

Probabilidades de transición parametrizadas por una política

$$\pi_0 = \begin{bmatrix} \neg \text{Anuncio} \\ \neg \text{Anuncio} \\ \neg \text{Anuncio} \end{bmatrix} \quad T(\pi_0) = \begin{bmatrix} 0.7 & 0.2 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.6 \end{bmatrix} \quad V_0(s) = \begin{bmatrix} 91.07 \\ 104.19 \\ 135.10 \end{bmatrix}$$

Encontramos la política π_1 para la iteración siguiente

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

$$\pi_1(s=d) = \operatorname{argmax}_a \{$$

$$\begin{matrix} 101.19, \\ 103.81 \end{matrix}$$

Seleccionamos la acción Anuncio que corresponde con el máximo

