In The Name Of God

HW06

Advanced Neuroscience

MohammadAmin Alamalhoda

97102099

■ Part1 - Model Free RL

□ Explanation on the used Method

I have used the **Model-Free** method for implementing the RL algorithm. The used equations for updating the values of the state are written down in the following:

$$lr_t = \frac{0.1}{1+\log t}$$
$$\delta_t = r_t + \gamma V(S_{t+1}) - V(S_t)$$

Where r_t is:

$$r_t(x,y) = \begin{cases} 2 & \text{if reward in x,y} \\ -2 & \text{if punishment in x,y} \\ 0 & \text{otherwise} \end{cases}$$

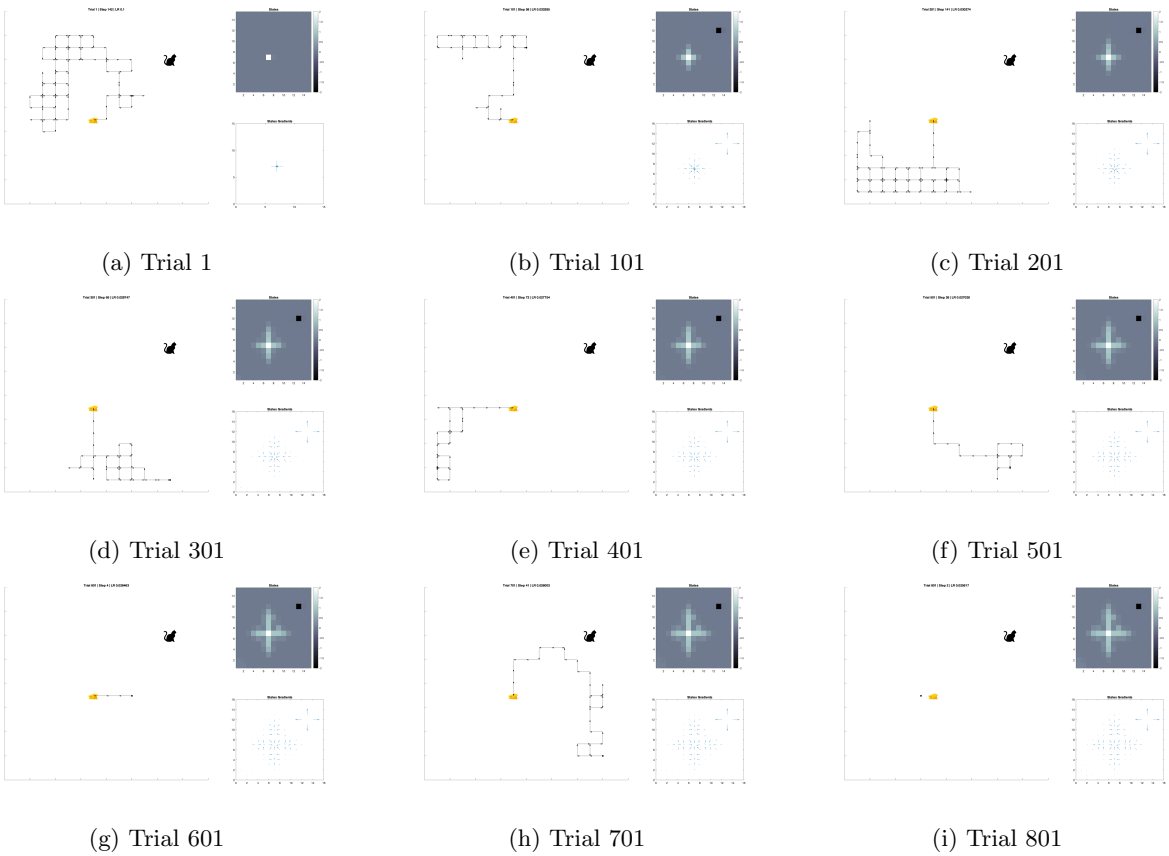
and

$$V(S_t) = V(S_t) + \delta_t lr_t$$
$$m_i = m_i + \delta_t(\delta_i - P(a_i))$$

Where δ_i and $P(a_i)$ are:

$$\delta_i = \begin{cases} 1 & \text{if } a_i \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$
$$P(a_i) = \frac{e^{m_i}}{\sum_{j=1}^{4} e^{m_j}}$$

