CME241 Notes

Saakaar Bhatnagar

Value Function Geometry

1 Aim

- This section is useful in understanding the effects of the different methods of updating w to arrive at approximations to the true value function for a prediction problem
- The approximations are linear spans using columns vectors of the dimension of the state space

2 Finding W_{BE} Value

2.1 Definitions and interpretation

Mathematically, we have:

$$W_{BE} = argmin_w [E_\pi[\delta]]^2$$

or

$$W_{BE} = argmin_w [E_\pi[B_\pi V_w - V_w]]^2$$

where

$$B_\pi = R_\pi + \gamma P_\pi$$

Graphically, it refers to finding the constants that span the feature function space such that the length of the vector connecting the original V_w and the $B_\pi V_w$ that is knocked out of the span of Φ, is minimal. It is not, however the same as w_π

Because of the expectation involved, we have to weight by the likelihood of seeing the state in the MRP. This can be calculated from the frequency of visitation of each state of the MRP.

2.2 Solution

$$W_{BE} = argmin_w d(B_\pi V_w, V_w)$$

$$= argmin_w d(R_\pi, (\phi - \gamma P_\pi \phi).w)$$

$$= argmin_w [(R_\pi + \gamma P_\pi \phi w - \phi w)^T D(R_\pi + \gamma P_\pi \phi w - \phi w)]$$

We can find the solution to the above by takng derivative w.r.t w, since it is quadratic in w:

$$(\gamma P_\pi \phi - \phi)^T D(R_\pi + \gamma P_\pi \phi w - \phi w) + (R_\pi + \gamma P_\pi \phi w - \phi w)^T D(\gamma P_\pi \phi - \phi) = 0$$

The above expression is symmetric, using that property we can expand and write:

$$(\gamma P_\pi \phi - \phi)^T D R_\pi + (\gamma P_\pi \phi - \phi)^T D(\gamma P_\pi \phi - \phi).w + (R_\pi)^T D(\gamma P_\pi \phi - \phi) + (\gamma P_\pi \phi w - \phi w)^T D(\gamma P_\pi \phi - \phi) = 0$$

Rearranging:

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

3 Finding W_{PBE}

3.1 Definitions and interpretation

Mathematically, we have:

$$W_{PBE} = argmin_w d(\Pi_\phi B_\pi V_w, V_w)$$

where

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

The graphical interpretation of the projection operator, is that it projects an arbitrary vector v onto the plane defined as the linear span of ϕ.

Therefore, as we solve the above, the value of the minima will be 0, since at W_{PBE} the bellman operator gives a vertical projection of V_w, and the projection operator projects it right back to the same point, making the distance 0.

Mathematically,

$$\phi w_{PBE} = \Pi B_\pi \phi w_{PBE}$$

3.2 Solution

Using the above statement, and noting that ϕ is a matrix of full rank(the feature functions are linearly independent),

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

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

Rearranging,

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

4 Notes

- It should be noted that the above solutions are obtainable when the MRP is fully known.
- If the MRP is not known, then we will have to resort to sampling and approximate methods. These often reduce to the methods we save in RL and value function approximation.