Hw4

14 de março de 2024 16:57

a) $\dot{J} = \mathbb{E}_\pi \left[c_t + (\gamma - 1) J \right]$

assuming $J = J^\pi$

$= \mathbb{E}_\pi \left[c_t + (\gamma - 1) J^\pi \right]$

from the first equation

$= \mathbb{E}_\pi \left[c_t + (\gamma - 1) \dfrac{c_\pi}{1-\gamma} \right]$

$= \mathbb{E}_\pi \left[c_t - c_\pi \right]$

$= \mathbb{E}_\pi \left[c_t \right] - c_\pi$, valor esperado do custo segundo π é c_π

$= c_\pi - c_\pi = 0$ //

b) $\dot{E}_t = \dfrac{d}{dt} E_t = \dfrac{1}{2} \times 2 \times (J^{(t)} - J^\pi) \times \dfrac{d}{dt}(J^{(t)} - \underbrace{J^\pi}_{\text{constante no tempo}})$

$= (J^{(t)} - J^\pi)(\dot{J} - 0) = \dot{J}\,(J^{(t)} - J^\pi)$

$J^\pi = \dfrac{c_\pi}{1-\gamma} \iff c_\pi = (1-\gamma) J^\pi$

$\dot{E}_t = \mathbb{E}_\pi \left[c_t + (\gamma-1) J^{(t)} \right] (J^{(t)} - J^\pi) = (c_\pi + (\gamma - 1) J^{(t)})(J^{(t)} - J^\pi) = \left((1-\gamma) J^\pi + (\gamma-1) J^{(t)}\right)(J^{(t)} - J^\pi) = (1-\gamma)(J^\pi - J^{(t)})(J^{(t)} - J^\pi) = -(1-\gamma)(J^\pi - J^{(t)})^2$

$0 \leq \gamma \leq 1$, logo $(1-\gamma) \geq 0$

$\dot{E}_t \leq 0$ //

c) No exercício anterior, concluímos que $\dot{E}_t \leq 0$, o que significa que E_t é decrescente.
Sendo $E_t = \frac{1}{2}(J^{(t)} - J^\pi)^2$, E_t pode ser vista como o erro entre o cost-to-go obtido e o da política π.
Como o $\dot{E}_t$ é calculado utilizando TD(0), concluímos que este tem ótimos resultados, convergindo para o valor de cost-to-go ótimo, mesmo sem considerar qualquer observação, ao contrário dos restantes TD(λ).