COMP3411 Week 07 Tutorial

Yifan He

z5173587@unsw.edu.au
https://github.com/hharryyf/COMP3411-24T1-tutoring

Probability Revision

Bayes Rule

- $P(\theta)$: Prior distribution of parameters
- D is the set of data (observations)
- $P(\theta|D)$: Posterior distribution
- $P(D|\theta)$: likelihood

Probability Revision

Bayes Rule

- $P(\theta)$: Prior distribution of parameters
- D is the set of data (observations)
- $P(\theta|D)$: Posterior distribution
- $P(D|\theta)$: likelihood
- $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$

Conditionally Independent

- Suppose A, B, C are events, A is conditionally independent of B given C iff $P(C) > 0$ and $P(A|B, C) = P(A|(B \wedge C)) = P(A|C)$

Q1

Assume 50% of the population are men and 50% are women. Only 4% of the population are color blind, but 7% of men are color blind. What percentage of color blind people are men?

- Formulate the question mathematically
- $P(men) = 50\%$, $P(women) = 50\%$
- $P(color\_blind) = 4\%$

Q1

Assume 50% of the population are men and 50% are women. Only 4% of the population are color blind, but 7% of men are color blind. What percentage of color blind people are men?

- Formulate the question mathematically
- $P(men) = 50\%$, $P(women) = 50\%$
- $P(color\_blind) = 4\%$
- $P(color\_blind|men) = 7\%$

Q1

Assume 50% of the population are men and 50% are women. Only 4% of the population are color blind, but 7% of men are color blind. What percentage of color blind people are men?

- Formulate the question mathematically
- $P(men) = 50\%$, $P(women) = 50\%$
- $P(color\_blind) = 4\%$
- $P(color\_blind|men) = 7\%$
- $P(men|color\_blind) = ?$

Q1

- $P(men) = 50\%$, $P(women) = 50\%$
- $P(color\_blind) = 4\%$, $P(color\_blind|men) = 7\%$
- $P(men|color\_blind) = ?$

- We apply Bayes Rule
- $P(men|color\_blind) = \frac{P(color\_blind|men)P(men)}{P(color\_blind)}$
- $P(men|color\_blind) = \frac{7\% \cdot 50\%}{4\%} = 87.5\%$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- Example: $P(toothache, catch, cavity) = 0.108$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (i) Calculate $P(toothache \wedge \neg catch)$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (i) Calculate $P(toothache \wedge \neg catch)$
- Just see which entry in the table satisfies $toothache \wedge \neg catch$, and sum up the probability

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (i) Calculate $P(toothache \land \neg catch)$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (i) Calculate $P(toothache \wedge \neg catch)$
- Just see which entry in the table satisfies $toothache \wedge \neg catch$, and sum up the probability

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (i) Calculate $P(toothache \wedge \neg catch)$
- Just see which entry in the table satisfies $toothache \wedge \neg catch$, and sum up the probability
- $P(toothache \wedge \neg catch) = 0.012 + 0.064 = 0.076$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (ii) Calculate $P(catch)$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (ii) Calculate $P(catch)$
- Just see which entry in the table satisfies *catch*, and sum up the probability

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (ii) Calculate $P(catch)$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (ii) Calculate $P(catch)$
- Just see which entry in the table satisfies *catch*, and sum up the probability
- $P(catch) = 0.108 + 0.016 + 0.072 + 0.144 = 0.34$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (iii) Calculate $P(cavity|catch)$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- (iii) Calculate $P(cavity|catch)$
- Use the definition of conditional probability
- $P(cavity|catch) = \frac{P(cavity \wedge catch)}{P(catch)}$
- $P(cavity|catch) = \frac{0.108+0.072}{0.108+0.072+0.016+0.144} = 0.53$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (iii) Calculate $P(cavity|catch)$
- Use the definition of conditional probability
- $P(cavity|catch) = \frac{P(cavity \wedge catch)}{P(catch)}$
- $P(cavity|catch) = \frac{0.108+0.072}{0.108+0.072+0.016+0.144} = 0.53$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (iv) Calculate $P(cavity \mid toothache \vee catch)$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (iv) Calculate $P(cavity|toothache \vee catch)$
- Use the definition of conditional probability
- $P(cavity|toothache \vee catch) = \frac{P(cavity \wedge (toothache \vee catch))}{P(toothache \vee catch)}$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (iv) Calculate $P(cavity|toothache \vee catch)$
- Use the definition of conditional probability
- $P(cavity|toothache \vee catch) = \frac{P(cavity \wedge (toothache \vee catch))}{P(toothache \vee catch)}$
- $P(cavity \wedge (toothache \vee catch)) = 0.108 + 0.012 + 0.072 = 0.192$

Q2

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- (iv) Calculate $P(cavity|toothache \vee catch)$
- Use the definition of conditional probability
- $P(cavity|toothache \vee catch) = \frac{P(cavity \wedge (toothache \vee catch))}{P(toothache \vee catch)}$
- $P(cavity \wedge (toothache \vee catch)) = 0.192$
- $P(toothache \vee catch) = 0.416$
- $P(cavity|toothache \vee catch) = \frac{0.192}{0.416} = 0.46$

Q2

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Verify the conditional independence claim: catch is conditionally independent of toothache given cavity.
- $P(catch|toothache, cavity) = P(catch|cavity)$

Q2

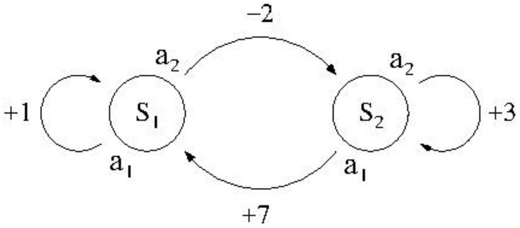
	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Verify the conditional independence claim: catch is conditionally independent of toothache given cavity.
- $P(catch|toothache, cavity) = P(catch|cavity)$
- $LHS = \frac{P(catch \wedge toothache \wedge cavity)}{P(toothache \wedge cavity)} = \frac{0.108}{0.108+0.012} = 0.9$
- $RHS = \frac{P(catch \wedge cavity)}{P(cavity)} = \frac{0.108+0.072}{0.108+0.072+0.012+0.008} = 0.9$

Q-learning Revision

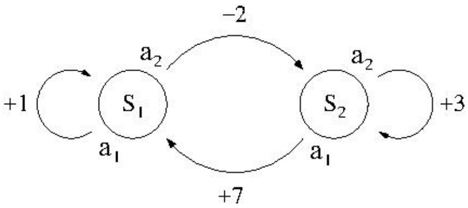
- Given an environment with a set of states, find the optimal policy at each state that will give you the highest expected reward
- $\delta(S, a)$, the transition function
- $r(S, a)$, the reward of performing a at S
- $V^*(S)$, the optimal expected reward at state S
- $Q(S, a)$, the Q function
  - $Q(S, a) = r(S, a) + \gamma V^*(\delta(S, a))$
- $\pi^*(S)$, the optimal policy at S
  - $V^*(S) = r(S, \pi^*(S)) + \gamma V^*(\delta(S, \pi^*(S)))$

Q4 ($\gamma = 0.7$)



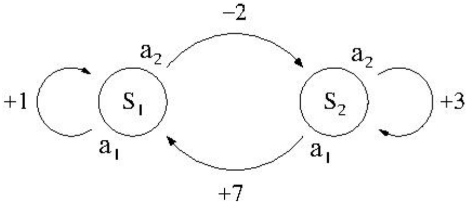
-
  - $\delta(S_1, a_1) = S_1, r(S_1, a_1) = +1$
  - $\delta(S_1, a_2) = S_2, r(S_1, a_2) = -2$
  - $\delta(S_2, a_1) = S_1, r(S_2, a_1) = +7$
  - $\delta(S_2, a_2) = S_2, r(S_2, a_2) = +3$

Q4 ($\gamma = 0.7$)



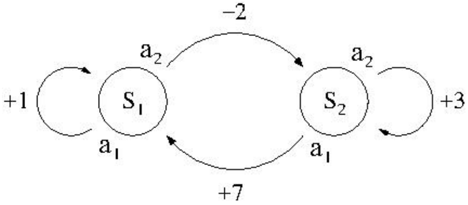
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$

Q4 ($\gamma = 0.7$)