□ Q01 and Q02



(a) Trial 1 (b) Trial 101 (c) Trial 201

(d) Trial 301 (e) Trial 401 (f) Trial 501

(g) Trial 601 (h) Trial 701 (i) Trial 801

Figure 1: The agent passed way, States values, and States gradients in some of the trials. $LR = \frac{0.1}{1+\log t}$, $\gamma = 1$

As can be seen in Figure 1, the agents learn the value of the states over the trials. It is noteworthy to mention that because my method of implementing the RL is Model-Free, the agent needs more trials to learn the value of the states and the transition probabilities.

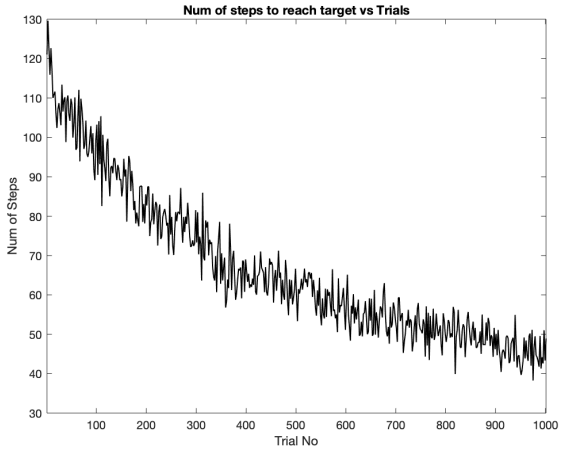


Figure 2: The number of steps to reach the target vs the number of trials of learning. I run the model 200 times and average over all of the vectors of the number of the steps in each of the iterations.

Figure 2 shows the relationship between the number of training trials and the passed steps to reach the target. As can be seen, the passed steps to reach the target decrease over the trials.

☐ Q03

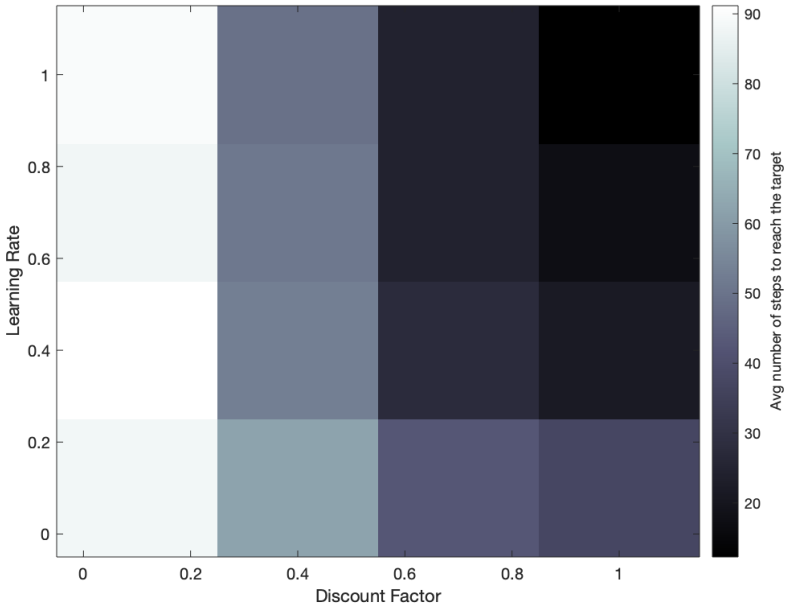


Figure 3: Average of passed steps to reach one target of the last 50 trials for different learning rates and discounting factors

