

Live Demo of Dynamic Programming & Sample-Based Methods

Rupam Mahmood

February 7, 2020







Live demo of policy evaluation and Monte Carlo estimation



Iterative Policy Evaluation, for estimating $V \approx v_{\pi}$

Input π , the policy to be evaluated

Loop: $\Delta \leftarrow 0$ Loop for each $s \in S$: $v \leftarrow V(s)$ $V(s) \leftarrow \sum_{a} \pi(a|s) \sum_{a} \nabla(a|s) \sum_{a} \nabla(a$ $\Delta \leftarrow \max(\Delta, |v - V$ until $\Delta < \theta$

 $v_{\pi}(0) \doteq E_{\pi} \left[G_0 \,|\, S_0 = 0 \right] \approx \sum_{n=1}^{\infty} -1$

- Algorithm parameter: a small threshold $\theta > 0$ determining accuracy of estimation Initialize V(s), for all $s \in S^+$, arbitrarily except that V(terminal) = 0

$$\sum_{\substack{s',r\\(s)|}} p(s',r|s,a) \left[r + \gamma V(s')\right]$$