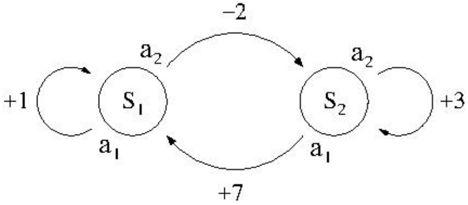
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Two possibilities $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - Or, $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$

Q4 ($\gamma = 0.7$)



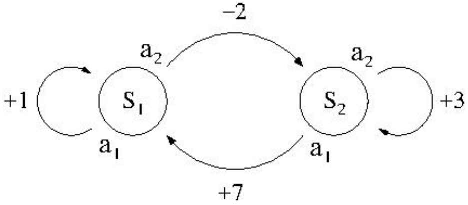
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$

Q4 ($\gamma = 0.7$)



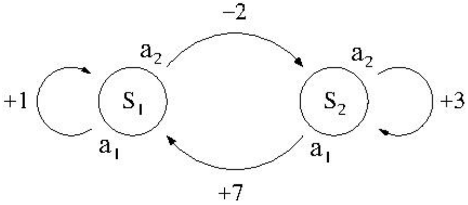
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$

Q4 ($\gamma = 0.7$)



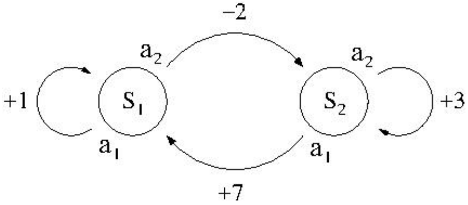
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - $V(S_1) = 1 + \gamma V(S_1)$
  - $V(S_2) = 3 + \gamma V(S_2)$

Q4 ($\gamma = 0.7$)



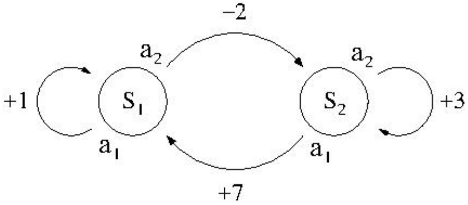
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - $V(S_1) = 1 + \gamma V(S_1)$
  - $V(S_2) = 3 + \gamma V(S_2)$
  - $V(S_1) = 3.33$ and $V(S_2) = 10$

Q4 ($\gamma = 0.7$)



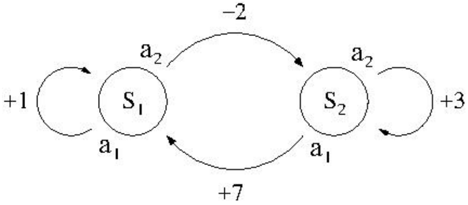
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - $V(S_1) = 1 + \gamma V(S_1)$
  - $V(S_2) = 3 + \gamma V(S_2)$
  - $V(S_1) = 3.33$ and $V(S_2) = 10$
  - Suppose $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$

Q4 ($\gamma = 0.7$)



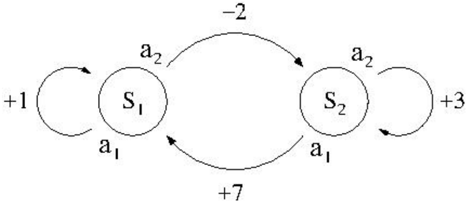
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - $V(S_1) = 1 + \gamma V(S_1)$
  - $V(S_2) = 3 + \gamma V(S_2)$
  - $V(S_1) = 3.33$ and $V(S_2) = 10$
  - Suppose $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$
  - $V(S_1) = -2 + \gamma V(S_2)$
  - $V(S_2) = 7 + \gamma V(S_1)$

Q4 ($\gamma = 0.7$)



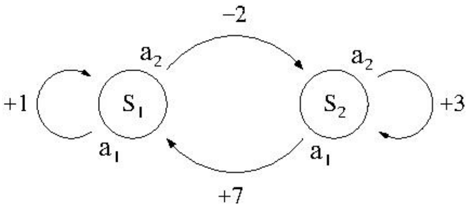
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
  - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
  - $V(S_1) = 1 + \gamma V(S_1)$
  - $V(S_2) = 3 + \gamma V(S_2)$
  - $V(S_1) = 3.33$ and $V(S_2) = 10$
  - Suppose $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$
  - $V(S_1) = -2 + \gamma V(S_2)$
  - $V(S_2) = 7 + \gamma V(S_1)$
  - $V(S_1) = 5.69$ and $V(S_2) = 10.98$

