



# ELEC-E8125 - Reinforcement learning D, Lecture, 4.9.2023-29.11.2023

Syllabus

Started on	Thursday, 14 September 2023, 3:42 PM
State	Finished
Completed on	Thursday, 14 September 2023, 9:06 PM
Time taken	5 hours 23 mins
Grade	10.00 out of 10.00 (100%)

Question 1

🚩 Flag questionMark 1.00 out of 1.00Correct

Consider a 3-element linear world with value function shown in table below.

-16	?	16
-----	---	----

The agent is in the middle state with unknown value. Two actions are available, corresponding to left and right motions. Immediate rewards are zero for both actions when taken in middle state ( $R(\text{middle, left})=R(\text{middle, right})=0$ ) and discount factor is 0.5.

What is the value of the middle state if policy is uniformly random (left and right equally probable)?

Answer: ✔

The correct answer is: 0.00

Question 2

🚩 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 3

🚩 Flag questionMark 1.00 out of 1.00Correct

Value iteration

Consider a 3-element linear world with value function shown in table below.

-2	?	5
----	---	---

In the middle state two actions are available  $U = \{ a, b \}$ . Immediate rewards are zero for both actions when taken in middle state ( $R(\text{middle},a)=R(\text{middle},b)=0$ ) and discount factor is 0.5. Transition probabilities are  $p(\text{left}|\text{middle},a) = p(\text{right}|\text{middle},a) = 0.5$ ,  $p(\text{left}|\text{middle},b)=1$ ,  $p(\text{right}|\text{middle},b)=0$ .

Value iteration has been chosen as the method to find an optimal value function for the middle state. Initial value estimate for the middle state is  $V(\text{middle})=0$ .

What is the updated value of the middle state  $V(\text{middle})$  after one step of value iteration? Update only the value of the middle state.

Answer: ✔

The correct answer is: 0.75

Question 4

🚩 Flag questionMark 1.00 out of 1.00Correct

Do a second iteration of value iteration (only for the middle state). Does the value of the middle state change?

Select one:

☐ True

☒ False✔

That's right. Since the values of the two other states have not changed, value iteration does not change the value.

The correct answer is 'False'.

Question 5

🚩 Flag questionMark 1.00 out of 1.00Correct

**Monte-Carlo estimation for episodic tasks**

Chapter 5.1-5.2

Consider a RL problem, with state space  $X = \{ A, B, C \}$  and action space  $U = \{ a, b \}$ . Discount factor  $\gamma = 1$ .

The following state and action sequences have been obtained by running a policy.

episode 1 (state, action, reward)

$$S_0 = A, A_1 = a, r_1 = 3$$
$$S_1 = B, A_2 = b, r_2 = 2$$
$$S_2 = C, A_3 = a, r_3 = 1$$

episode 2

$$S_0 = A, A_1 = a, r_1 = 4$$
$$S_1 = B, A_2 = a, r_2 = 2$$
$$S_2 = C, A_3 = b, r_3 = 1$$

What is the Monte-Carlo estimate of the state-value function  $V(A)$ ?

Answer: ✔

The correct answer is: 6.50

Question 6

🚩 Flag questionMark 1.00 out of 1.00Correct

What is the Monte-Carlo estimate of the state-value function  $V(B)$ ?

Answer: ✔

The correct answer is: 3.00

Question 7

🚩 Flag questionMark 1.00 out of 1.00Correct

What is the Monte-Carlo estimate of the action-value function  $Q(A,a)$ ?

Answer: ✔

The correct answer is: 6.50

Question 8

🚩 Flag questionMark 1.00 out of 1.00Correct

What is the Monte-Carlo estimate of the action-value function  $Q(B,a)$ ?

Answer: ✔

The correct answer is: 3.00

Question 9

🚩 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  $V(E)=4$ ,  $V(F)=2$ . 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, 3, F, b

Answer: ✔

The correct answer is: 4.00

Question 10

🚩 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  $Q(E,a)=4$ ,  $Q(E,b)=2$ ,  $Q(F,a)=1$ ,  $Q(F,b)=3$ . 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: 2.75

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

