Carolyn Kao (chkao831@stanford.edu)

CME 241, Winter 2021

HW14

1.LSTD Idea of using LSTD for approximate policy evaluation in PI

```
Start with random weights w (i.e.  value funcUon)
Repeat until Convergence
π(s) = greedy(Φw)
Evaluate π using LSTD
```

- `Generate sample trajectories of` P_π
- `Use LSTD to produce new weights w (w gives an approximated value function of` π)

Linear model approximation $\boldsymbol{W} = \left(I - \gamma \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Phi}'\right)^{-1} \left(\Phi^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T R \quad = \left(\Phi^T \Phi - \gamma \Phi^T \Phi'\right)^{-1} \Phi^T R$ aligns with LSTD solution $\boldsymbol{W} = \left(\boldsymbol{\Phi}^T \Phi - \gamma \Phi^T \Phi'\right)^{-1} \Phi^T R$.

2. LSPI LSPI is similar to previous loop by replaces LSTD with a new algorithm LSTDQ