IEMS 490 Reinforcement Learning: Value and Policy Iteration

Yiming Peng

Department of IEMS

Northwestern University

October 28, 2018

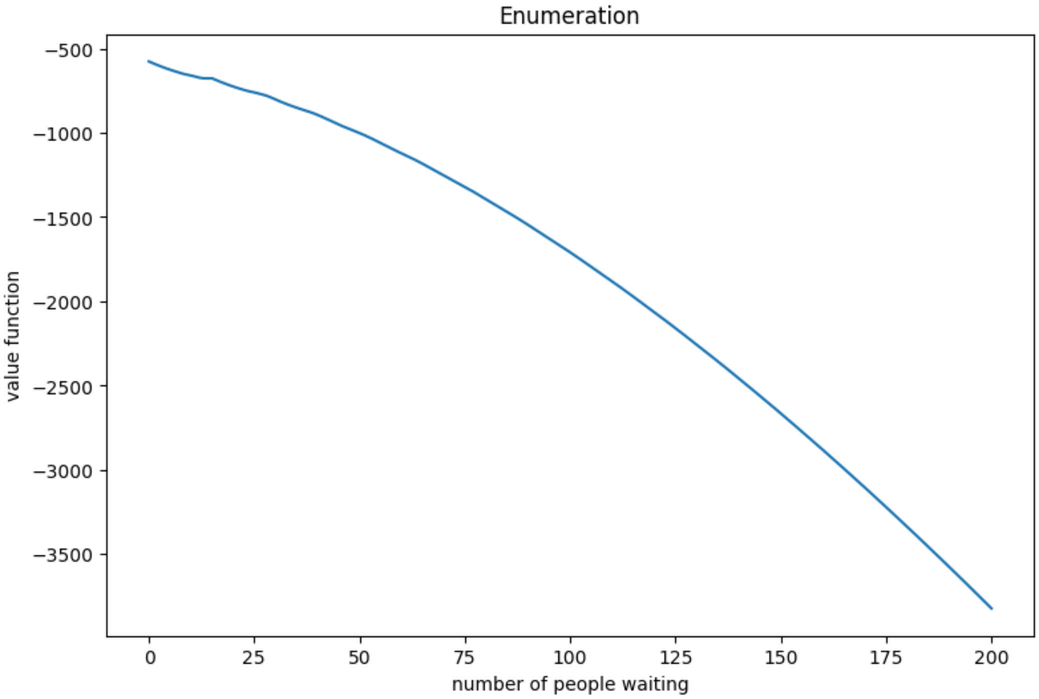
1. $\mathcal{S} = \{0, 1, ..., S\}$ where $S = 200$; $\mathcal{A} = \{0, 1\}$ where 1 stands for dispatching a shuttle and 0 for not dispatching.

   If $a_t = 1$, then $s_{t+1} = \begin{cases} s_t - K + A_t, & \text{if } s_t > K. \\ A_t, & \text{otherwise.} \end{cases}$, $r(s_t, 1) = \begin{cases} -(s_t - K)c_h - c_f, & \text{if } s_t > K. \\ -c_f, & \text{otherwise.} \end{cases}$

   If $a_t = 0$, then $s_{t+1} = \min(S, s_t + A_t)$, $r(s_t, 0) = -s_t c_h$.

