Assignment 8 by Haobin Tang

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

Bellman Error (BE)-minimizing:

Vector $v_w = \Phi \cdot w$ in this subspace has coordinates $[v_w(s_1), v_w(s_2), \ldots, v_w(s_n)]$

Bellman Error (BE)-minimizing: $w_{BE} = argmin_w d(B_\pi v_w, v_w)$

1. This can be expressed as the solution of a linear system $Aw = b$
2. Matrix A and Vector b comprises of $P_\pi, R_\pi, \Phi, \mu_\pi$
3. In model-free setting, A and b can be estimated with batch data
4. Based on observation: $w_{BE} = argmin_w (\mathbb{E}\pi[\delta])^2$, where δ is TD Error
5. Cannot learn if we can only access features, and not underlying states

$$\begin{aligned} w_{BE} &= argmin_w d(B_\pi v_w, v_w) \\ &= argmin_w d(R_\pi, \Phi w - \gamma P_\pi \Phi w) \\ &= argmin_w d(R_\pi, (\Phi - \gamma P_\pi \Phi) w) \end{aligned}$$

This is a weighted least-squares linear regression of R_π versus $\Phi - \gamma P_\pi \Phi) w$ with weights $\mu \pi$, whose solution is:

$$w_{BE} = ((\Phi - \gamma P_\pi \Phi)^T D (\Phi - \gamma P_\pi \Phi))^{-1} (\Phi - \gamma P_\pi \Phi)^T D R_\pi$$

Projected Bellman Error (PBE)-minimizing: $w_{PBE} = argmin_w d((\Pi_\Phi \cdot B_\pi) v_w, v_w)$

1. The minimum is 0, i.e., $\Phi \cdot w_{PBE}$ is the fixed point of operator $\Pi_\Phi \cdot B_\pi$
2. Starting with an arbitrary VF vector v and repeatedly applying B_π (taking it out of the subspace) followed by Π_Φ (projecting it back to the subspace), we will reach the fixed point $\Phi \cdot w_{PBE}$
3. Also,w_{PBE} can be expressed as the solution of a linear system $Aw = b$
4. In model-free setting, A and b can be estimated with batch data

$$\Pi_\Phi = \Phi \cdot (\Phi^T \cdot D \cdot \Phi)^{-1} \cdot \Phi^T \cdot D$$
$$B_\pi v = R_\pi + \gamma P_\pi \cdot v$$

Therefore

$$\Phi \cdot (\Phi^T \cdot D \cdot \Phi)^{-1} \cdot \Phi^T \cdot D \cdot (R_\pi + \gamma P_\pi \cdot \Phi \cdot w_{PBE}) = \Phi \cdot w_{PBE}$$
$$\Phi \cdot w_{PBE} = (\Phi^T \cdot D \cdot (\Phi - \gamma P_\pi \Phi))^{-1} \Phi^T \cdot D \cdot R_\pi$$