Jianxun Zhou
zhou-jx23@mails.tsinghua.edu.cn

February 19, 2026
cs285 – Homework 2

1 MULTISTEP Q-LEARNING

TD-LEARNING BIAS

No.

TABULAR LEARNING

True:
1a3, 1b3, 2a3, 3a1, 3a3

VARIANCE OF Q ESTIMATE

Highest Variance: $N \to \infty$
Lowest Variance: $N = 1$

FUNCTION APPROXIMATION

CD

MULTISTEP IMPORTANCE SAMPLING

$$\phi_{k+1} \leftarrow argmin_{\phi \in \Phi} \sum_{j,t} \left(\frac{\pi(a_{j,t}|s_{j,t})}{\pi'(a_{j,t}|s_{j,t})} (y_{i,t} - Q_\phi(s_{j,t}, a_{j,t})) \right)^2$$

No change when $N = 1$.
When $N \to \infty$, 3b3 is right.