In your report for Part 2:

1. Show and discuss the results after:
   * Applying Q-learning to solve the deterministic environment defined in Part 1. Plots should include epsilon decay and total reward per episode.

Epsilon: 0.9, Gamma: 0.6, lr : 0.5

Shape

Description automatically generatedA picture containing icon

Description automatically generated

Applying Q-learning to solve the stochastic environment defined in Part 1. Plots should include epsilon decay and total reward per episode.

Epsilon: 0.7, Gamma: 0.6, lr : 0.5

Shape

Description automatically generatedA picture containing antenna, chime

Description automatically generated

* + Applying any other algorithm of your choice to solve the deterministic environment defined in Part 1. Plots should include total reward per episode.

Shape

Description automatically generatedA picture containing chart

Description automatically generated

* + Applying any other algorithm of your choice to solve the stochastic environment defined in Part 1. Plots should include total reward per episode.

Timeline

Description automatically generated

Shape

Description automatically generated

* + Provide the evaluation results. Run your environment for at least 10 episodes, where the agent chooses only greedy actions from the learnt policy. Plot should include the total reward per episode.

1. Compare the performance of both algorithms on the same deterministic environment (e.g. show one graph with two reward dynamics) and give your interpretation of the results. Chart, box and whisker chart

   Description automatically generated
2. Compare how both algorithms perform in the same stochastic environment (e.g. show one graph with two reward dynamics) and give your interpretation of the results.

A picture containing text, pencil, writing implement, chime

Description automatically generated

1. Briefly explain the tabular methods, including Q-learning, that were used to solve the problems. Provide their update functions and key features.