☐ Q04

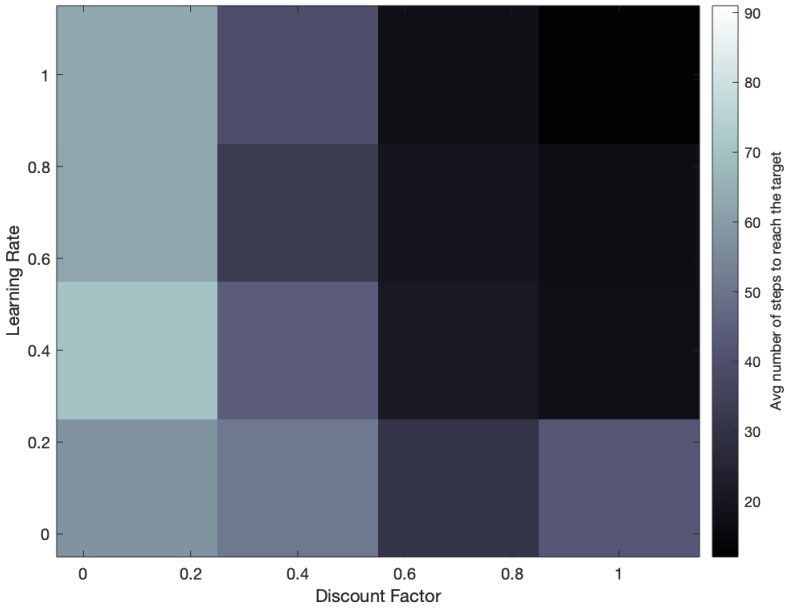


Figure 4: Average of passed steps to reach one of the two targets of the last 50 trials for different learning rates and discounting factors

As can be seen in Figures 3 and 4, bigger learning rates or discount factors tend to decrement in the number of passed steps to reach the target. Also, by adding another target the chance of the agent reaching a target increases, and as can be seen in the Figure 4, the overall number of passed steps to reach one of the targets has been decreased in comparison with Figure 3.

■ Part2 - Model Free RL using TD rule

□ Explanation on the used Method

I have used the **Model-Free** method of the TD rule for implementing the RL algorithm. The used equations for updating the values of the state are written down in the following:

$$lr_t = \frac{0.1}{1+\log t}$$
$$\delta_t = r_t + \gamma V(S_{t+1}) - V(S_t)$$
$$\lambda = 0.99$$

Where r_t is:

$$r_t(x,y) = \begin{cases} 2 & \text{if reward in x,y} \\ -2 & \text{if punishment in x,y} \\ 0 & \text{otherwise} \end{cases}$$

