Q-learning

Initialize Q(s, a) arbitrarily

Repeat (for each episode):

Initialize s

Repeat (for each step of episode):

Choose a from s using policy derived from Q

Take action a, observe r, s’

Q(s, a) = Q(s, a) + α [r + γ maxα’ Q(s’, a’) – Q(s, a)]

s = s’

Until s is terminal

SARSA

Initialize Q(s, a) arbitrarily

Repeat (for each episode):

Initialize s

Choose a from s using policy derived from Q

Repeat (for each step of episode):

Take action a, observe r, s’

Choose a’ from s’ using policy derived from Q

Q(s, a) = Q(s, a) + α[r + γQ(s’, a’) – Q(s, a)]

s = s’, a = a’

until s is terminal