Homework #14

Junwu Zhang
CME 241: Reinforcement Learning for Finance

Februray 26, 2020

Problem 1.
Write with proper notation, the derivations to solutions of Linear Systems for Bellman Error-minimization and Projected Bellman Error-minimization

Solution.
In this problem, we are discussing Value Function Geometry. First, let's list down the important components for this problem with proper mathematical notations:

- *States:* $\mathcal{S} : \{s_1, s_2, \cdots, s_n\}$($n$ states)
- *Actions:* $\mathcal{A} : \{a_1, a_2, \cdots, a_n\}$(finite amount of actions)
- *Policy:* fixed (often stochastic) policy denoted $\pi(a|s)$
- *Feature functions:* there are m feature functions $\phi_1, \phi_2, \cdots, \phi_m : \mathcal{S} \to \mathbb{R}$
- *Value Function (VF):* for given policy π, $\mathbf{v}_\pi : \mathcal{S} \to \mathbb{R}$
- *Weights:* $\mathbf{w} = (w_1, w_2, \ldots, w_m)$
- The corresponding VF with linear function approximation can be defined as:

$$\mathbf{v}_{\mathbf{w}}(s) = \mathbf{w}^T \cdot \phi(s) = \sum_{j=1}^{m} w_j \cdot \phi_j(s), \forall s \in \mathcal{S} \tag{1}$$

- *Probability distribution:* $\mu_\pi : \mathcal{S} \to [0, 1]$
- *Expected Reward:* $r(s, a)$

$$\mathbf{R}_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \cdot r(s, a) \tag{2}$$

  where $\mathbf{R}_\pi(s)$ is the vector $\left[\mathbf{R}_\pi(s_1), \mathbf{R}_\pi(s_2), \ldots, \mathbf{R}_\pi(s_n)\right]$
- *Transition probability:* $p(s, s', a)$ is the probabilty that state s transition to state s' given action a

$$\mathbf{P}_\pi\left(s, s'\right) = \sum_{a \in \mathcal{A}} \pi(a|s) \cdot p\left(s, s', a\right) \tag{3}$$

  where $\mathbf{P}_\pi$ is the matrix $\left[\mathbf{P}_\pi\left(s_i, s_{i'}\right)\right], 1 \leq i, i' \leq n$
- γ is the Markov Decision Process (MDP) discount factor

Given these notations, we know a few more things:

- Bellman operator $\mathbf{B}_\pi$ on $\mathbf{v}$ given policy π can be defined as:

$$\mathbf{B}_\pi \mathbf{v} = \mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{v} \tag{4}$$

- Subspace of VF can be denoted as $\mathbf{\Phi}$, and the projection operator that does orthogonal project of VF onto $\mathbf{\Phi}$ can be written as:

$$\mathbf{\Pi}_{\mathbf{\Phi}} = \mathbf{\Phi} \cdot \left(\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{\Phi}\right)^{-1} \cdot \mathbf{\Phi}^T \cdot \mathbf{D} \tag{5}$$

- The *distance* between VF vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is $d(\mathbf{v}_1, \mathbf{v}_2)$. Weighted by μ_π across n dimensions, the distance can be written as:

$$d\left(\mathbf{v}_1, \mathbf{v}_2\right) = \sum_{i=1}^{n} \mu_\pi\left(s_i\right) \cdot \left(\mathbf{v}_1\left(s_i\right) - \mathbf{v}_2\left(s_i\right)\right)^2 = \left(\mathbf{v}_1 - \mathbf{v}_2\right)^T \cdot \mathbf{D} \cdot \left(\mathbf{v}_1 - \mathbf{v}_2\right) \tag{6}$$

  where $\mathbf{D}$ is the square diagonal matrix.

Therefore we can solve for Linear Systems for Bellman Error-minimization as:

$$\begin{aligned}
\mathbf{w}_{BE} &= \arg\min_{\mathbf{w}} d\left(\mathbf{v}_{\mathbf{w}}, \mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{v}_{\mathbf{w}}\right) \\
&= \arg\min_{\mathbf{w}} d\left(\mathbf{\Phi} \cdot \mathbf{w}, \mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi} \cdot \mathbf{w}\right) \\
&= \arg\min_{\mathbf{w}} d\left(\mathbf{\Phi} \cdot \mathbf{w} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi} \cdot \mathbf{w}, \mathbf{R}_\pi\right) \\
&= \arg\min_{\mathbf{w}} d\left(\left(\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi}\right) \cdot \mathbf{w}, \mathbf{R}_\pi\right)
\end{aligned} \tag{7}$$

With this, we can solve for a weighted least-square linear regression and get the solution as:

$$w_{BE} = \left(\left(\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi}\right)^T \cdot \mathbf{D} \cdot \left(\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi}\right)\right)^{-1} \cdot \left(\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi}\right)^T \cdot \mathbf{D} \cdot \mathbf{R}_\pi \tag{8}$$

We can also solve for Linear Systems for Projected Bellman Error-minimization. Writing out the formulation for $\mathbf{\Pi}_\Phi$ and $\mathbf{B}_\pi$, we have:

$$\mathbf{\Pi}_{\mathbf{\Phi}} = \mathbf{\Phi} \cdot \left(\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{\Phi}\right)^{-1} \cdot \mathbf{\Phi}^T \cdot \mathbf{D} \tag{9}$$

$$\mathbf{B}_\pi \mathbf{v} = \mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{v} \tag{10}$$

Since $\mathbf{\Phi} \cdot \mathbf{w}_{PBE}$ is the fixed point of operator $\mathbf{\Pi}_\Phi \cdot \mathbf{B}_\pi$, combining the two equations above, we can write:

$$\mathbf{\Phi} \cdot \left(\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{\Phi}\right)^{-1} \cdot \mathbf{\Phi}^T \cdot \mathbf{D} \cdot \left(\mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi} \cdot \mathbf{w}_{PBE}\right) = \mathbf{\Phi} \cdot \mathbf{w}_{PBE} \tag{11}$$

If columns of Φ are assumed to be independent, we can expand the above equations to be:

$$\begin{aligned}
\left(\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{\Phi}\right)^{-1} \cdot \mathbf{\Phi}^T \cdot \mathbf{D} \cdot \left(\mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi} \cdot \mathbf{w}_{PBE}\right) &= \mathbf{w}_{PBE} \\
\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \left(\mathbf{R}_\pi + \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi} \cdot \mathbf{w}_{PBE}\right) &= \mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{\Phi} \cdot \mathbf{w}_{PBE} \\
\mathbf{\Phi}^T \cdot \mathbf{D} \cdot \left(\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi}\right) \cdot \mathbf{w}_{PBE} &= \mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{R}_\pi
\end{aligned} \tag{12}$$

We can see that this is a square linear system with the form $\mathbf{A} \cdot \mathbf{w}_{PBE} = \mathbf{b}$, and the solution to equations with this form is:

$$\mathbf{w}_{PBE} = \mathbf{A}^{-1} \cdot \mathbf{b} = \left(\mathbf{\Phi}^T \cdot \mathbf{D} \cdot (\mathbf{\Phi} - \gamma \mathbf{P}_\pi \cdot \mathbf{\Phi})\right)^{-1} \cdot \mathbf{\Phi}^T \cdot \mathbf{D} \cdot \mathbf{R}_\pi \tag{13}$$

■