Introducción al uso de MDP pero para Estados Continuos

Hasta ahora: $MDP(S, A, P_{sa}, \gamma, R)$

$$V^{\pi}(s) = E[R(s_0) + \gamma R(s_1) \dots]$$

$$V^{*}(s) = \max_a V^{\pi}(s)$$

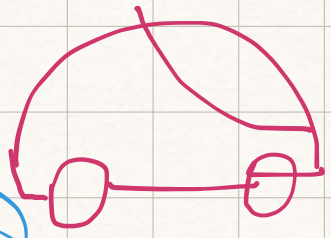
$$\pi^{*}(s)$$

$$\pi^{*} = \arg\max_a \sum_{s'} P_{sa}(s') V^{*}(s')$$

$$V(s) = R(s) + \max_a \gamma \sum_{s'} P_{sa}(s') V(s')$$

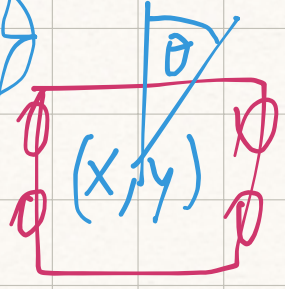
$$V(s) = V(s') \Rightarrow V^{*}(s)$$

Ejemplos: AA



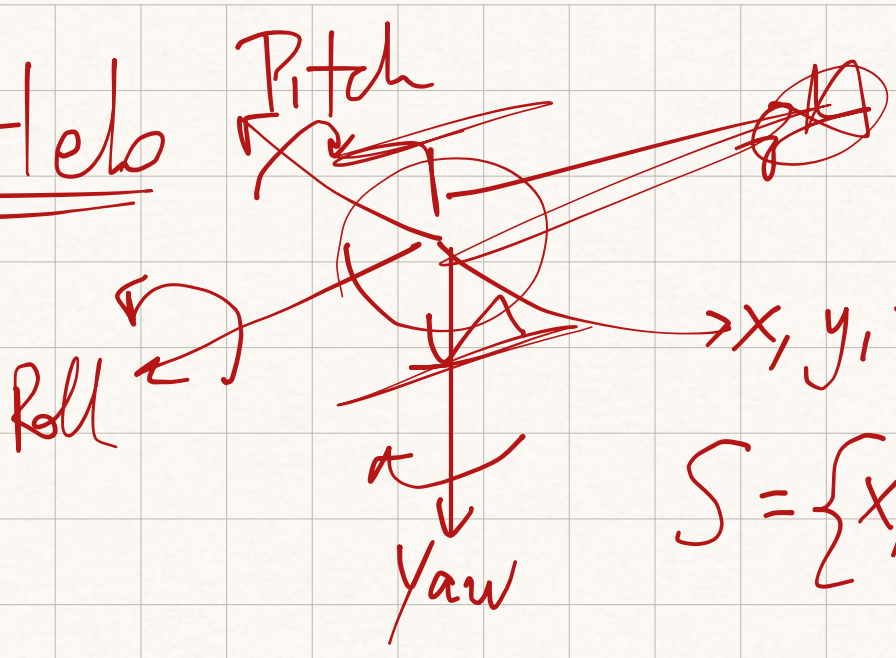
$$S = \{x, y, \theta, \dot{x}, \dot{y}, \dot{\theta}\}$$

