

Model-Based Planning

Evaluation

Bellman expectation equation

Policy evaluation

+ greedy policy improvement

Control

Bellman optimality equation

Value iteration

of loop(s) for policy evaluation = 1

Generalized Policy iteration

Policy iteration

Model-Free Learning

Sampling over an entire episode

MC

Update at each step

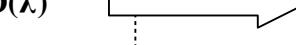
($\lambda=1$)

λ -discounted weighting for all n-step TD

TD(0)

Evaluate $q(s,a)$

+ ϵ -greedy ($\epsilon \rightarrow 0$)



Update at each step

($\lambda=1$)

λ -discounted weighting for all n-step TD

GLIE MC control

SARSA(0)

Both off-policy.
Learn from a different policy

Q-learning

Importance sampling TD

Illustrated by Lili Mou

Email: doublepower.mou at the mail server provided by Google (gmail)