Programming Assignment 2

**Part 1**

Q1:

1. The global consistency problem is that the association of metric coordinates to place cell activation involves using self-motion, which is inherently relative to an origin point. However, changes in the origin position would mean that the coordinates learnt are relative to different positions and thus inconsistent in the environment.
2. Any navigational model that attempts to directly associate place cells with metric values representing locations in the environment. Self-motion estimates are used to integrate a path from an origin, which is how an agent would know where it is. However, this origin is relative to where it started, and its view of the coordinates is tied to its motion from that origin point. Thus, if it is picked up and moved to a different location it no longer is where its path integration information suggests it is, and any path integration done thereafter would be inconsistent.

Q2:

1. The distal reward problem is that in most cases the reward is far away from the origin of the task, and thus the origin and its vicinity will not be associated to the reward like the configurations closer to the goal.
2. A navigation model that represents rewards using place cells would suffer from this. This is because even though the place cells in the immediate proximity of the reward could be associated with it, any other place cell that is far away from it would not. Considering that odds are an agent would start at a place cell far from the reward, it would not know where to go as the place cells there themselves do not know of the distant reward.

Q3:

1. The role of the actor is to determine how the agent should act. That is, it encodes the policy that the agent uses to select the direction in which it should move next based on the activation pattern of the place cells.
2. The role of the critic is to criticize and instruct the agent whether the action taken by it was good or bad relative to the task at hand. It does so by encoding the estimated value of where the agent is based on the activation pattern of the place cells.
3. The error signal is

where is the reward perceived by the agent at time *t*, is the constant discounting factor, and and are the values estimated by the critic of the positions of the agent at time *t+1* and *t* respectively.

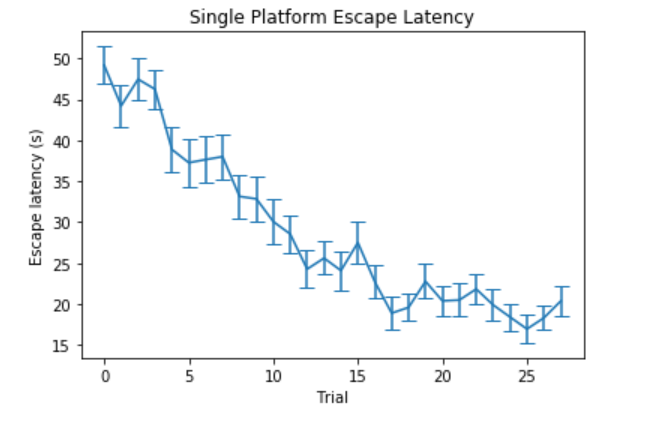
1. The place cells encode the position of the agent, forming a basis function representation of location. Their encoding of location is then used to inform the critic and actor of the location so that they may calculate their values.
2. The place cells do not encode any information about distance, direction or any other spatial or navigational information that could be used by the agent to know how to act.

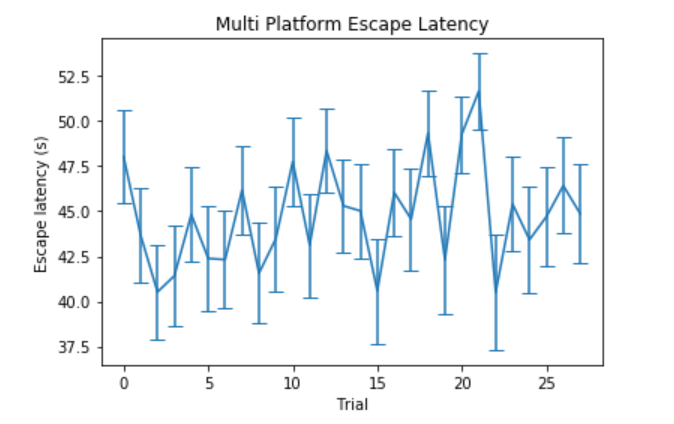
Q4:

1. TD learning solves the distal reward problem by encoding a continuous reward function using place cell activities. This means that a place cell is associated with a value that represents the expected return that could be obtained by moving from that place, essentially anticipating the reward from neighbouring. This predictive conditioning of place cells anticipating the expected reward in their vicinity allows the notion of a reward to reach distant place cells, even though they are not directly in the vicinity of the reward itself. The agent then at the other edge of the environment will therefore still know how to move since it can anticipate which direction would provide it the highest expected return.
2. The main information the rat needs to be able to accurately have is its own self-motion estimates, which involves also having a head direction. This is needed because the entire model relies on these estimates being accurate to make the global coordinates being learned converge to a consistent system. Whether this is realistic or not depends on the conditions the rat is in. The rats need to be able to calculate this heading in a consistent way, and place cells do not encode these kinds of metrics. Therefore, if the rat has a reliable way of calculating this information through other means, such as visual and vestibular cues, then they use utilize it to inform the coordinate model. In the absence or removal of the information or capabilities they must calculate this information, such as possibly if they were disoriented or had some of their other systems such as the vestibular system compromised, then learning a global consistent coordinate system would also be compromised.

**Part 2**

Q2: With a single platform location the model does successfully reduce the latency over trials. The experiment was conducted by running 4 trials a day for 7 days, totaling in 28 separate trials. The experiment was repeated 50 times, and the figure below shows the mean and standard errors over those 50 iterations for each trial, with a clear downward trend to the escape latencies. After trial 17, the behaviour seems to be asymptotic.

****

Q3: Running the same experiment as described above, if we change the locations of the platform every day, then the escape latencies do not seem to reduce over trials. This is because although our actor critic model is good at the distal reward problem, the values it learned on previous days interfere with the ones it would like to learn for the new platform. Another issue is that it does not seem to be able to use its previous days’ knowledge of the environment to quickly learn the new platform position. Both issues are less about the distal reward problem, which the Actor-Critic model solves, and more about the global consistency problem which is not yet implemented in this model.

**Part 3**

Q2: The model does successfully reduce the latency over the trials in the case of multiple platform locations. The same experiment as in part 2 is conduction where each day sees the rat trying to find a platform at a novel location. The graph below shows the escape latencies of the combined model in the DMP task.

Q3: