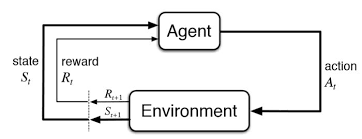
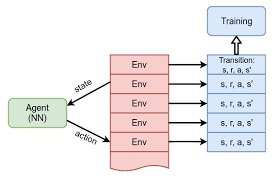
**Explain how the cartpole problem can be solved using the REINFORCE algorithm**.

When looking at a problem such as the cartpole and the way that it learns as you run the program. This is the perfect placement for the Reinforce algorithm, this is a goal oriented type of algorithm that makes decisions based on the ability to receive reward. Similar to teaching kids to do chores using the reward of an allowance. Besides the reward this algorithm runs with the sole purpose of solving a problem, solving the problem and getting the reward goes hand in hand. Looking at the cartpole problem there are two rewards that I can identify. The first is the reward given for when the pole's location is close to 0. The second is given when the pole moves up, aka the direction we want it in. The decision maker or the Agent is going to need to lift the pole as the pole wants to fall towards the ground. All of these factors over time, the algorithm will learn to keep the pole upright for extended periods of time. Below is a diagram of how the reinforce learning algorithm works.



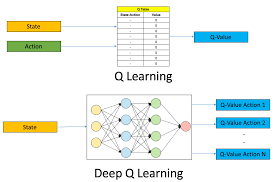
There is a reward given. In this case the user is the deciding factor in the actions taken to be successful in maintaining an upward position of the pole. The critic in this case is the thing that will tell the actor what is the most effective way of accomplishing the goal through suggestions on ways that the actor caLooking at this the breakdown of how it works is kind of simple. Setting the rules is the first step, then we use those rules to lay out the steps, followed by calculating points and reward. Make adjustments as things are learned and repeat starting at step 2.

**Explain how the cartpole problem can be solved using the A2C algorithm.** Another way that the cartpole can be approached is using the Advantage Actor Critic or the A2C. This takes two types of reinforced learning algorithms and combines them, the only difference here is that the goal and overall solution are different. Here we know that the problem is the pole being attached to a cart that moves without friction, the pole stands on it straight up and down. The End goal is to keep the pole from falling as the cart moves -1 or +1, each time this is successfully improved through movement. Below is a diagram explaining the process of the A2C.



**Explain how policy gradient approaches differ from value-based approaches, such as Q-learning Multiple**

With this there are very apparent in the above methods they are value based types of learning algorithms and the gradient methods. When looking at Q-learning it is in most cases used for things dealing with different value based approaches. But it is a method that is generally based on estimating the Q-function. This can truly maximize the best action and also rewards inside of the space and is generally used in high dimensional spaces or in a situation where actions are continuous throughout the space.



Sharma, S. (2020, June 13). *The ultimate beginner's Guide to Reinforcement Learning*. Medium. Retrieved August 6, 2022, from <https://towardsdatascience.com/the-ultimate-beginners-guide-to-reinforcement-learning-58>

8Packt subscription. (n.d.). Retrieved August 6, 2022, from https://subscription.packtpub.com/book/data/9781788834247/11/ch11lvl1sec43/adding-an-

*Deep Q-learning: An introduction to deep reinforcement learning*. Analytics Vidhya. (2020, April 27). Retrieved August 6, 2022, from https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/