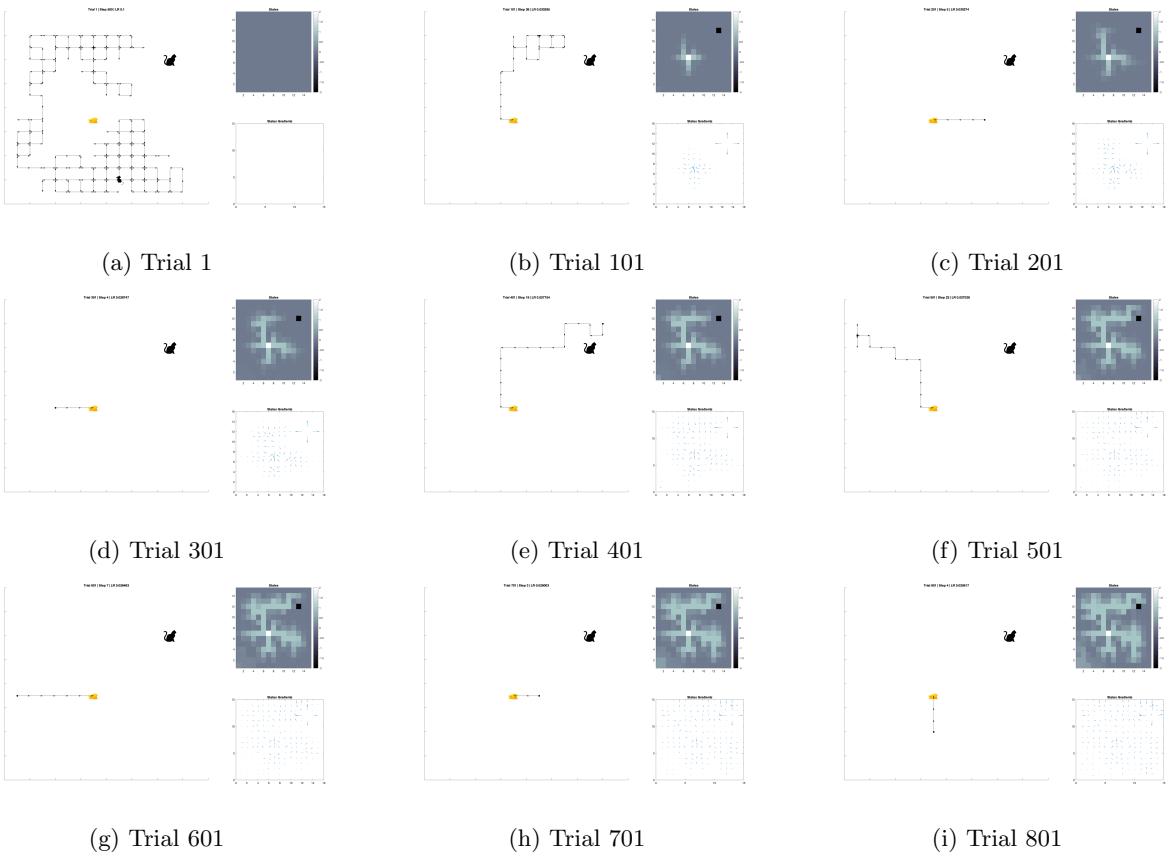
and

$$V(S_{t-t'}) = V(S_{t-t'}) + \lambda^{t'}\delta_t lr_t,\ \forall t' = 1,2,...$$
$$m_{i_(S_{t'})} = m_{i_(S_{t'})} + \lambda^{t'}\delta_t(\delta_i - P(a_i))$$

Where δ_i and $P(a_i)$ are:

$$\delta_i = \begin{cases} 1 & \text{if } a_i \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$
$$P(a_i) = \frac{e^{m_i}}{\sum_{j=1}^{4} e^{m_j}}$$

□ Q05



(a) Trial 1 (b) Trial 101 (c) Trial 201

(d) Trial 301 (e) Trial 401 (f) Trial 501

(g) Trial 601 (h) Trial 701 (i) Trial 801

Figure 5: TD Rule - Agent passed way, States values and States gradients in some of the trials. $LR = \frac{0.1}{1+\log t}$, $\gamma = 1$, $\lambda = 0.99$

As can be seen in Figure 9, the agents learn the value of the states over the trials. By comparing Figures 1 and 9 we can infer that the TD rule can learn the values of the states faster. This result is more obvious in Figure 6.

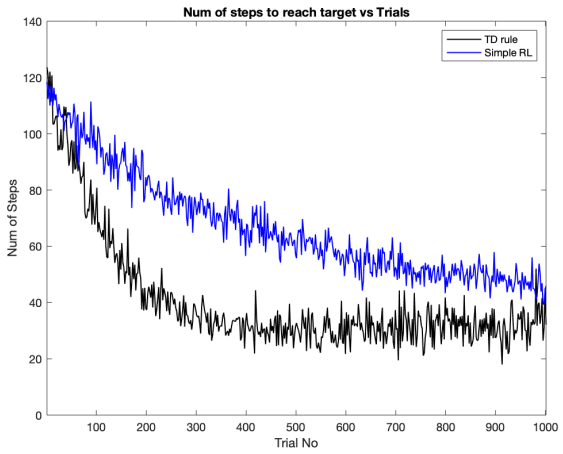


Figure 6: TD-Rule - Number of steps to reach the target vs the number of trials of learning. I run the model 200 times and average over all of the vectors of the number of the steps in each of the iterations.

