CS285 hw3

oliver song

October 2023

1

1.1

No. Because the $max()$ operation is non-linear

1.2

1.2.1

(a)13 (b)1

1.2.2

(a)13 (b)1

1.2.3

(a)123 (b)13

1.3

Highest variance: $N = 1$, lowest variance: $N \to \infty$

1.4

Statements 1 and 4 are true.

1.5

$$\phi_{k+1} \leftarrow arg\min_{\phi'} \sum_{j,t} (y_{j,t} - \frac{\pi_{\phi'(a_{j,t}|s_{j,t})}}{\pi_{\phi(a_{j,t}|s_{j,t})}} Q_\phi(s_{j,t}, a_{j,t}))^2$$

We do not need to change the update rule when $N = 1$, because action is determined as the $argmax$ of q-values, which is deterministic regardless of the policy.