6-dim



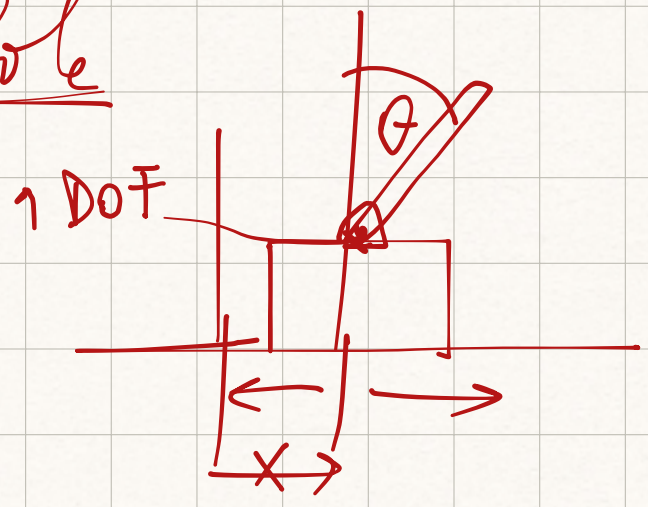
$$S = (3, 1)$$

Helo



$$S = \{x, y, z, \phi, \theta, \psi\}$$

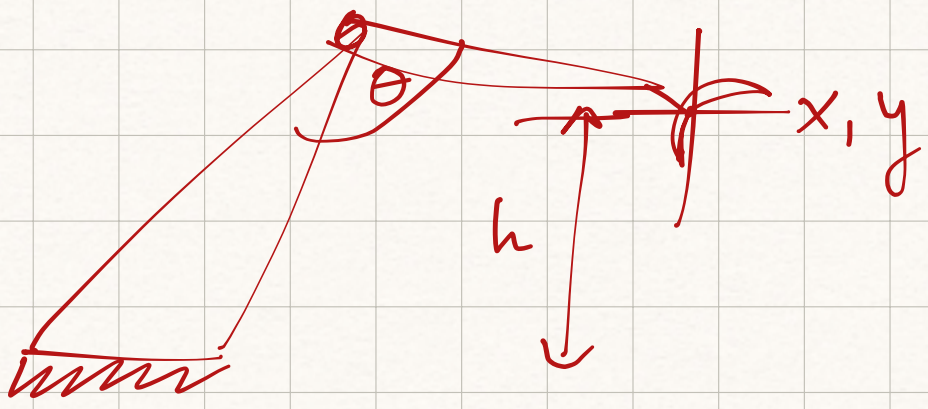
Can-Pole

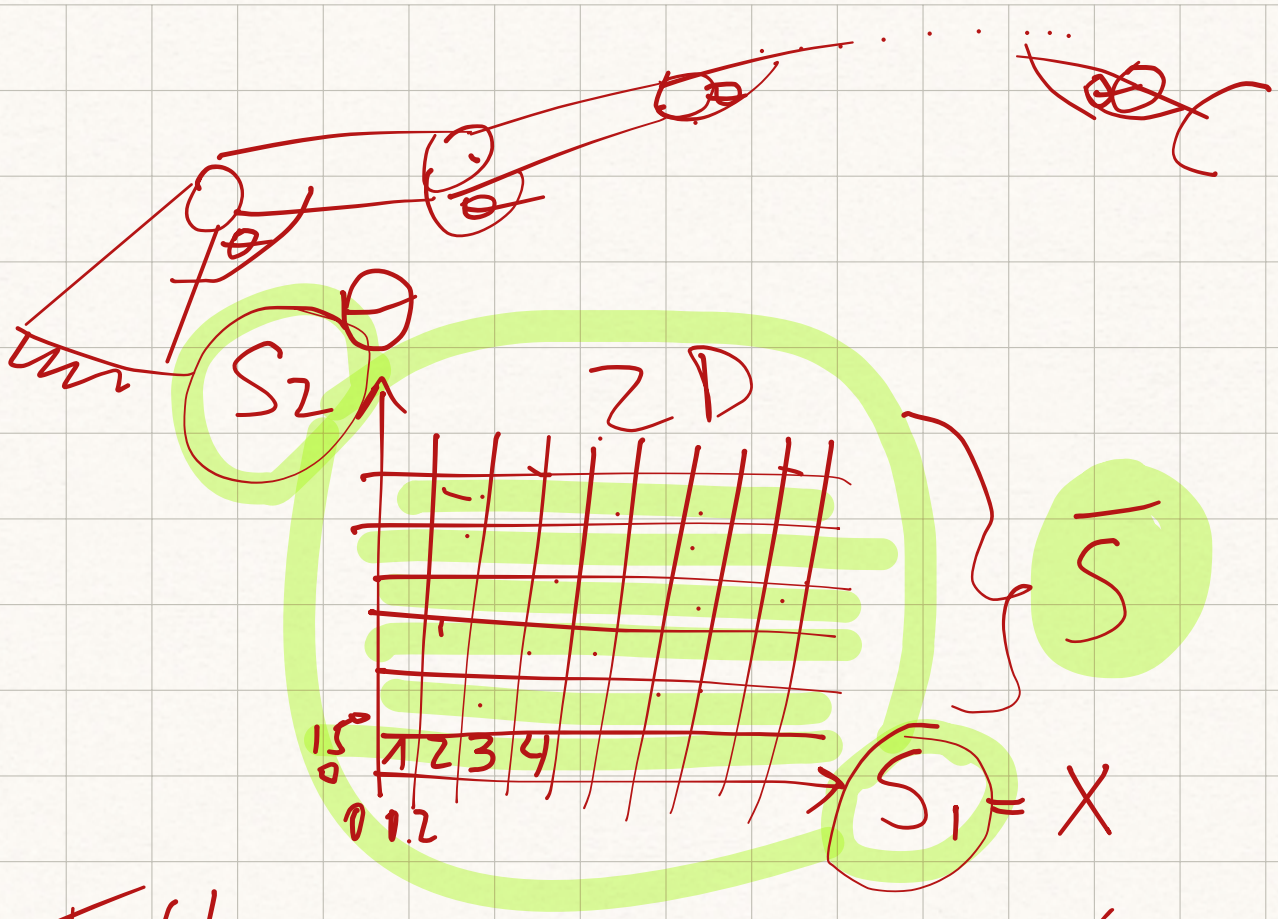


$$S = \{x, \theta, \dot{x}, \dot{\theta}\}$$

$$S = \{-, -, -, -, \ldots\ldots, n\}$$

$$S = \mathbb{R}^n$$





Fabrica 100 robots

cada robot ejecuta 2 acciones

Conjunto de Estados de la fabrica

$$2^{100}$$

r_1	r_2	r_3		r_{100}
2 ·	2 ·	2	. . . . .	· 2 =

Aproximación de la fn de Valor

$$V^* \approx \theta^T \phi(s)$$

Modelos

Regresiones lineales

s → Modelo → $s' \sim P_{sa}$