Q4 ($\gamma = 0.7$)



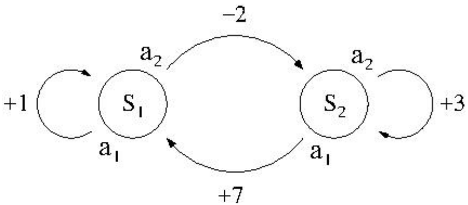
- Determine $\pi^*(S_1)$ and $\pi^*(S_2)$
    - Suppose $\pi^*(S_1) = a_1$ and $\pi^*(S_2) = a_2$
    - $V(S_1) = 1 + \gamma V(S_1)$
    - $V(S_2) = 3 + \gamma V(S_2)$
    - $V(S_1) = 3.33$ and $V(S_2) = 10$
    - Suppose $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$
    - $V(S_1) = -2 + \gamma V(S_2)$
    - $V(S_2) = 7 + \gamma V(S_1)$
    - $V(S_1) = 5.69$ and $V(S_2) = 10.98$
- $\pi^*(S_1) = a_2$ and $\pi^*(S_2) = a_1$

Q4 ($\gamma = 0.7$)



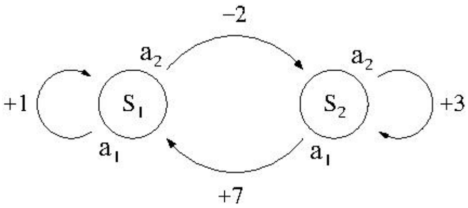
- Determine the value function V^*
  - The expected reward when following the optimal policy

Q4 ($\gamma = 0.7$)



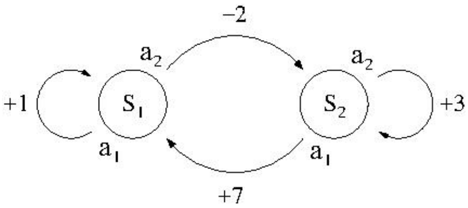
- Determine the value function V^*
  - The expected reward when following the optimal policy
  - $V^*(S_1) = 5.69$ and $V^*(S_2) = 10.98$

Q4 ($\gamma = 0.7$)



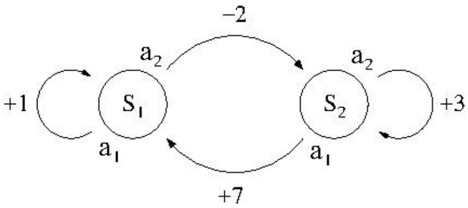
- Determine the Q-function
  - $Q(S_1, a_1)$
  - $Q(S_1, a_2)$
  - $Q(S_2, a_1)$
  - $Q(S_2, a_2)$

Q4 ($\gamma = 0.7$)



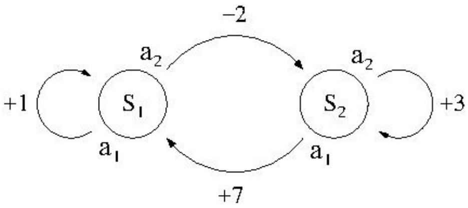
- Determine the Q-function
  - $Q(S_1, a_1) = 1 + \gamma V^*(S_1)$
  - $Q(S_1, a_2) = -2 + \gamma V^*(S_2)$
  - $Q(S_2, a_1) = 7 + \gamma V^*(S_1)$
  - $Q(S_2, a_2) = 3 + \gamma V^*(S_2)$

Q4 ($\gamma = 0.7$)



- Determine the Q-function
  - $Q(S_1, a_1) = 1 + \gamma V^*(S_1) = 1 + 0.7 \cdot 5.69 = 4.98$
  - $Q(S_1, a_2) = -2 + \gamma V^*(S_2) = -2 + 0.7 \cdot 10.98 = 5.69$
  - $Q(S_2, a_1) = 7 + \gamma V^*(S_1) = 7 + 0.7 \cdot 5.69 = 10.98$
  - $Q(S_2, a_2) = 3 + \gamma V^*(S_2) = 3 + 0.7 \cdot 10.98 = 10.69$

