fn. lineales y no lineales, estocásticas y determinísticas para encontrar

Det: $$\boxed{s_{t+1} = \underline{\underline{A}}\, s_t + \underline{\underline{B}}\, a_t}$$

Estocástica: $$\boxed{s_{t+1} = A s_t + B a_t + \xi_t}$$

noise

$$\boxed{\xi_t \sim N(\mu, \underset{\Sigma}{\sigma^2})}$$

ξ (ruido) → epsilon

→ viento en un helo

Algoritmos basados en Modelo

$$\begin{array}{c} s \rightarrow \boxed{f} \rightarrow s' \\ \uparrow \\ a \end{array}$$

→ sin física

→ aprende x datos

— funciones lin

— " no lin

$s'_{t+1} = As_t + Ba_t$

- Teniendo el Modelo, lo puedo alimentar
- Mejora algo RL — Model-based
- Model-Free → Simulador
  → Juego de Video

* Modelo híbridos (M-based + M Free)

* Modelar el ruido Eg

SIM 2 REAL

→ Modelo Det. (sin ruido) son muy frágiles

→ 

Actor-Crítico → A2C / A3C

Actor → Acciones, según estado

Critic → [Evaluando esas acciones] (Q-Valor)

Funciones No Lineales poseen estos "vectores" de atributo ← Det / Entrada

$\phi_1(s)$, $\phi_2(a)$

ϕ_s, ϕ_a

+ LVA

Es un algo para aproxima la función Valor en un espacio de Est. Cont. (MDP), asumiendo $S = \mathbb{R}^n$, A (discreta y pequeños)

$$\boxed{V(s) = \theta^T \phi(s)}$$

→ conj. parametros

$$\phi(s) = \begin{bmatrix} x \\ x \\ x^2 \\ xx \end{bmatrix}$$

$\measuredangle\ x\theta$

Conclusión

1.

$\theta \leq 24^0 \longrightarrow R = -1$

$\dot{x}\theta \longrightarrow R \rightarrow = -1$

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

$V(s) := R(s) + \gamma \max_a E_{s' \sim P_{sa}} [V(s')]$

ojo

$E = x P(x)$

Generar muestra de estado aleatorio

IVA → (1) $\boxed{S^{(1)}, \ldots, S^{(n)}}$

(2) Inicialize $\theta := 0$

(3) Repeat {

⇒ for $i = 1, \ldots, n$

⇒ for each action $a \in A$

— { sample $\boxed{S_1', S_2', \ldots, S_k'} \sim P_{sa}^{(i)}$

Q-Values:

Set $q(a) = \frac{1}{k} \sum \left[R(s^{i}) + \gamma V(s') \right]$

$Q(s,a)$.

$q(a) = \frac{1}{K}\sum_{j=1} [R(s) + \gamma V(s_j')]$

$$\boxed{q(a) = E_{s'}[R(s) + \gamma V(s')]}$$

$$\underbrace{\max_a E_{s'}[R(s) + \gamma V(s')]}_{y^{(i)}}$$

$$y^{(i)} = \max_a q(a)$$

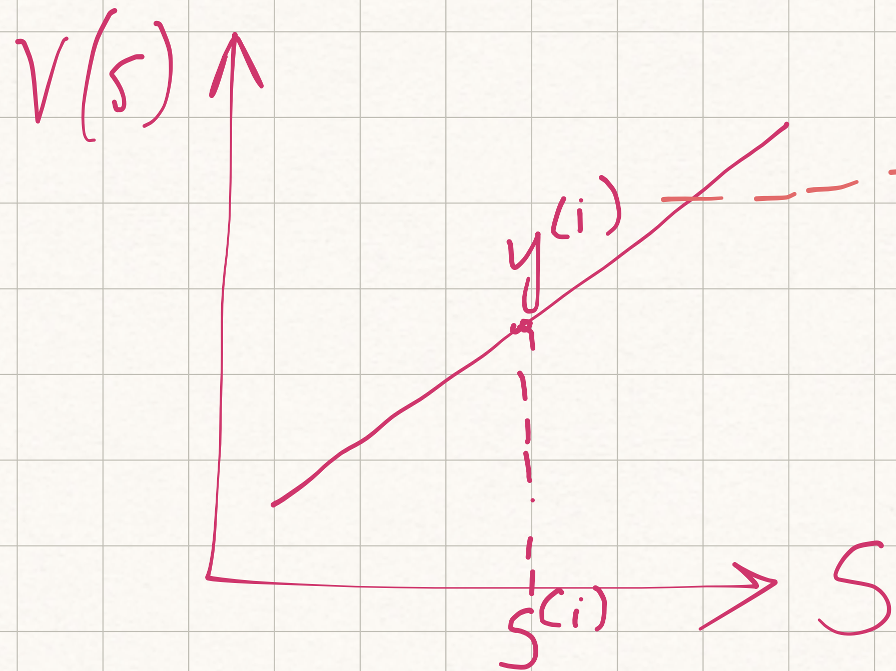
$$VI \Rightarrow V(s^{(i)}) := y^{(i)}$$

$$FVA \Rightarrow \boxed{V(s^{(i)}) \approx y^{(i)}}$$

aprox

$$\theta^T \phi(s^{(i)}) \approx y^{(i)}$$

Set $\theta := \arg\min_{\theta} \underbrace{\frac{1}{2} \sum_{i=1}^{n} \left(\theta^T \phi(s^i) - y^{(i)}\right)^2}_{\text{función de costo}}$



función de mapeo

$s \to V(s)$

Generar el Modelo

EQUIPO 1

$s \to \boxed{M} \to s'$

a

SIM DATA

Aplicar el Método de It. Valor Ajustado

EQUIPO 2

V^*

En la práctica

$\sum^{K} [R(s^i) + \gamma \boxed{V(s^i)}]$

$$q(a) = \frac{1}{K}\sum_{j=1}^{K}\left[R(s') + \gamma \boxed{V(s_j)}\right]$$

$$\underline{\theta^T \phi(s_j)}$$

El último valor de $V(s')$ sirve para actualizar $V(s)$

DRL / $f_\theta(s,a) = nn(s,a)$

RL / $f_\theta(s,a) = A\phi_1(s) + B\phi_2(a) + \xi$

NN / → Evitas crear vectores de atributos

VA / → $V^* \longrightarrow \pi^*$

$$\pi^*(s) = \arg\max_a E_{s'}\left[V^*(s')\right]$$

$\boxed{\text{Generalización}}$