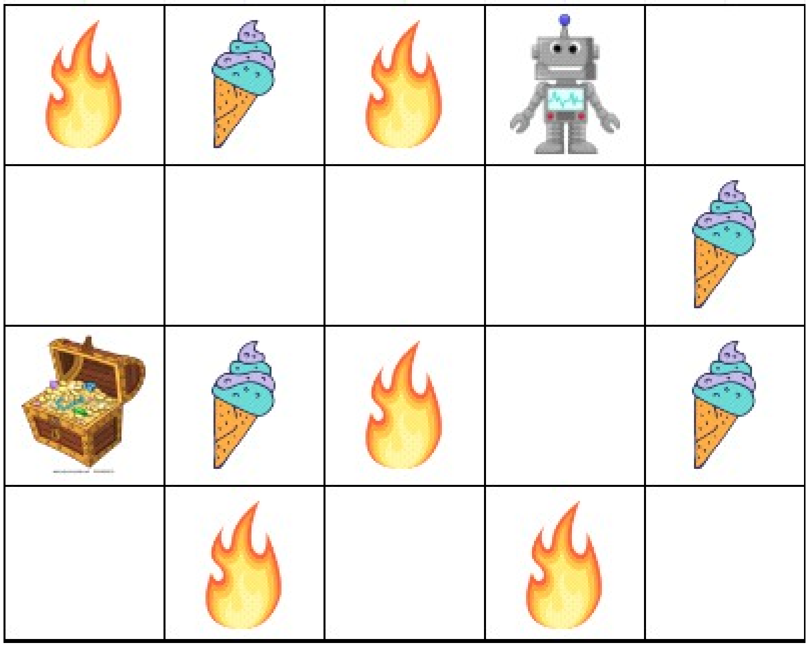
Machine Learning
Coursework 2
Mattia Gennaro

1 - The Domain



The domain is made by 20 states.

The goal is represented by the chest, and gives a reward of 200 points.

It presents two different types of reward:

- The fire, which will give a negative reward of -5 points.
- The ice cream, which will give a reward of 5 points.

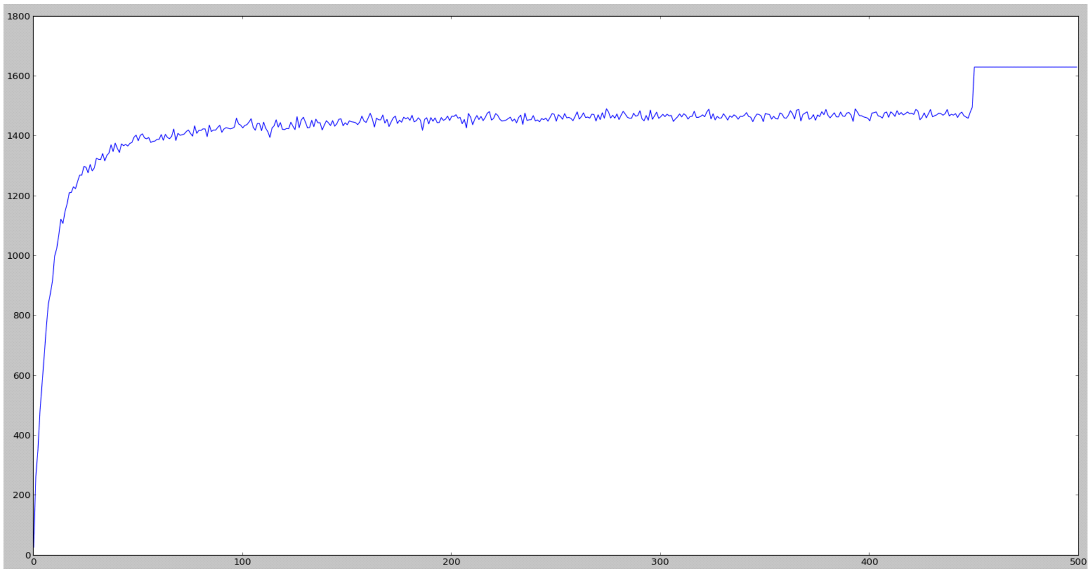
2 – Experiments

All the experiments are done using 500 for Epochs and Episodes, to have a smoother plot, and so that the lack of convergence is evident