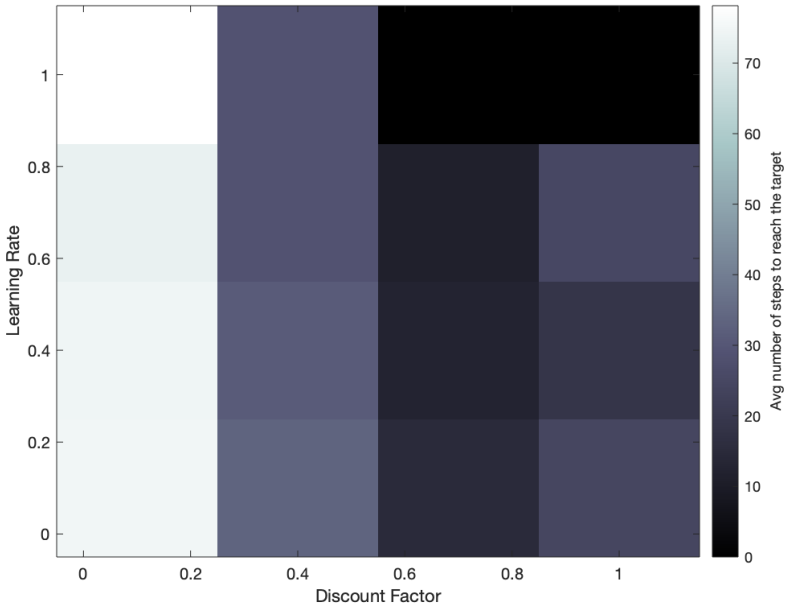
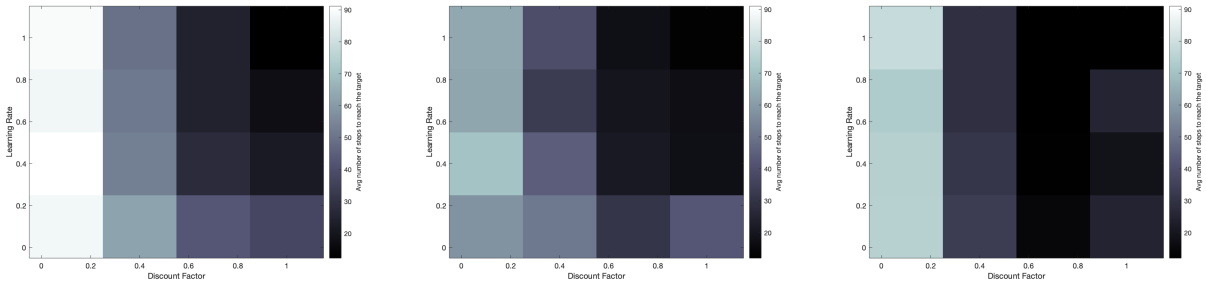


Figure 7: TD-Rule - Average of passed steps to reach one target of the last 50 trials for different learning rates and discounting factors

As can be seen in Figure 7, bigger learning rates or discount factors tend to decrement in the number of passed steps to reach the target. In the following, you can see the effect of the learning rate and discounting factor on the number of passed steps to reach the target for different RL rules and conditions.

