ECEN 689: RL: Fall 2018
Homework 2

---

1. (5 points) Let $\mathcal{X}$ be a given set, and $g_1 : \mathcal{X} \to \mathbb{R}$ and $g_2 : \mathcal{X} \to \mathbb{R}$ be two real-valued functions on $\mathcal{X}$. Also assume that both function are bounded. Then, show that

$$|\max_x g_1(x) - \max_x g_2(x)| \leq \max_x |g_1(x) - g_2(x)|$$

2. Consider a finite MDP $(\mathcal{X}, \mathcal{A}, P, R, \gamma)$. Assume that $\max_x \max_a R(x, a) = C$. Let V^* be the optimal value function. Also, $||\cdot||$ indicates the sup norm.

   (a) (2 points) Give an upper bound and lower bound for the optimal value function V^*.

   (b) (4 points) Consider the Value Iteration algorithm $V_{k+1} = TV_k$, where T is the Bellman operator. Show that for any given $\epsilon > 0$, there exists a finite integer k_0 such that $||V_k - V^*|| \leq \epsilon$ for all $k \geq k_0$. Also give a lower bound for k_0.

   (c) (4 points) Consider the Value Iteration algorithm $V_{k+1} = TV_k$, where T is the Bellman operator. Let m be an integer such that

$$||V_{m+1} - V_m|| \leq \frac{\epsilon(1-\gamma)}{2\gamma}$$

   Then, show that

$$||V_{m+1} - V^*|| \leq \epsilon/2.$$

   Discuss how this result can be used as a stopping criteria for the Value Iteration algorithm

3. (8 points) Suppose you terminate the Value Iteration algorithm after n steps and let $\hat{V} = V_n = T^n V_0$ be the corresponding approximate value function. Assume that $||\hat{V} - V^*|| < \epsilon$. Now, compute the approximate policy $\hat{\pi}$ as

$$\hat{\pi}(x) = \arg\max_a (R(x, a) + \gamma \sum_{y \in \mathcal{X}} P(y|x, a)\hat{V}(y))$$

   Let $V_{\hat{\pi}}$ be the value corresponding to the policy $\hat{\pi}$. Show that

$$||V_{\hat{\pi}} - V^*|| \leq \frac{2\gamma\epsilon}{(1-\gamma)}$$

4. (17 points) **Frozen Lake MDP**: Implement value iteration and policy iteration for the Frozen Lake environment from OpenAI Gym (Link Here). Use discount factor $\gamma = 0.9$.

   To be submitted: IPython file which shows: (i) value iteration and policy iteration convergence plots, (ii) optimal value function marked on the grid, (iii) optimal policy marked on the grid