Assignment 5.1

Q.1 What is reinforcement learning & why is it called so? Can we use Reinforcement Learning (RL) to detect facial emotions?

Reinforcement Learning (RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. Reinforcement learning is called so because the "reinforcement" in reinforcement learning refers to how certain behaviors are encouraged, and others discouraged. Behaviors are reinforced through rewards which are gained through experiences with the environment. The focus of RL methods is essentially to learn a policy for robot behaviour i.e. learning to interact with the environment in a way that maximises expected returns. Facial emotion detection is a supervised learning (classification) problem which requires a somewhat different methodology. Due to this reason, Reinforcement Learning might not be a well-suited methodology for detecting facial emotions

Q.2 What is the goal of exploitation and what is the goal of exploration?

In simple words, exploitation refers to the building of capabilities and resources to manage the present (short-term organizational goals and stay ahead of current competition). Exploitation basically exploits the agent's current estimated value and chooses the greedy approach to get the most reward. However, the agent is being greedy with the estimated value and not the actual value, so chances are it might not get the most reward. Exploration refers to the building of capabilities and resources for managing the future (strategic goals and stay ahead of future. Exploration is more of a long-term benefit concept where it allows the agent to improve its knowledge about each action which could lead to long term benefit.

Q.3 What is greedy agent? Does a greedy agent always find an optimal policy? Explain it.

A greedy agent means the Agent constantly performs the action that is believed to yield the highest expected reward. Obviously, such a policy will not allow the Agent to explore at all. In order to still allow some exploration, an ε-greedy policy is often used instead: a number (named ε) in the range of [0,1] is selected, and prior selecting an action, a random number in the range of [0,1] is selected. if that number is larger than ε, the greedy action is selected — but if it's lower, a random action is selected. Note that if ε=0, the policy becomes the greedy policy, and if ε=1, always explore. There is always at least one such optimal policy. The so called greedy policy is following the currently best path of actions. During learning however, for the values to converge into good estimates it is required that the agent visits all available states to gain information about them.

Q.4 What is reward and return in bellman equation?

Reward(R): treat which the agent gets after performing an action(a).

R(s): reward for being in the state s

R(s,a): reward for being in the state and performing an action a

R(s,a,s'): reward for being in a state s, taking an action a and ending up in s'

Return

$R_t = \sum_{k=0}^{T} \Upsilon^{k} r_{t+k+1}$ where ϒ can be 1 only if a zero reward absorbing state is always reached.

Q.5 Explain Monte Carlo method in brief.

Any method which solves a problem by generating suitable random numbers, and observing that fraction of numbers obeying some property or properties, can be classified as a Monte Carlo method The Monte Carlo method for reinforcement learning learns directly from episodes of experience without any prior knowledge of MDP transitions. Here, the random component is the return or reward. One caveat is that it can only be applied to episodic MDPs. It's fair to ask why, at this point. The reason is that the episode has to terminate before we can calculate any returns. Here, we don't do an update after every action, but rather after every episode. It uses the simplest idea – the value is the mean return of all sample trajectories for each state.