$s_{t+1} = As_t + Ba_t$

a

✗ Simulador físico

✗ Aprender para Data

Piloto:

$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \dots\dots s_T$

$s_0^{(1)} \xrightarrow{a_0^{(1)}} s_1^{(1)} \xrightarrow{a_1^{(1)}} s_2^{(1)} \xrightarrow{a_3^{(1)}} \dots\dots s_T^{(1)}$

$s_0^{(n)} \xrightarrow{a_0^{(n)}} s_1^{(n)} \xrightarrow{a_1} s_2^{(n)} \xrightarrow{a_3} \dots\dots s_T^{(n)}$

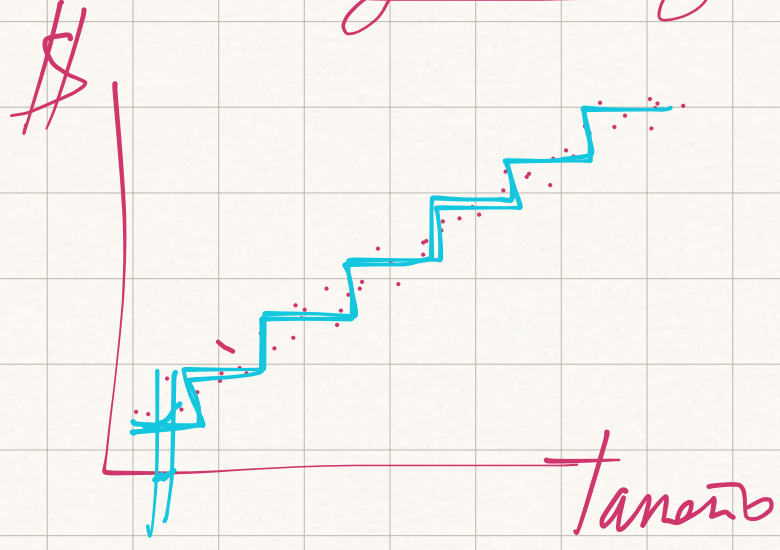
Carla

Obtener Modelo

$(s_{t+1} = As_t + Ba_t)$

$(s_{t+1}, r_{t+1}, s_t, a_t)$

Algoritmos de RL basados en Model → ↓ generalización



Model - Free