ULiège
Prof. Damien Ernst

INFO8003-1
Optimal decision making
for complex problems

Pratical session

Reconstructing an MDP

QUESTION 1

You work in an automobile factory as a software engineer. Your company has merged with your main competitor. As a consequence, you receive their turbo production machine. This new machine was custom-made by your concurrent's former engineers and the documentation is very sparse. What you get from the documentation is that (i) there are 4 states in the machine, s_1, s_2, s_3 and s_4 and (ii) there are two buttons in the machine which increase or decrease pressure by a fixed amount, respectively. The performance metrics of your machine is the number of pieces produced by hour. A change of state lasts one hour. You are asked to do a series of experiments to better understand the machine. More specifically, your task is to compute an estimate of the probability of transitioning to each state given the previous state action pair and an estimate of the reward you get per state action pair. The output of your simulations are the one-step trajectories bellow :

- s_1, INCREASE, one turbo, s_2
- s_1, INCREASE, two turbos, s_3
- s_1, INCREASE, one turbo, s_2
- s_1, INCREASE, one turbo, s_2
- s_3, INCREASE, two turbo, s_4
- s_3, INCREASE, two turbo, s_4
- s_4, INCREASE, zero turbo, s_1
- s_2, INCREASE, one turbo, s_1
- s_1, DECREASE, zero turbo, s_1
- s_1, DECREASE, zero turbo, s_1

- s_2, DECREASE, five turbos, s_4
- s_2, DECREASE, zero turbo, s_3
- s_3, DECREASE, four turbo, s_3
- s_3, DECREASE, four turbo, s_3
- s_3, DECREASE, zero turbo, s_2
- s_4, DECREASE, zero turbo, s_1
- s_4, DECREASE, zero turbo, s_3

QUESTION 2

Compute the policy μ_2^* over 2 timesteps starting at state s_0 with $\gamma = 0.95$.

QUESTION 3

Provide a bound on the suboptimality of μ_2^* with respect to μ^*. Is the bound a good one? Compute the value of N such that this bound is equal to 0.1.