Homework 1

10-403 Recitation 3

Homework 1

Clarifications

- Boltzman exploration:
  - $\mathbb{P}(A_t = a) = \pi_t(a) = \frac{\exp(Q_t(a)T)}{\sum_{a'} \exp(Q_t(a')T)}$
  - As $T \to 0$, behaviour converges to uniformly at random policy
  - As $T \to \infty$, behaviour converges to pure greedy policy
- ϵ-greedy exploration:
  - w.p. $1 - \epsilon : A_t = \arg\max_a Q_t(a)$
  - w.p. $\epsilon : A_t \sim \text{Uniform}(\{a_1, \ldots, a_N\})$

Homework 1

Clarifications

- Boltzman exploration:
  - $\mathbb{P}(A_t = a) = \pi_t(a) = \frac{\exp(Q_t(a)T)}{\sum_{a'} \exp(Q_t(a')T)}$
  - As $T \to 0$, behaviour converges to uniformly at random policy
  - As $T \to \infty$, behaviour converges to pure greedy policy
- ϵ-greedy exploration:
  - w.p. $1 - \epsilon : A_t = \arg\max_a Q_t(a)$
  - w.p. $\epsilon : A_t \sim \text{Uniform}(\{a_1, \ldots, a_N\} - \{\arg\max_a Q_t(a)\})$

Both forms or ϵ-greedy are acceptable
But previous slide is easier to implement

Homework 1

Clarifications

- “Fraction of population to keep at each iteration: 10%”
  - $\Longrightarrow$ Elite Size = 10

Homework 1

Reminder

- Please respond to Team Information poll on Piazza!