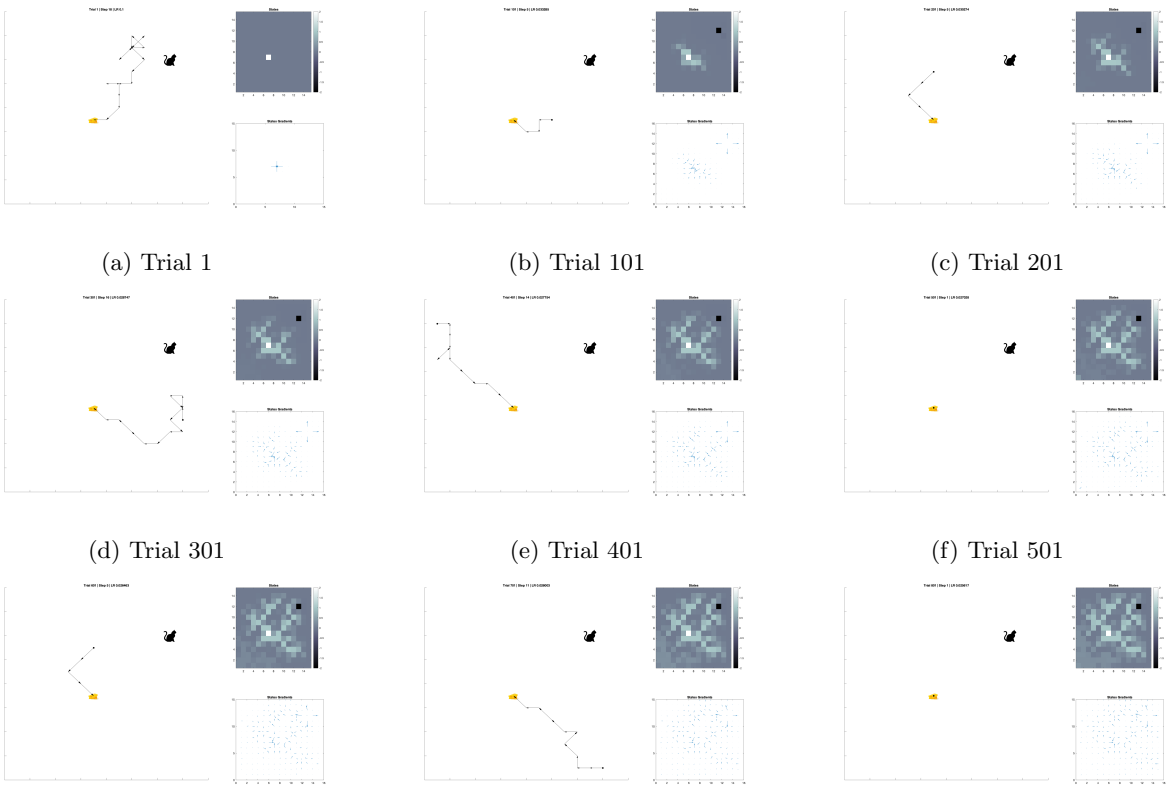


(a) Simple RL - 1 Target (b) Simple RL - 2 Targets (c) TD-Rule - 1 Target

Figure 8: Average of passed steps to reach one target of the last 50 trials for different learning rates and discounting factors during different conditions

TD-rule needs very lower trials to reach the target in comparison with the simple RL method and even simple RL with 2 targets! (Figure 8).

□ TD-Rule with 8 Directions



(a) Trial 1 (b) Trial 101 (c) Trial 201

(d) Trial 301 (e) Trial 401 (f) Trial 501

(g) Trial 601 (h) Trial 701 (i) Trial 801

Figure 9: TD-Rule with 8 directions - Agent passed way, States values and States gradients in some of the trials. $LR = \frac{0.1}{1+\log t}$, $\gamma = 1$, $\lambda = 0.99$

