

Probabilidades de transición parametrizadas por una política

$$\pi_0 = \begin{bmatrix} \neg \text{Anuncio} \\ \neg \text{Anuncio} \\ \neg \text{Anuncio} \end{bmatrix} \quad T(\pi_0) = \begin{bmatrix} 0.7 & 0.2 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.6 \end{bmatrix} \quad V_0(s) = \begin{bmatrix} 91.07 \\ 104.20 \\ 135.10 \end{bmatrix}$$

Encontramos la política π_1 para la iteración siguiente

$$\pi_1(s) = \underset{a}{\operatorname{argmax}} P(s' \mid s, a) V_0(s')$$

$$\begin{aligned} \pi_1(s=d) = \underset{a}{\operatorname{argmax}} \{ & P(s'=d \mid s=d, a=\neg \text{Anuncio}) V_0(s'=d) + \\ & P(s'=b \mid s=d, a=\neg \text{Anuncio}) V_0(s'=b) + \\ & P(s'=c \mid s=d, a=\neg \text{Anuncio}) V_0(s'=c), \\ & P(s'=d \mid s=d, a=\text{Anuncio}) V_0(s'=d) + \\ & P(s'=b \mid s=d, a=\text{Anuncio}) V_0(s'=b) + \\ & P(s'=c \mid s=d, a=\text{Anuncio}) V_0(s'=c) \} \end{aligned}$$

