CMPS 4660/6660 Reinforcement Learning – Fall 2020

Homework 2 (15 points)

Due 09/17/20 before class

Note: Please clearly justify your answer to each of the following questions. **Questions marked with *** are required for graduate students only.**

All the problems below assume a given Markov decision process $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$. V is the normed vector space of uniformally bounded value functions over the state space $\mathcal{S}$ with the infinity norm.

1. **Monotonicity of Bellman Operators (5 points)**

   Consider the Bellman operator $T^\pi : V \to V$ for a given stationary policy π, where $T^\pi v = r^\pi + \gamma P^\pi v$. Prove that for any $u, v \in V$, if $u \leq v$, then $T^\pi u \leq T^\pi v$. Recall that $u \leq v$ iff $u(s) \leq v(s)$ for all $s \in \mathcal{S}$.

2. ***** Policy Improvement (5 points)**

   Let π_0 be a stationary policy and π be the greedy policy with respect to v_{π_0}. That is, $\pi(s) = \text{argmax}_{a \in \mathcal{A}(s)}[r(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{ss'}(a) v_{\pi_0}(s')]$. Show that $v_\pi \geq v_{\pi_0}$.

3. **Policy Iteration for Action Values (5 points)**

   Give a complete policy iteration algorithm for computing q^*, analogous to that for computing v^* given in Section 4.3 of Sutton and Barto's book.