TBMI26 – Computer Assignment Reports  
Reinforcement Learning

Deadline – March 15 2019

Authors: Anubhav Dikshit <anudi287>

Hector Plata <hecpl268>

In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. **You will also need to upload all code in .m-file format**. We will correct the reports continuously so feel free to send them as soon as possible. If you meet the deadline you will have the lab part of the course reported in LADOK together with the exam. If not, you’ll get the lab part reported during the re-exam period.

1. **Define the V- and Q-function given an optimal policy. Use equations and describe what they represent. (See lectures/classes)**
2. **Define a learning rule (equation) for the Q-function and describe how it works. (Theory, see lectures/classes)**
3. **Briefly describe your implementation, especially how you hinder the robot from exiting through the borders of a world.**
4. **Describe World 1. What is the goal of the reinforcement learning in this world? What parameters did you use to solve this world? Plot the policy and the V-function.**
5. **Describe World 2. What is the goal of the reinforcement learning in this world? What parameters did you use to solve this world? Plot the policy and the V-function.**
6. **Describe World 3. What is the goal of the reinforcement learning in this world? What parameters did you use to solve this world? Plot the policy and the V-function.**
7. **Describe World 4. What is the goal of the reinforcement learning in this world? How is this world different from world 3, and why can this be solved using reinforcement learning? What parameters did you use to solve this world? Plot the policy and the V-function.**
8. **Explain how the learning rate α influences the policy and V-function in each world. Use figures to make your point.**

1. **Explain how the discount factor γ influences the policy and V-function in each world. Use figures to make your point.**
2. **Explain how the exploration rate ε influences the policy and V-function in each world. Use figures to make your point. Did you use any strategy for changing ε during training?**

1. **What would happen if we instead of reinforcement learning were to use Dijkstra's cheapest path finding algorithm in the ''Suddenly Irritating blob'' world? What about in the static ''Irritating blob'' world?**
2. **Can you think of any application where reinforcement learning could be of practical use? A hint is to use the Internet.**
3. **(Optional) Try your implementation in the other available worlds 5-12. Does it work in all of them, or did you encounter any problems, and in that case how would you solve them?**