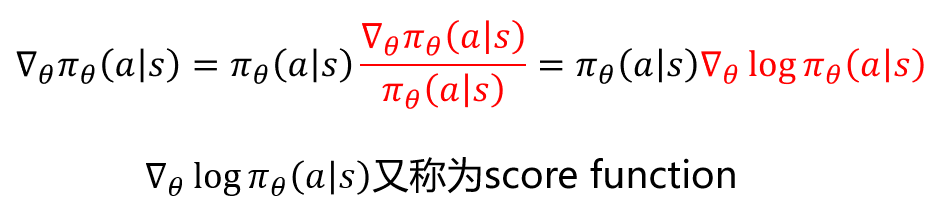
基于策略RL：常收敛到局部最优、评估策略时存在较大的方差

策略评估函数：

策略的梯度：