MPD

Přechodový model: $P(s'|s,a)$ pravděpodobnosti přechodů do různých stavů

Odměny: $R(s)$

Utilitní funcke: $U(s_0,\ldots,s_n) = R(s_0) + \gamma R(s_1) + \ldots \gamma^n R(n_n)$ pro $0 < \gamma < 1$

Řešení: π policy dává distrubici přes stavy $\pi(s)$. Přáli bychom si tu nejlepší $argmax_\pi U^\pi(s)$

Optimální policy by pak měla být tak, která dává $\pi^*(s) = argamx_a \sum_s P(s|s',a)U(s)$

DEF Bellmanova rovnice:

$$U(s) = R(s) + \gamma max_a \sum_s P(s|s',a)U(s)$$

Vypočítat takové hodnoty by bylo těžké můžeme použít nějakou interační metodu:

- Iterace Hodnot
  - Pro každý stav updatetujeme jeho hodnotu v závislosti na všech stavech, do kterých se z něho můžeme dostat zvolením nejlepší akce. Když tohle budeme opakova, tak to zkonverguje.
- Iterace Policy
  - Každý stav má takovou utilitu, kterou mu přiřadí naše utilitní funkce. Pokud Existuje lepší akce, než nám říká policy, updatujeme policy a podle ní i všechny utility