

Lecture 13: Function Approximation I: Features and Linear Values

CS 639: Introduction to Reinforcement Learning

Tengyang Xie

Thursday, October 15, 2026

Housekeeping

- HW2 is due Wed 10/21 on Gradescope; hard close 48 hours later.
 - Four slip days shared by HW1–HW4, at most two on one assignment; applied automatically.
- Midterm: Tue 10/27 in class, 75 minutes, Lectures 1–14; more details to come.
- Reading for today: Sutton and Barto, Sections 9.1–9.4 (Tsitsiklis and Van Roy 1997, 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.

Two destinations, by computer

mean VE, after n episodes =	10	20	50	100	500	2,000
TD(0) on the table, $\alpha = 0.1$	25.29	19.67	7.11	1.22	0.53	0.49
semi-gradient TD(0), the layer class, $\alpha = 0.1$	20.08	13.58	5.44	4.58	4.54	4.53
gradient Monte Carlo, the layer class, $\alpha = 0.1$	9.33	5.82	4.17	3.98	3.94	3.89
semi-gradient TD(0), the dead-end class, $\alpha = 0.1$	12.27	5.28	0.77	0.50	0.53	0.49
least squares on the returns, the layer class	5.97	4.47	3.56	3.26	3.04	3.01
the TD fixed point of the record, the layer class	6.20	5.13	4.45	4.21	4.04	4.01

The experiment: the prize tree (three steps, a fair coin at each, $\gamma = 1$; the prize 8 behind the rightmost last step), the coin rule run for 2,000 episodes; that is one run. Every number starts at 4. Each learner turns the first n episodes into a fit V_n ; the entry shown is the weighted error $VE(V_n) = \sum_s \mu(s) (V_n(s) - V^{\pi}(s))^2$, averaged over 200 independent runs (seed 639); the error at the start is 31. The layer class ties the bottom four states into one number; the dead-end class ties the four dead ends. Destinations: 0 for the table and the dead-end class, 3 and 4 for the layer class under Monte Carlo and TD.

Before you leave

- Reading: Sutton and Barto, Sections 9.1–9.4; Tsitsiklis and Van Roy (1997), optional.
- HW2 is due Wed 10/21 on Gradescope; hard close 48 hours later.
- Midterm Tue 10/27 in class, Lectures 1–14; more details to come.
- Lecture 14 (Tue 10/20): function approximation for control. Semi-gradient Q-learning, fitted Q-iteration, and when bootstrapping with features diverges.