Q4 ($\gamma = 0.9$)

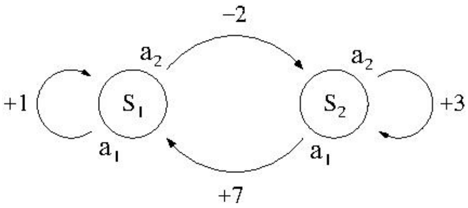


- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a), b)$

Q	a_1	a_2
S_1	0	-2
S_2	0	0

Current state	action	new Q value
S_1	a_2	$-2+\gamma \cdot 0 = -2$
S_2		

Q4 ($\gamma = 0.9$)

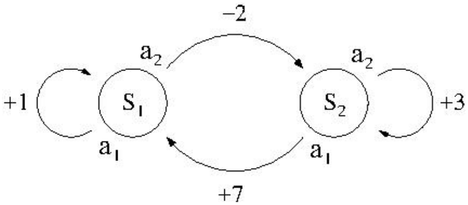


- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a),b)$

Q	a_1	a_2
S_1	0	-2
S_2	7	0

Current state	action	new Q value
S_1	a_2	$-2 + \gamma \cdot 0 = -2$
S_2	a_1	$7 + \gamma \cdot 0 = 7$

Q4 ($\gamma = 0.9$)

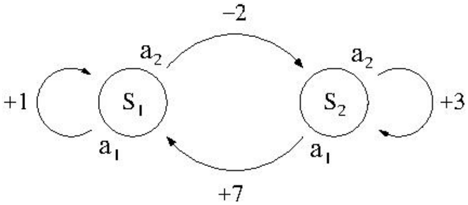


- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a),b)$

Q	a_1	a_2
S_1	0	-2
S_2	7	0

Current state	action	new Q value
S_1	a_2	$-2+\gamma\cdot 0=-2$
S_2	a_1	$7+\gamma\cdot 0=7$
S_1	a_1	$1+\gamma\cdot 0=1$

Q4 ($\gamma = 0.9$)

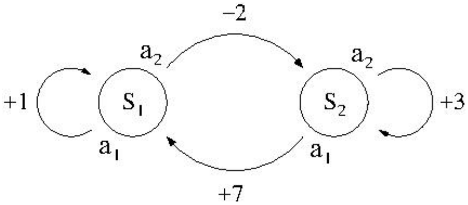


- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a),b)$

Q	a_1	a_2
S_1	1	-2
S_2	7	0

Current state	action	new Q value
S_1	a_2	$-2+\gamma\cdot 0=-2$
S_2	a_1	$7+\gamma\cdot 0=7$
S_1	a_1	$1+\gamma\cdot 0=1$
S_1		

Q4 ($\gamma = 0.9$)

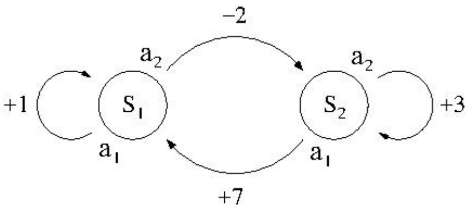


- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a),b)$

Q	a_1	a_2
S_1	1	-2
S_2	7	0

Current state	action	new Q value
S_1	a_2	$-2+\gamma\cdot 0 = -2$
S_2	a_1	$7+\gamma\cdot 0 = 7$
S_1	a_1	$1+\gamma\cdot 0 = 1$
S_1	a_2	$-2+\gamma\cdot 7 = 4.3$

Q4 ($\gamma = 0.9$)



- Trace the Q-learning algorithm (initial state S_1)
- $Q(S,a) = r(S,a) + \gamma max_b Q(\delta(S,a), b)$

Q	a_1	a_2
S_1	1	4.3
S_2	7	0

Current state	action	new Q value
S_1	a_2	$-2 + \gamma \cdot 0 = -2$
S_2	a_1	$7 + \gamma \cdot 0 = 7$
S_1	a_1	$1 + \gamma \cdot 0 = 1$
S_1	a_2	$-2 + \gamma \cdot 7 = 4.3$