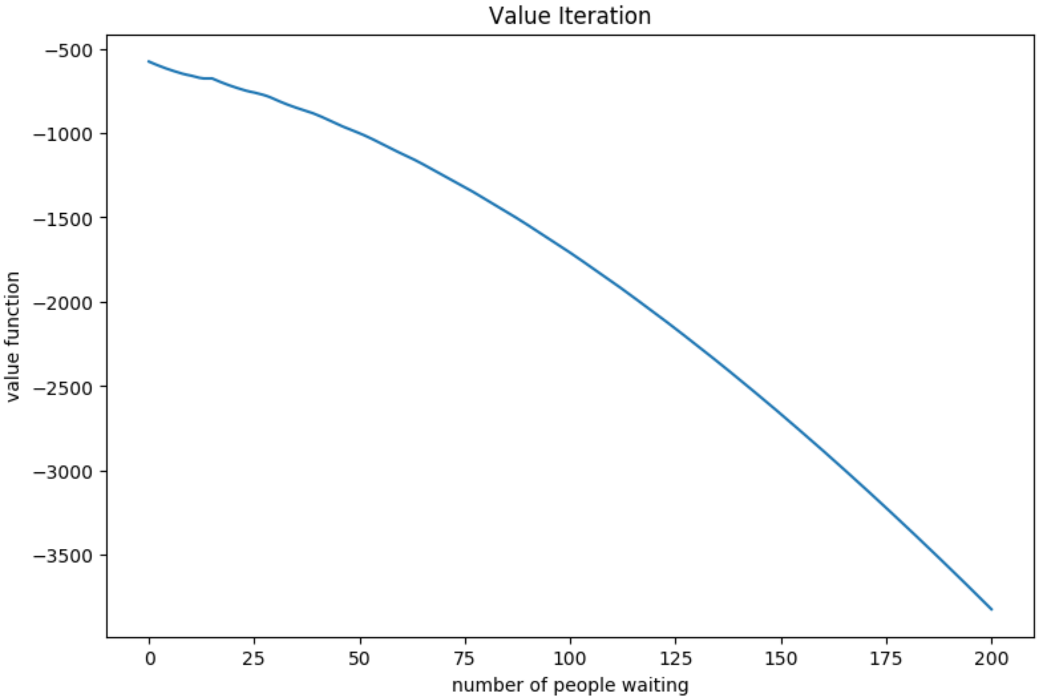
   (a) Enumeration (T=500)

   Assume that $V_{T+1}(s) = 0 \ \forall s \in \mathcal{S}$.



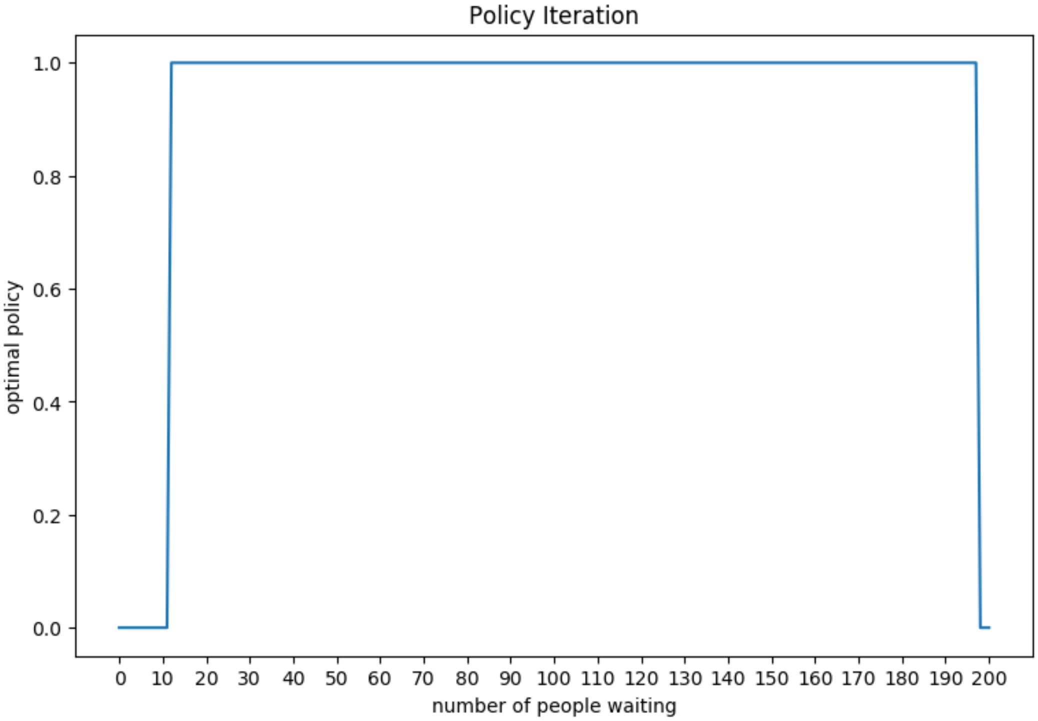
(b) Value Iteration

Initial $V_0(s) = 0 \ \forall s \in \mathcal{S}$.



(c) Policy Iteration

Initial $\pi_0(s) = 0 \ \forall s \in \mathcal{S}$.



2. In terms of modeling, the multiclass case is similar to the single class case as above. If we have K clasees, then $\mathcal{S} = \{0, 1, ..., S\}^K$ which, in this case ($K = 5, S = 100$), becomes too big for these tabualr algorithms. Therefore, we leave it to be solved by other algorithms to be covered later in the course.