

Linear Systems Bellman Error Minimization

The motivation behind this derivation is to investigate how we might best be able to solve an RL problem using linear value function approximation. In a general setting, we know we have converged at the true value function when we reach the stationary point of the Bellman operator. At this point, updating our value function with new experiences provides no change to the value function. However, with function approximation, we are of course "approximating" the true value function, and thus are unlikely to arrive at the perfectly defined corollary to the tabular setting. Instead, we'd like to investigate different interpretations of defining/minimizing the Bellman Error.

We first try to minimize the distance between the value function and the Bellman updated value function.

$$w_{BE} = \operatorname{argmin}_w d(v_w, R_\pi + \gamma P_\pi v_w)$$

We can rewrite our value function in terms of our linear approximation.

$$= \operatorname{argmin}_w d(\phi \cdot w, R_\pi + \gamma P_\pi \cdot \phi \cdot w)$$

Since the RHS is a sum of vectors, we can subtract one component to both elements of the distance function in order to rearrange terms and then re-factor

$$\begin{aligned} &= \operatorname{argmin}_w d(\phi \cdot w - \gamma P_\pi \cdot \phi \cdot w, R_\pi) \\ &= \operatorname{argmin}_w d((\phi - \gamma P_\pi \cdot \phi) \cdot w, R_\pi) \end{aligned}$$

Notice that R_π is a vector of scalars, as are ϕ , P_π (matrix), and γ (single value). The only variable we are minimizing over is w . Thus, this is equivalent to fitting a set of features $(\phi - \gamma P_\pi \cdot \phi) = X$ via a parameter w to fit a target $R_\pi = Y$. This can be solved with the common least squares formula. Recall that our definition of distance is weighted by μ_π for each state, requiring us to weight the regression by diagonal matrix of weights D .

$$\begin{aligned} w_{BE} &= (X^T X)^{-1} X^T Y \\ &= ((\phi - \gamma P_\pi \cdot \phi)^T D (\phi - \gamma P_\pi \cdot \phi))^{-1} \cdot (\phi - \gamma P_\pi \cdot \phi)^T \cdot D \cdot R_\pi \end{aligned}$$

Another approach is to minimize the distance between the value function approximation and the projected Bellman operator applied to the value function. Visually, the projected Bellman operator applies the Bellman operator to the value function, then projects it back into the subspace spanned by the approximations feature space. This differs from before by this notion of projection - it keeps us in the space of the function approximation while still applying the Bellman update. First we define the projection operator

$$\Pi_\phi = \phi \cdot (\phi^T \cdot D \cdot \phi)^{-1} \cdot \phi^T \cdot D$$

This is simply the projection onto the weighted subspace of ϕ (derived from the least squares solution minimizing the distance from a point on subspace to v). Recall we want to find the point where applying the projected Bellman update yields no difference in our value function (stationary point).

$$\begin{aligned}
 \phi w_{PBE} &= \Pi_{\phi} \cdot B\pi \cdot \phi \cdot w_{PBE} \\
 \phi w_{PBE} &= \phi \cdot (\phi^T \cdot D \cdot \phi)^{-1} \cdot \phi^T \cdot D \cdot (R_{\pi} + \gamma P_{\pi} \cdot \phi \cdot w_{PBE}) \\
 w_{PBE} &= (\phi^T \cdot D \cdot \phi)^{-1} \cdot \phi^T \cdot D \cdot (R_{\pi} + \gamma P_{\pi} \cdot \phi \cdot w_{PBE}) \\
 (\phi^T \cdot D \cdot \phi) \cdot w_{PBE} &= \phi^T \cdot D \cdot (R_{\pi} + \gamma P_{\pi} \cdot \phi \cdot w_{PBE}) \\
 (\phi^T \cdot D \cdot \phi) \cdot w_{PBE} - \phi^T \cdot D \cdot \gamma P_{\pi} \cdot \phi \cdot w_{PBE} &= \phi^T \cdot D \cdot R_{\pi} \\
 \phi^T \cdot D \cdot (\phi - \gamma P_{\pi} \cdot \phi) \cdot w_{PBE} &= \phi^T \cdot D \cdot R_{\pi}
 \end{aligned}$$

First note that we are able to invert ϕ because we assume it's columns to be linearly independent. Notice the last line from above is simply a linear system. $\phi^T \cdot D \cdot (\phi - \gamma P_{\pi} \cdot \phi)$ is a scalar matrix, $\phi^T \cdot D \cdot R_{\pi}$ is a scalar vector, and we are trying to solve for w . Thus,

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

This provides a direct solution to minimizing the projected Bellman error for a linear function approximation.