AI Theory, L6/1

Review of what MDPs and policies and utilities are.

Assignment:

using reinforce.js , deploy a GUI version of Bridge crossing MDP.

Algorithm: Policy Evaluation

$\rightarrow$ Initialize $V_n^0 = (0)$ for all states

→ for iteration $t = 1, 2, 3, \ldots, t_p$

for each state s

$$V_n^t = \sum_{s'} T(s, a, s') \left[R(s, a, s') + \gamma V_n^{(t-1)}s \right]$$

for this, assume Q-value

$= V$ because we have

only one action.