

Lecture 4: Optimal Values and Policies

CS 639: Introduction to Reinforcement Learning

Tengyang Xie

Tuesday, September 15, 2026

Quiz 1

- 15 minutes, 11:00–11:15. Closed book; the formulas you need are printed on the sheet.
- Name, UW NetID, and Campus ID at the top.
- Final answers go in the table at the top: numbers for 1(a)–(c), one letter for 2(a)–(b).
 - Exact fractions or decimals are fine. No partial credit.
- Work alone. No notes, phones, or laptops.
- Finished early? Turn the sheet face down and wait. Sheets are collected at 11:15.

Housekeeping

- Quiz 1 is in. Solutions and scores follow on Canvas.
 - Best 5 of 6 quizzes count; no make-ups.
- HW1 comes out Thu 09/17, due Fri 10/02 on Gradescope.
- Office hours
 - Tengyang Xie: Wed 1–2pm, MH 5530
 - Sarina Xi: Mon 9–11am, MH 5683
 - Yurun Yuan: Fri 9–11am, MH 5683
- Questions: Piazza first; private posts are fine.

All eight

$\pi(G), \pi(W), \pi(B)$		$V^\pi(G)$	$V^\pi(W)$	$V^\pi(B)$
run, run, run	never repair	27.5	10	0
run, run, repair	repair when broken	44.30	32.41	29.87
run, repair, run		38.46	24.62	0
run, repair, repair	repair when worn	38.46	24.62	24.62
repair, run, run		-100	10	0
repair, run, repair		-100	-65	-100
repair, repair, run		-100	-100	0
repair, repair, repair	always repair	-100	-100	-100

One linear solve per row, $\gamma = 0.9$. Repairing a good machine restarts the repair bill forever: -100 from G. One row is largest in every column, as the theorem promises.

Before you leave, and next time

- Next time, Thu 09/17: value iteration. Computing V^* by iteration with the max inside, then reading π^* off it.
- Reading: Sutton & Barto, Section 3.6 (today); Section 4.4 for next time.
- HW1 comes out Thu 09/17 after class (planned).
- Quiz 1 solutions and scores follow on Canvas.