Experiment **1:**

- epsilon = 0.1
- alpha = 0.1
- Move: deterministic
- Learning: Sarsa

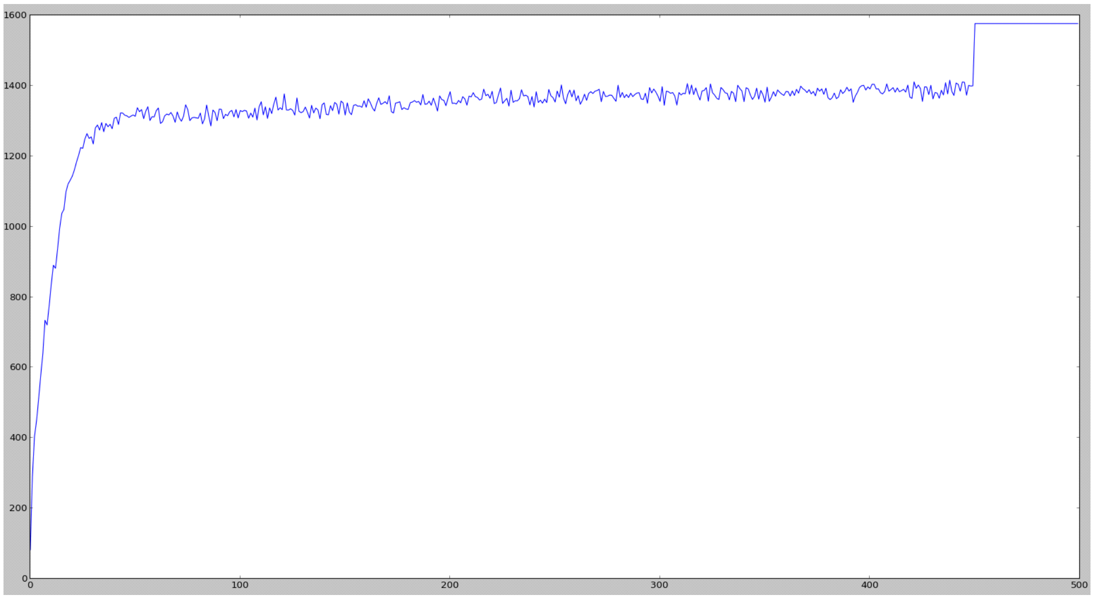
Result:



Experiment **2**:

- epsilon = 0.1
- alpha = 0.1
- Move: deterministic
- Learning: q

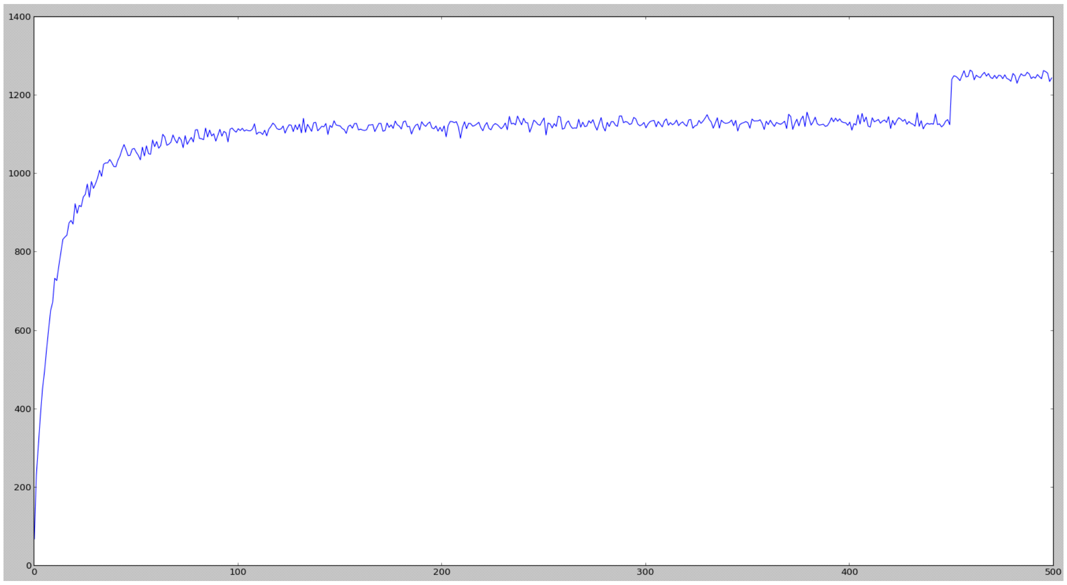
Result:



