10/3/25

Continuación de procesos de Markov

Distribución de probabilidad para todos los valores:

X_t V. A. en el tiempo t

Valores$(X_t) = \{S_1, \dots, S_n\}$

$$Pr(X_t) = \begin{bmatrix} Pr(X_t = S_1) \\ \vdots \\ Pr(X_t = S_n) \end{bmatrix} \rightarrow \sum_{i=1}^{n} Pr(X_t = S_n) = 1$$

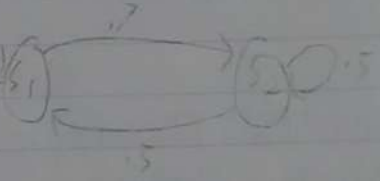
X_t es un proceso de Markov de primer orden

$$Pr(X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \dots, X_0) = Pr(X_{t+1} \mid X_t)$$

$$Pr(X_{t+1} = S_i \mid X_t = S_j) = P_{ij} \rightarrow \sum_{i=1}^{n} P_{ij} = 1$$

Ejemplo:

$$P = \begin{bmatrix} 0.3 & 0.5 \\ 0.7 & 0.5 \end{bmatrix}$$



$$P(X_1) = \begin{bmatrix} .4 \\ .6 \end{bmatrix}$$

$$P(X_{t+1} = S_i) = \sum_{j=1}^{n} P_{ij} * P(X_t = S_j)$$

$$P(X_{t+1} = S_1) = .3 * .4 + .5 * .6$$

$$P(X_{t+1} = S_2) = .7 * .4 + .5 * .6$$

$$P(X_1) = P\,P(X_0)$$

$$P(X_2) = P\,P(X_1) = P\,P\,P(X_0)$$

$$\therefore P(X_t) = P^t P(X_0)$$

10/3/25

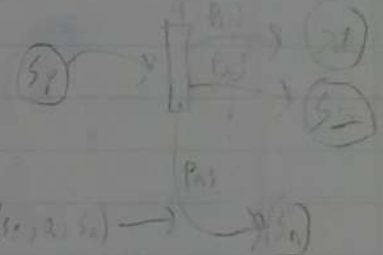
Sp el proceso es estacionario, entonces:
$P(X_{t+1}) = P(X_t)$ para una t donde
$P(X_{t+1}) = P\,P(X_t) \rightarrow$

Si es estacionario entonces hay una t tal que:
$P(X_t = S_1) = .3\,P(X_t = S_1) + .5\,P(X_t = S_2)$
$P(X_t = S_2) = .7\,P(X_t = S_1) + .5\,P(X_t = S_2)$

$-.7\,P(X_t = S_1) + .5\,P(X_t = S_2) = 0$
$-.5\,P(X_t = S_1) - .5\,P(X_t = S_2) = -.5$

$-1.2\,P(X_t = S_1) = -0.5 \rightarrow P(X_t = S_1) = \frac{5}{12}$
$P(X_t = S_2) = 7/12$

Concepto Ejemplo:
S conjunto estados
A conjunto acciones
$S_0 \in S$ estado principal
$S_F \subseteq S$ conjunto de estados finales



Acciones legales $S \rightarrow P(A)$
transición: $S \times A \rightarrow S$
costo_local: $S \times A \rightarrow \mathbb{R}^+ - \{0\}$

Modelo de decisión de Markov (MDP)

11/3/28

Procesos de Markov:

MDP: Modelo matemático para toma de decisiones bajo incertidumbre. Fue introducido en 1950's-1960's. El término "Markov" se refiere a Andrey Markov, pues MDP son extensiones de las cadenas de Markov.

Este se usa en áreas de Robótica, Almacenamiento y Agricultura, como ejemplo.

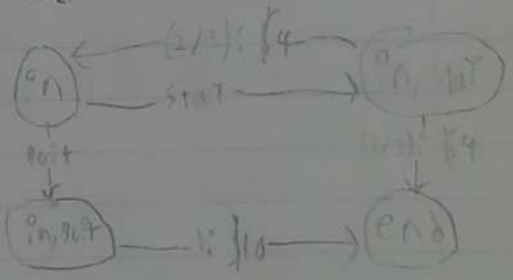
Ejemplo de Markov: Juego de Dados

En cada ronda:
- Eliges quedarte o salir
- Si te quedas, ganas $4
- Si te sales, ganas $10

El juego termina si te sales o cae 1 o 2.

