

Deterministic optimal policy for DYNA-Q+ AS 0.01 after 6000 iterations

