

Started on	Friday, 22 September 2023, 8:31 PM
State	Finished
Completed on	Friday, 22 September 2023, 9:06 PM
Time taken	34 mins 55 secs
Grade	10.00 out of 10.00 (100%)

Question 1

🚩 Flag questionMark 1.00 out of 1.00Correct

Temporal difference estimation  
Chapter 6.1

Temporal difference methods estimate the value function incrementally.

Consider a RL problem, with state space  $X = \{ E, F \}$  and action space  $U = \{ a, b \}$ . Discount factor  $\gamma = 0.5$ .

Let the current state-value estimate for a state be  $V(E)=2$ ,  $V(F)=2$ . Let us use TD(0) with  $\alpha = 0.5$  (see Eq. 6.2) to update the state-value estimate.

What is the state-value  $V(E)$  after TD(0) update if the following state-action-reward-state-action sequence is observed:

E, a, 1, F, b

Answer: ✔

The correct answer is: 2.00

Question 2

🚩 Flag questionMark 1.00 out of 1.00Correct

Chapter 6.4

Consider the RL problem from the previous task.

Let the current action value estimates be  $Q(E,a)=2$ ,  $Q(E,b)=2$ ,  $Q(F,a)=0$ ,  $Q(F,b)=4$ . Let us use SARSA with  $\alpha = 0.5$  (see Eq. 6.7) to update the action-value estimate.

What is the action-value  $Q(E,a)$  after SARSA update if the following state-action-reward-state-action sequence is observed:

E, a, 1, F, a

Answer: ✔

The correct answer is: 1.50

Question 3

🚩 Flag questionMark 1.00 out of 1.00Correct

Consider RL problem with  $X=(E,F)$ ,  $U=\{a,b\}$ ,  $\gamma=0.5$ .

Let the current action value estimates be  $Q(E,a)=1$ ,  $Q(E,b)=4$ ,  $Q(F,a)=3$ ,  $Q(F,b)=4$ . Let us use Q-learning with  $\alpha = 0.5$  (see Eq. 6.8) to update the action-value estimate.

What is the action-value  $Q(E,a)$  after Q-learning update if the following state-action-reward sequence is observed:

$x_0 = E, u_0 = a, r_1 = 0, x_1 = F, u_1 = a, r_2 = 4$

Answer: ✔

The correct answer is: 1.50

Question 4

🚩 Flag questionMark 1.00 out of 1.00Correct

Monte-Carlo (MC) estimate of discrete (tabular) state value function  $V$  is guaranteed to converge to the true value.

Select one:

☒ True✔

☐ False

The correct answer is 'True'.

Question 5

🚩 Flag questionMark 1.00 out of 1.00Correct

TD(0) estimate of discrete (tabular) state value function  $V$  is guaranteed to converge to the true value.

Select one:

☒ True✔

☐ False

The correct answer is 'True'.

Question 6

🚩 Flag questionMark 1.00 out of 1.00Correct

TD/SARSA approach can be used to estimate discrete action-value function  $Q(x,u)$  in continuing (non-episodic) environments.

Select one:

☒ True✔

☐ False

The correct answer is 'True'.

Question 7

🚩 Flag questionMark 1.00 out of 1.00Correct

Monte-Carlo (MC) approach can be used to estimate discrete action-value function  $Q(x,u)$  in continuing (non-episodic) environments.

Select one:

☐ True

☒ False✔

The correct answer is 'False'.

Question 8

🚩 Flag questionMark 1.00 out of 1.00Correct

Tabular methods allow handling larger state spaces than function approximation methods.

Select one:

☐ True

☒ False✔

The correct answer is 'False'.

Question 9

🚩 Flag questionMark 1.00 out of 1.00Correct

Optimal action value function is always unique for a task/environment.

Select one:

☒ True✔

☐ False

The correct answer is 'True'.

Question 10

🚩 Flag questionMark 1.00 out of 1.00Correct

Q-learning can be used to estimate discrete action-value function  $Q(x,u)$  in continuing (non-episodic) environments.

Select one:

☒ True✔

☐ False

The correct answer is 'True'.

Finish review



Tuki / Support  
Opiskelijoille / Students

- MyCourses instructions for students
- email: mycourses(at)aalto.fi

Opettajille / Teachers

- MyCourses help
- MyTeaching Support form

Palvelusta

- MyCourses rekisteriseloste
- Tietosuojailmoitus
- Palvelukuvaus
- Saavutettavuusseloste

About service

- MyCourses protection of privacy
- Privacy notice
- Service description
- Accessibility summary

Service

- MyCourses registerbeskrivning
- Dataskyddsmeddelande
- Beskrivning av tjänsten
- Sammanfattning av tillgängligheten

