

Lecture 9: Temporal-Difference Learning

CS 639: Introduction to Reinforcement Learning

Tengyang Xie

Thursday, October 1, 2026

Housekeeping

- HW1 is due Fri 10/02 on Gradescope; hard close 48 hours later.
 - Four slip days shared by HW1–HW4, at most two on one assignment; applied automatically.
- HW2 comes out Tue 10/06.
- Reading for today: Sutton and Barto, Sections 6.1–6.3 (Szepesvári, Section 3.1.1, optional).
- 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.

Error against weeks, by computer

mean error, week t =	50	100	200	500	1,000	2,000
TD(0), $\alpha = 0.1$	32.05	28.08	20.85	9.18	3.56	2.47
TD(0), $\alpha = 0.3$	25.75	18.52	11.08	6.56	6.21	6.29
TD(0), $\alpha = 1/n(S_t)$	26.37	24.22	22.37	20.17	18.70	17.37
Monte Carlo, 44-week cut	18.29	4.74	3.27	1.74	1.25	0.96
certainty equivalence	4.17	3.35	2.45	1.30	0.97	0.73
batch TD(0) on the first t weeks	= certainty equivalence, by the theorem					

The experiment: start the machine at G and run π = repair-when-broken ($\gamma = 0.9$) for 2,000 weeks; that is one stream. Each method turns the first t weeks into a table V_t , starting from $V_0 = 0$; the entry shown is the error $\max_s |V_t(s) - V^\pi(s)|$, averaged over 200 independent streams. Monte Carlo averages 44-week cut returns, each counted once complete; certainty equivalence solves the chain counted from the same t weeks; batch TD(0) on those weeks equals it, by the theorem.

Before you leave

- Reading: Sutton and Barto, Sections 6.1–6.3; Szepesvári, Section 3.1.1, optional.
- HW1 is due Fri 10/02 on Gradescope; hard close 48 hours later.
- HW2 comes out Tue 10/06.