ULiège
Prof. Damien Ernst
TA : Samy Aittahar

INFO8003-1
Optimal decision making
for complex problems
9th April 2018

Examination

Theoretical examination 4

1 QUESTIONS

1. Describe the online Q-iteration algorithm. What are the main issues of this algorithm? Explain how to overcome them.
2. Explain the overestimation problem in Q-learning. Propose an approach to avoid it.
3. Describe at least three approaches to use the Q-learning with problems having continuous action spaces. Discuss their pros/cons.
4. Provide the direct policy differentiation formula, and derive the corresponding reinforcement learning algorithm. How does it relate to maximum likelihood?
5. Explain drawbacks of the vanilla policy gradient, and propose approaches to overcome them.
6. Why is policy gradient an on-policy method? How to derive the off-policy variant? Derive the policy gradient in this variant.
7. Describe the K-armed bandit setting. Explain how to assess the performance of a given strategy in this setting.
8. How does a K-armed bandit problem differs from a classical reinforcement learning problem in an MDP?
9. Describe at least three approaches which address the exploration-exploitation problem in the K-armed bandit setting and discuss their drawbacks.
10. Describe the *Upper Confidence Bound* (*UCB*), and explains how it follows the *optimism in the face of uncertainty* principle.
11. When does the *UCB* algorithm should switch from exploration to exploitation? Provide and prove bounds.
12. Describe the *Upper Confidence Trees* (*UCT*) algorithm, and explain why, despite the fact that *UCT* is consistent with respect to *UCB*, it may demonstrate poor performance in practice.