Exam for course 5

Question 1 (G-5.1) : Exploration / Exploitation

2

Question 2 (G-5.2) : Supervised versus Reinforcement

1. reinforcement
2. reinforcement
3. reinforcement

Question 3 (G-5.3) : The reward hypothesis

3

Question 4 (B-5.1) : Gamma

1, 2, 4

Question 5 (B-5.2) : Policy and Value function

The policy function associates the state of the agent with an action, while the value function associates the state of the agent with the prediction of cumulated future rewards, weighted by γ.

Question 6 (B-5.3) : Q-learning

In Q-learning, we aim to find V^{π^*} as a solution to the recursive system of equations (Bellman equation) :

$$\forall s \in S^\alpha, \forall a \in A, Q(s,a) = r_{s,a} + \gamma \max_{a'} Q(s(a), a'),$$

where $r_{s,a}$ is the reward agent α performs action a in state s and $s(a)$ is the state observed by agent α after performing action a.