Reinforcement Learning
Week 6 – Wednesday

Error Projected Bellman

① doesn't involve true value, thus more stable than $\mathcal{E}_{ms}$.

Scalar $V(s) = r(s) - g + \mathbb{E}[V(s')]$

↳ with deterministic policy

matrix (vector) $\vec{V} = \vec{r} - g\mathbb{1} + P \cdot \vec{V}$

one-step state distribution

size: $|S|$ by $|S|$

$\vec{V} = \mathbb{B}[\vec{V}]$

↳ Bellman (evaluator) operation

$$e_{PB} \overset{\text{def}}{=} \| \hat{v}(w) - \mathbb{P}\,\mathbb{B}\,\hat{v}(w) \|^2_{p^*}$$

↳ norm vector

$$= (\Delta v)^T D_{p^*} \Delta v$$

where

$\cancel{\Delta v = \hat{v}(w) - \mathbb{P}\mathbb{B}\hat{v}_w}$

$\Delta v = \hat{v}(w) - PB\hat{v}_w$

$$= \sum_{s \in S} p^*(s) \cdot (\Delta v(s))^2$$

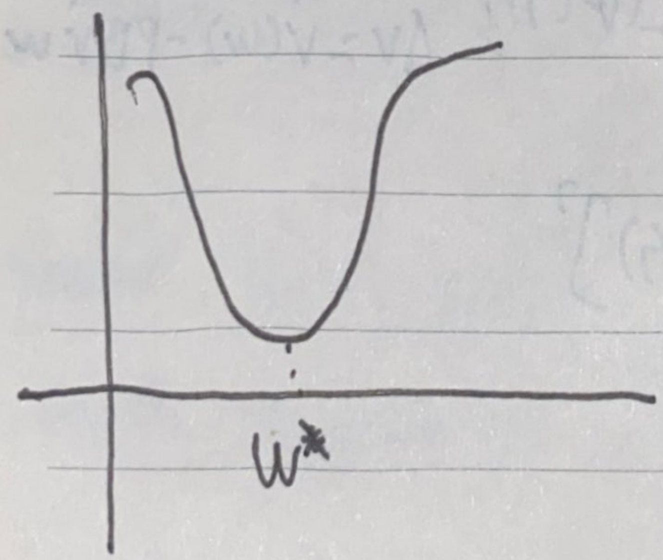
$$= \mathbb{E}_{p^*}[\Delta v(s)]^2$$

Exceed expectations, derive $\mathbb{P}$!

How to get w^*

① take the $\underbrace{\text{gradient of } \mathcal{E}_{PB} \quad w}_{\nabla \mathcal{E}_{PB} \text{ of } (w)}$

② set the grad to zero because in Local minima, gradient is 0



- Then we can show that

$w^* = X^{-1} y \rightarrow X =$

$$\sum_{s} p^*(s) \sum_{s'} P(s'|s) \left[f(s) \{ f(s) - f(s') \}^T \right]$$

$$= F^T D_{p^*} (I - P) F$$

$$y = \sum_{s} p^*(s) \left[(r(s) - g\mathbb{1}) \; f(s) \right]$$

$$= F^T D_{p^*} (r - g\mathbb{1})$$

LSTD (Least Square Temporal Diff)

↳ predict $\hat{w}^*$

↳ = $\boxed{\hat{X}^{-1} \hat{y}}$

$$\hat{X} = \frac{1}{n} \overset{\sum_{i=1}^{n}}{F(s_i)} \left[f(s_i) - F(s_i') \right]^T$$

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} (r(s_i) - g\mathbb{1}) . f(s_i)$$

$$\boxed{\hat{w}^* = \hat{X}^{-1} . \hat{y}}$$