 

Sarsa(𝜆)

Offline: online accumulated, offline update

Online: update at each step

MC = offline TD(1)

TD=TD(0)

Offline update is identical

Forward: Like MC, can only computed from complete episodes

average n-step returns over different n

TD(lambda)

Q-Learning

Importance Sampling for Off-Policy TD

Importance Sampling for Off-Policy MC

Importance Sampling

On-policy: Learn about policy 𝜋 from experience sampled from 𝜋

Off-policy: Learn about policy 𝜋 from experience sampled from 𝜇

Off-Policy Learning

Sarsa (TD(0) on Q)

On-Policy Temporal-Difference Learning

MC policy evaluation

GLIE MC control

On-Policy Monte-Carlo Learning

TD

Lec9 RL2