Experiment **3**:

- epsilon = 0.1
- alpha = 0.1
- Move: stochastic
- Learning: Sarsa

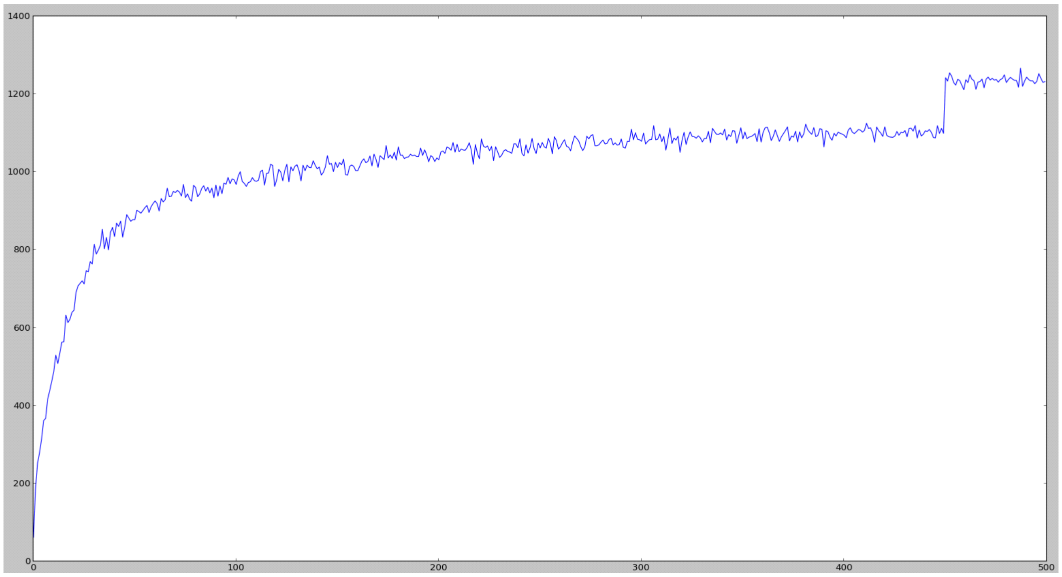
Result:



Experiment **4**:

- epsilon = 0.1
- alpha = 0.1
- Move: stochastic
- Learning: q

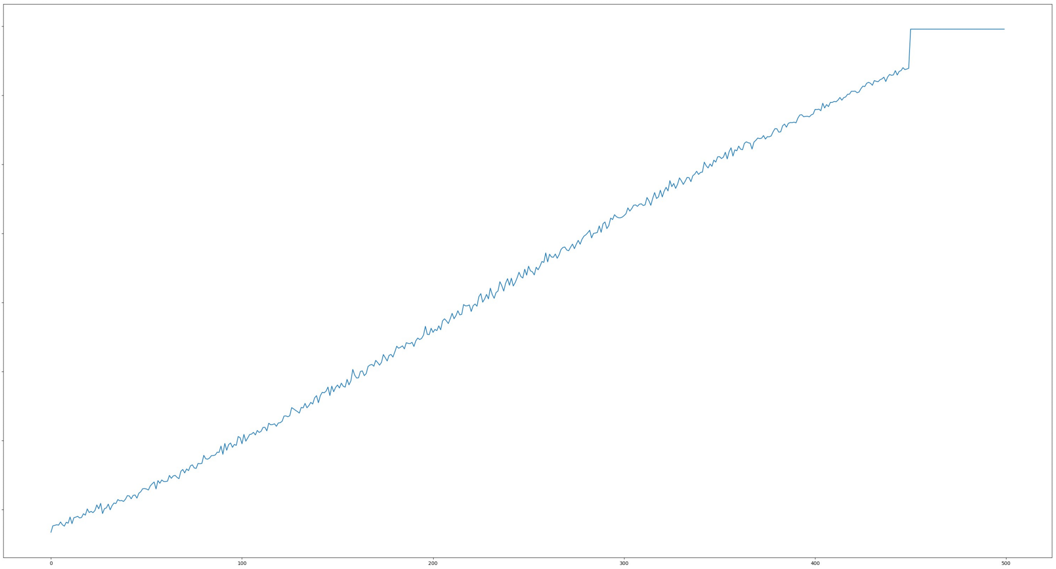
Result:



Experiment **5**:

- epsilon = linear
- alpha = 0.05
- Move: deterministic
- Learning: Sarsa

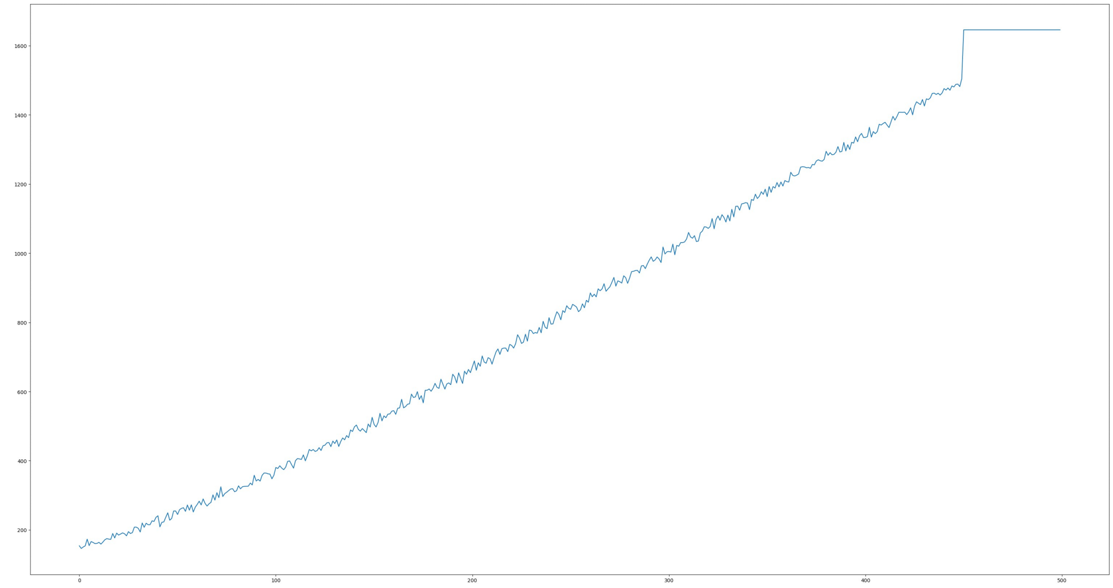
Result:



Experiment **6**:

- epsilon = linear
- alpha = 0.05
- Move: deterministic
- Learning: q

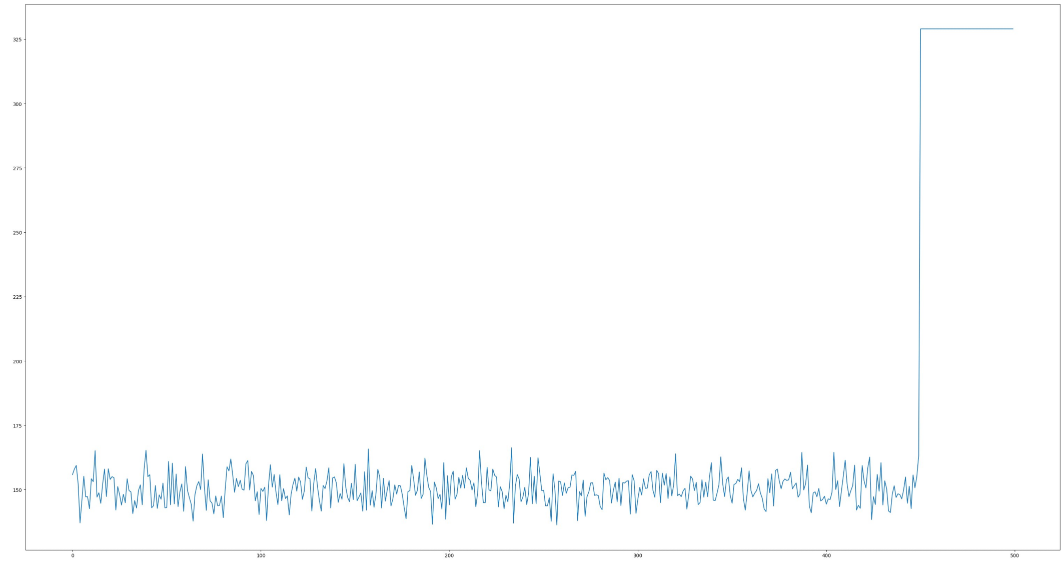
Result:



Experiment **7:**

- epsilon = 1
- alpha = 1
- Move: deterministic
- Learning: Sarsa

Result:

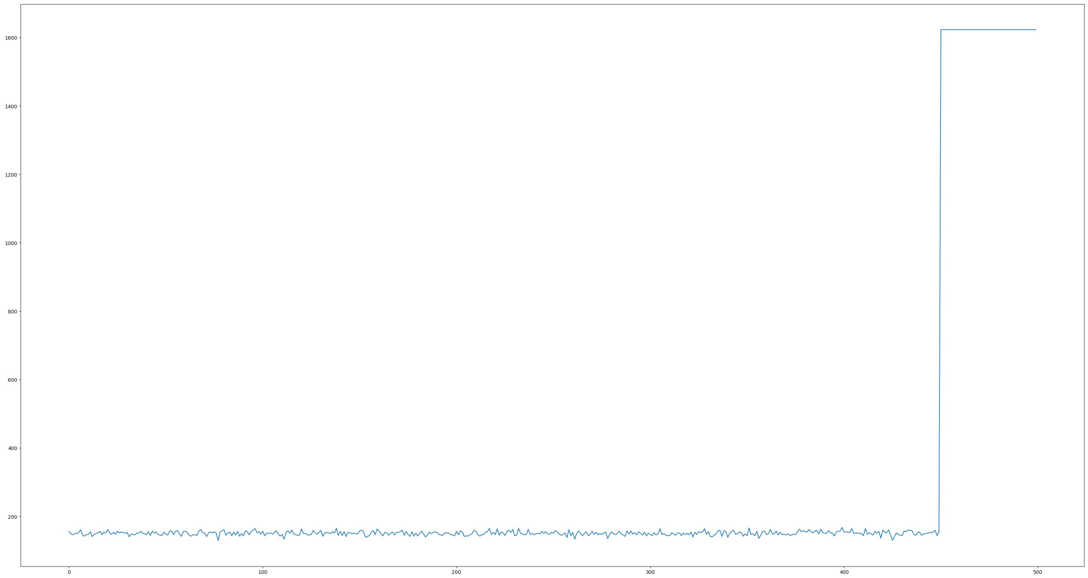


Experiment **8**:

- epsilon = 1
- alpha = 1
- Move: deterministic

- Learning: q

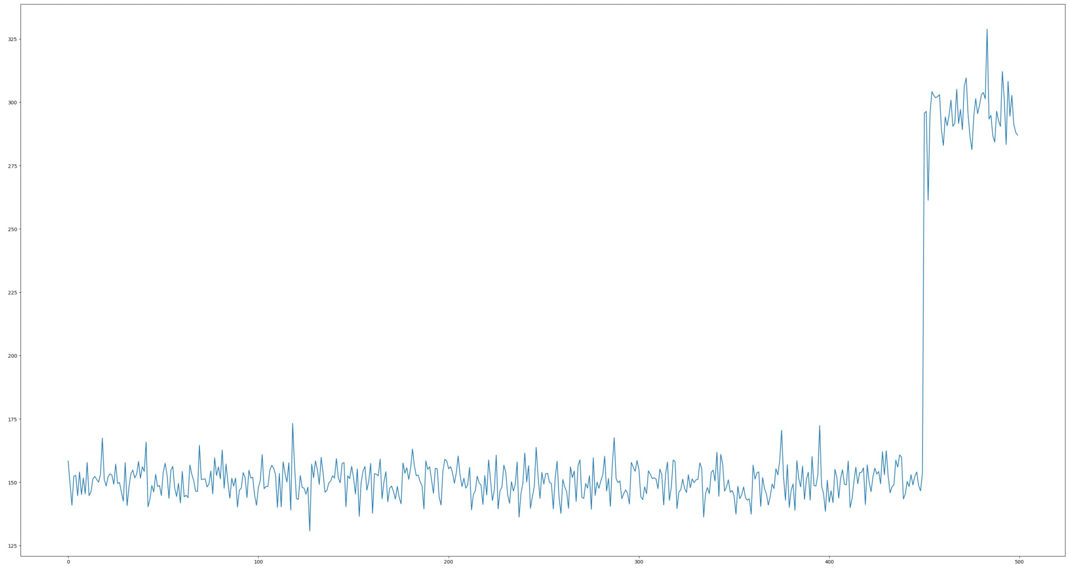
Result:



Experiment **9**:

- epsilon = 1
- alpha = 1
- Move: stochastic
- Learning: Sarsa

Result:

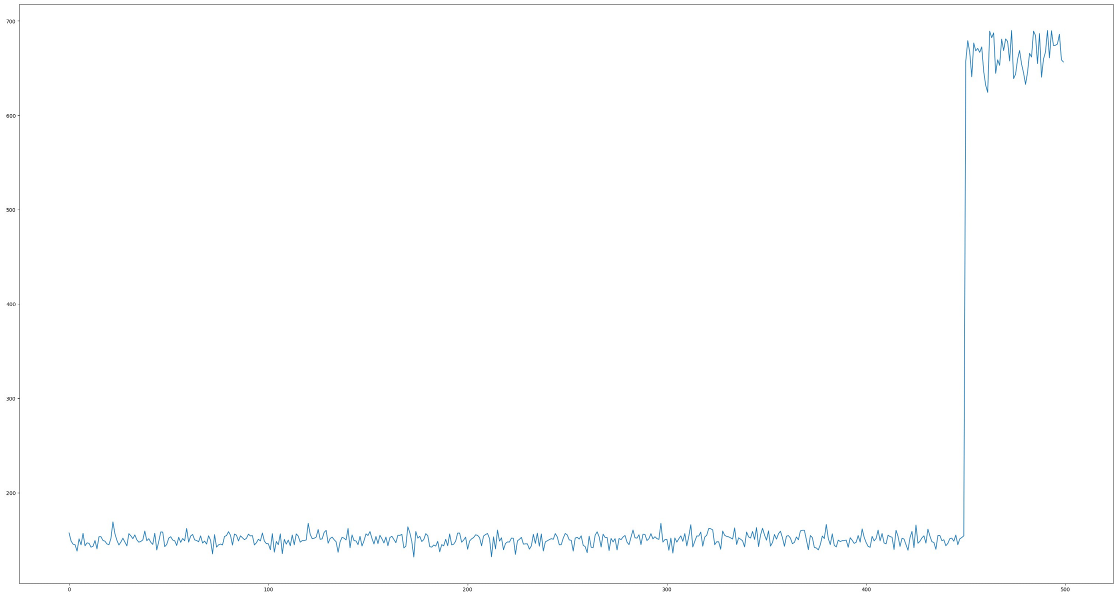


Experiment **10**:

- epsilon = 1
- alpha = 1

- Move: stochastic
- Learning: q

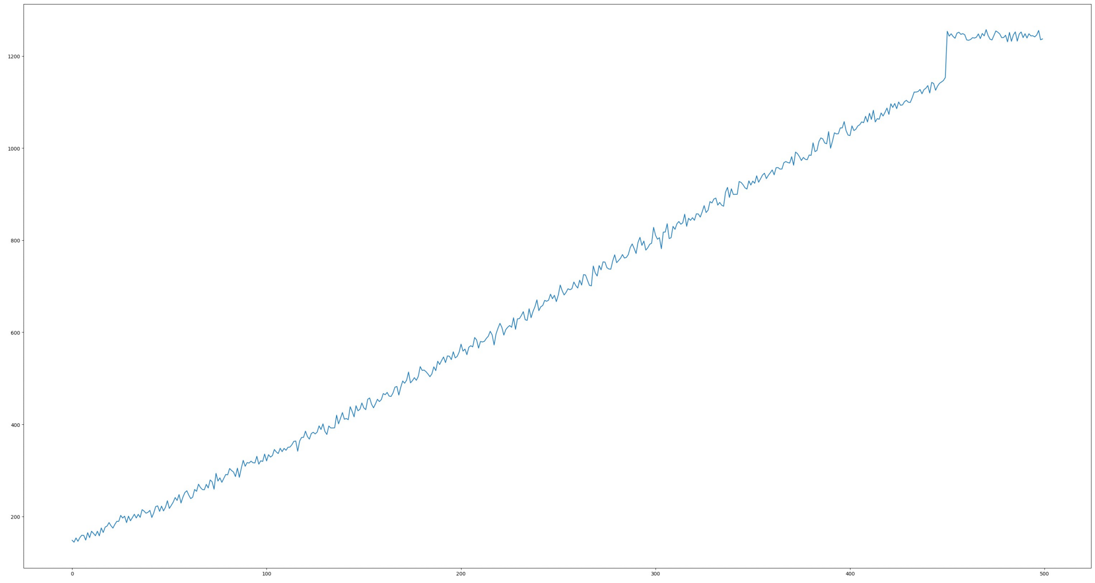
Result:



Experiment **11**:

- epsilon = linear
- alpha = 0.05
- Move: stochastic
- Learning: Sarsa

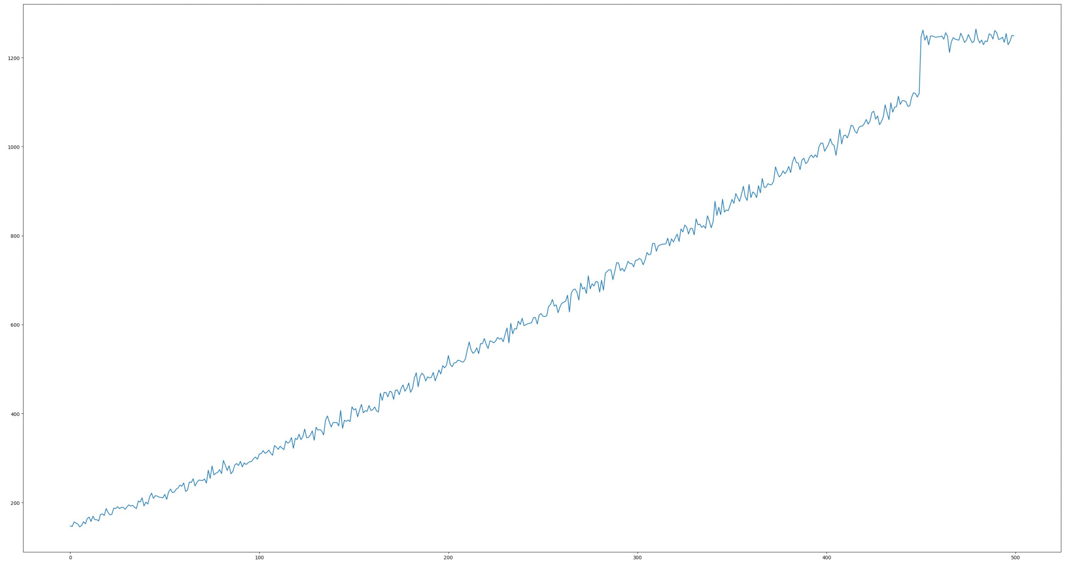
Result:



Experiment **12**:

- epsilon = linear
- alpha = 0.05
- Move: stochastic
- Learning: q

Result:



i)The quality of the learned policy is really high, and could not improve more because there is no better one.
ii)The effects of alpha and epsilon on Sarsa and q are really high, since they are both used as parameters in both functions.
Alpha is the learning rate, and it's passed directly, while epsilon is used to choose an agent, which will be passed to both functions as well.
With an higher value of alpha, we will have smaller values in qValues, with an higher value of epsilon, we will have the agent to choose the biggest value inside qValues, almost every time, avoiding random choices.
iii)