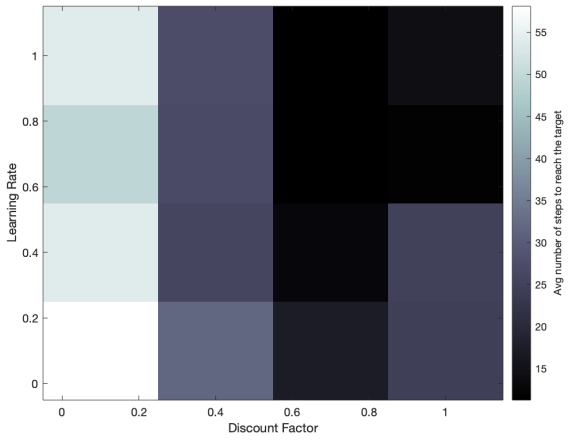
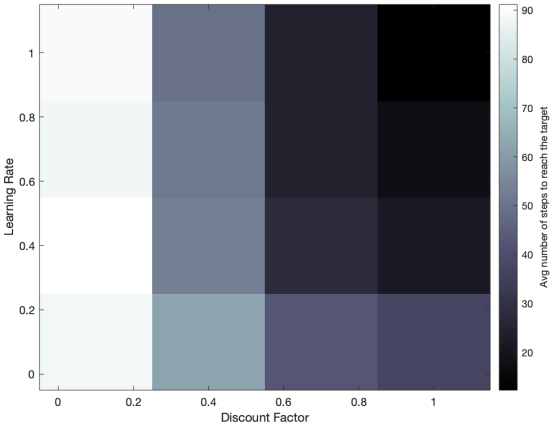
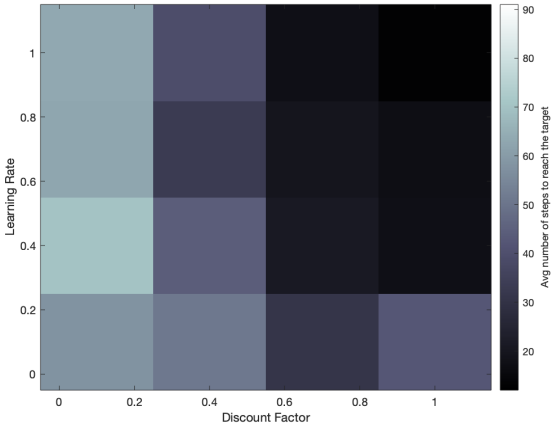


Figure 10: TD-Rule with 8 directions - Average of passed steps to reach one target of the last 50 trials for different learning rates and discounting factors

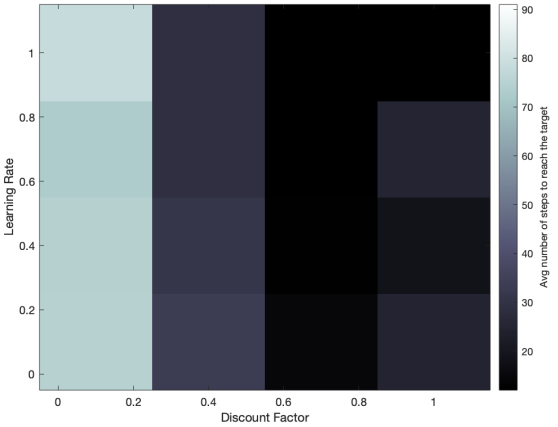
In Figure 11 you can see the effect of the learning rate and discounting factor on the number of passed steps to reach the target for different RL rules and conditions including the TD-rule with 8 directions.



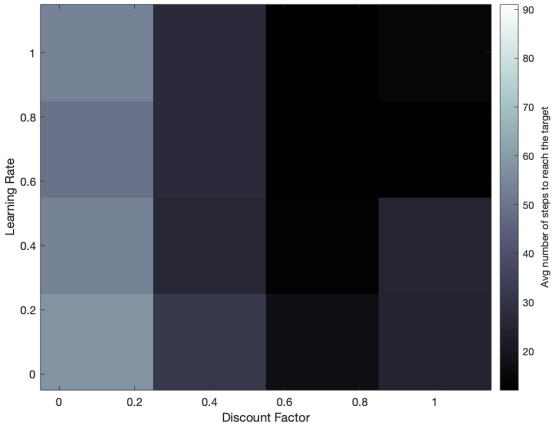
(a) Simple RL - 1 Target



(b) Simple RL - 2 Targets



(c) TD-Rule - 1 Target - 4 Direction



(d) TD-Rule - 1 Target - 8 Directions

Figure 11: Average of passed steps to reach one target of the last 50 trials for different learning rates and discounting factors during different conditions