Tsinghua-Berkeley Shenzhen Institute
Learning From Data
Fall 2020

Writing Assignment 5

Issued: Tuesday 15th December, 2020 **Due:** Wednesday 30th December, 2020

POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students **NOT** to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect to not to google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.
- **Required homework submission format:** You can submit homework either as one single PDF document or as handwritten papers. Written homework needs to be provided during the class in the due date, and PDF document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of due date.

  It is encouraged you LaTeX all your work, and we would provide a LaTeX template for your homework.
- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your own, **PLEASE STATE THIS**. Each student must understand, write, and hand in answers of their own.

5.1. (2 points) (Bellman's equation) In a dynamic decision problem, given a policy π, the value function satisfies the Bellman equation:

$$V^{\pi}(s) = R(s) + \gamma \sum_{s' \in S} P_{s\pi(a)}(s') V^{\pi}(s') \tag{1}$$

Now we play a simple game in a 3x3 block square. Our goal is to move the red object from the upper left $(0, 0)$ to the bottom right corner $(2, 2)$ (See Figure 1). The state s is represented by a tuple (x, y) where $x, y \in \{0, 1, 2\}$. Choosing $\gamma = 0.8$. The reward matrix satisfies $R((2, 2)) = 1$ and $R(s) = 0$ for other state s. There are four actions possible for each state $\mathcal{A} = \{\mathsf{up}, \mathsf{down}, \mathsf{left}, \mathsf{right}\}$, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the

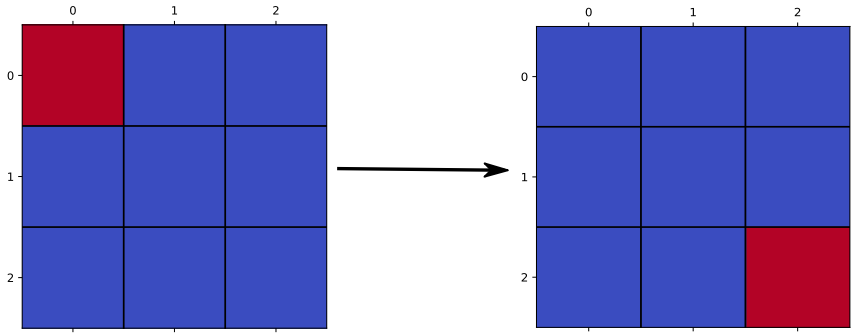

Figure 1: Moving a red object from upper left to bottom right

grid in fact leave the state unchanged. For example $P_{s=(1,1),a=\text{right}}(s' = (1,2)) = 1$ and $P_{s=(1,1),a=\text{right}}(s' = (0,2)) = 0$. Suppose π is a policy defined by

$$\pi((i,j)) = \begin{cases} (i+1,j) & i < 2 \\ (i,j+1) & i = 2, j < 2 \\ (2,2) & i = 2 \text{ and } j = 2 \end{cases}$$

Compute numerically the value function $V^\pi(s)$ for each s by solving the Bellman's equation.

5.2. (2 points) (Convergence of Value Iteration) You have learned in class value iteration algorithm updates the value function $V^{t+1}(s) = BV^t(s)$ for every state s, where B is the Bellman backup operator:

$$B\,V(s) = R(s) + \max_{a \in A} \gamma \sum_{s' \in S} P_{sa}(s')V(s') \tag{3}$$

(a) (1 point) Show that Bellman backup operator is a contraction operator. That is, for any value function V_1, V_2,

$$\max_{s \in S} |B\,V_1(s) - B\,V_2(s)| \leq \gamma \max_{s \in S} |V_1(s) - V_2(s)| \tag{4}$$

(b) (1 point) Assuming $R_{\max} = \max_{s \in S} R(s)$ and $V^0(s) = 0$ for all $s \in S$, show that

$$\max_{s \in S} |V^t(s) - V^*(s)| \leq \frac{\gamma^t R_{\max}}{1-\gamma} \tag{5}$$

From (5), we can see that $V^t(s)$ converges to $V^*(s)$.

5.3. (3 points) (Mean Square Error) We mentioned Bias-Variance Tradeoff in class. We define the MSE of $\hat{X}$, an estimator of X as $\text{MSE}(\hat{X}) \triangleq \mathbb{E}[(\hat{X}-X)^2]$. The variance of $\hat{X}$ is defined as $\text{Var}(\hat{X}) \triangleq \mathbb{E}[(\hat{X}-\mathbb{E}[\hat{X}])^2]$ and the bias is defined as $\text{Bias}(\hat{X}) \triangleq \mathbb{E}[\hat{X}]-X$.

(a) (1 point) Please prove that

$$\text{MSE}(\hat{X}) = \text{Var}(\hat{X}) + (\text{Bias}(\hat{X}))^2$$

(b) (2 points) Our data are added with an independent Gaussian noise, say, $X+N$, where $\mathbb{E}[N]=0$ and $\mathbb{E}[N^2]=\sigma^2$ and the estimator is $\hat{X}$. We define the empirical MSE as $\mathbb{E}[(\hat{X}-X-N)^2]$. Please prove that

$$\mathbb{E}[(\hat{X}-X-N)^2] = \text{MSE}(\hat{X}) + \sigma^2$$

The equation tells us that the empirical error is a good estimation of the true error. Thus, we can minimize the empirical error in order to properly minimize the true error.

5.4. (3 points) Important inequalities in Learning Theory.

(a) (1.5 points) (Markov's Inequality) Let X be a non-negative random variable, then for every positive constant a, please show that

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

(b) (1.5 points) (Chebyshev's inequality)For random variable X, if its expected value $\mathbb{E}(X)$ and variance $Var(X)$ are both finite, for every positive constant a, please show that

$$P(|X-\mathbb{E}(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

5.5. (3 points) (Bonus question: VC Dimension) Given some finite domain set, $\mathcal{X}$, and a number $k \leq |\mathcal{X}|$,

please figure out the VC-dimension of each of the following classes:

(a) (1.5 points) $\mathcal{H}_k^{\mathcal{X}} = \left\{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x)=1\}| = k\right\}$. That is, the set of all functions that assign the value 1 to exactly k elements of $\mathcal{X}$.

(b) (1.5 points) $\mathcal{H}_{\leq k}^{\mathcal{X}} = \left\{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x)=1\}| \leq k \text{ or } |\{x : h(x)=0\}| \leq k\right\}$