

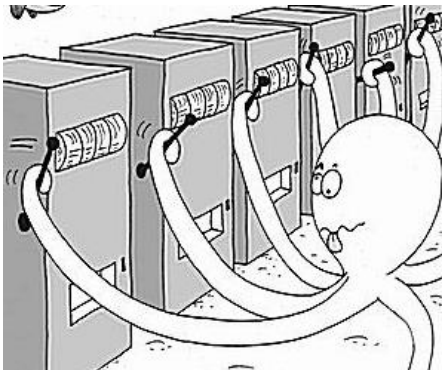
Machine Learning

Equations de Bellman

Equation de Bellman

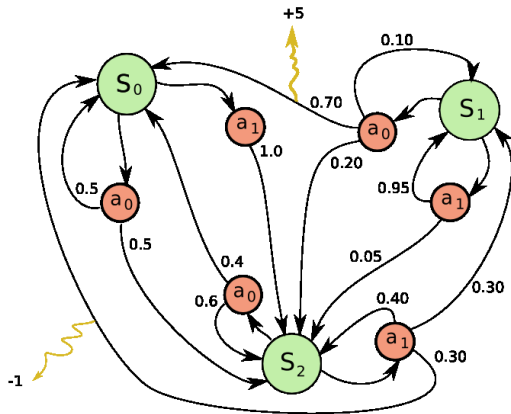
- Meilleure action, meilleure value :
- $V(S_0) = \max_{\{a_t\}} \sum_{t=0}^{\infty} \gamma^t \cdot R(s_t, a_t)$
- où $0 < \gamma < 1$ pénalise les récompenses lointaines
-
- Principe d'optimalité de bellman :
- “une séquence d'action optimale démarre par l'action optimale”
- (même concept que pour le chemin le plus court)
- $\Rightarrow V(S_0) = \max_{a_t \in A_t} [R(s_t, a_t) + \gamma V(M(s_t, a_t))]$

K-armed bandits



Apprentissage par Renforcement

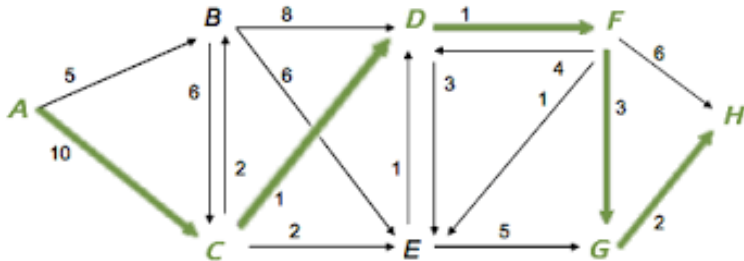
Markov Decision Process



Apprentissage par Renforcement

Programmation Dynamique (planification et ordonnancement)

- Solution optimale = combinaison de solutions optimale à des sous-problèmes
- programmation récursive suivant les equations de Bellman



TD-Learning

- Pas de modèle, estimation de la valeur uniquement
- $V(S_t) \leftarrow V(S_t) + \alpha(R_t + \gamma V(S_{t+1}) - V(S_t))$
- où :
- α est un pas d'apprentissage
- γ est le taux de dévaluation d'une récompense lointaine





Equations de Bellman



ML WEEK

■

■

■

■

■

■

■

■