* Métodos de programación dinámica
  + Requieren un conocimiento del modelo completo
  + Principio de la optimalidad de bellman
* Algoritmo de DP para encontrar pi\*
  + Iteración de política:
    - Dos pasos:
      * Evaluacion de la política
        + Dada una política pi[i] calcular Ui = Ui[πi] (la utilidad de cada estado si se ejecuta πi)
        + Mejora de la política.

Calcula una nueva máxima utilidad esperada πi+1 usando una vista hacia delante de un paso basada en Ui

* + - * + Π\*(s) = argmax [a€ A(s)] ∑P(s`|s,a) U(S`)
      * El algoritmo termina cuando no cambia las utilidades
        + Ui(s) = R(s) + Y∑P(s’|si πi(s)) Ui(s)
* Metodos libres de modelo