Recompensa (utilidad)

$E[Y \mid \pi_{stay}] = 12$



1/3/25

Definicion de MDP (Proceso de decisiones de Markov):

Class MDP:

S conjunto de estados

A conjunto de acciones

def acciones(self, s):

def transicion(self, s, a, s'): $T(s,a,s')$
regresa la $Pr(S_{t+1} = s' \mid S_t = s \wedge A_t = a)$

def recompensa(self, s, a, s'):
regresa un numero $R(s,a,s')$

def terminal(self, s):
true si s es final, si no False

$0 \leq \gamma \leq 1$

$R_t = \gamma^0 r_t + \gamma^1 r_{t+1} + \gamma^2 r_{t+2} + \ldots = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$

12/3/25

Definición alterna de MDPi

$MDP = \langle S, A, A, T, R, S_T, \gamma \rangle$

$S = \{s_1, s_2, \ldots, s_n\}$

$A = \{a_1, \ldots, a_m\}$

$A: S \to P(A)$ $A(s)$ acciones legales en $s \in S$

$T: S \times A \times S \to R$ $T(s, a, s') = P_r(S_{t+1} = s' \mid S_t = s, a_t = a)$

$\sum_{s' \in S} T(s, a, s') = 1$

$r: S \times A \times S \to R$ $r(s, a, s')$ es la recompensa de ir de s a s' con la acción a.

$S_T \subseteq S$ conjunto de estados terminales

$0 \leq \gamma \leq 1$ factor de descuento

$$R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$$

Retorno en el instante t

Objetivo: Encontrar la mejor política

$\pi: S \to A$ tal que $\pi(s) = a$

π_1 y π_2 dos políticas diferentes

$V^{\pi}(s) = E^{\pi}[R_t \mid S_t = s]$

$V^{\pi}(s) = \sum_{s' \in S} T(s, \pi(s), s') [r(s, \pi(s), s') + \gamma E^{\pi}[R_{t+1} \mid S_{t+1} = s']]$

$V^{\pi}(s) = \sum_{s' \in S} T(s, \pi(s), s') [r(s, \pi(s), s') + \gamma V^{\pi}(s')]$

12/3/25

Ejemplo: Juego de dados

$S = \{s_1, s_2\}$

$A = \{s, 1\}$

$T(s_1, 1, s_1) = 0$ $\quad R(s_1, 1, s_1) = 0$

$T(s_1, 1, s_2) = 1$ $\quad R(s_1, 1, s_2) = 10$

$T(s_1, s, s_1) = 2/3$ $\quad R(s_1, s, s_1) = 4$

$T(s_1, s, s_2) = 1/3$ $\quad R(s_1, s, s_2) = 4$

$$V^{\pi_2}(s_1) = T(s_1, s, s_1)[R(s_1, s, s_1) + \gamma V^{\pi_2}(s_1)] + T(s_1, s, s_2)[R(s_1, s, s_2) + \gamma V^{\pi_2}(s_2)]$$

$$V^{\pi_1}(s_1) = \frac{2}{3}\left[4 + V^{\pi_2}(s_1)\right] + \frac{1}{3}[4 + 0]$$

$$\frac{1}{3} V^{\pi_1}(s_1) = \frac{8}{3} + \frac{4}{3} = 12$$

$\pi_1 \leq \pi_2$ si y solo si

$V^{\pi_1}(s) \leq V^{\pi_2}(s) \quad \forall s \in S$

Y π^* es una política óptima si:

$V^{\pi}(s) \leq V^{\pi^*}(s) \quad \forall s \in S, \ \forall \pi \in \Pi$

$$V^*(s) = \max_{a \in A(s)} \left(\sum_{s' \in S} T(s, a, s')[R(s, a, s') + \gamma V^*(s')] \right)$$

↓

Ecuación de optimalidad de estado de Bellman

14/3/25

Ejemplo cambio multiple (doble)

1 2 3 4 5 6

$T(1, '(', 2) = 1$

$T(1, '(', x) = 0 \quad \forall x \neq 2$

$T(5, x, 5) = 1 \quad \forall x \in A, \forall \neq 5$

$T(5, x, 1) = 0$

$T(6, x, 6) = 1 \quad \forall x$

$T(6, x, 1) = 0$

$T(1, 6, 2) = \beta$

$T(1, 6, 1) = 1 - \beta$

$T(1, 6, x) = 0 \quad \forall x \neq 1, 2$