**Part 1: Policy-Based and Value-Based Algorithms**

**Chapter 2: REINFORCE:**

* Reinforce: during learning, actions that resulted in good outcome should become more probable and the one with bad outcomes should be less probable. If learning is successful, over the course of many iterations action probabilities produced by the policy shift to distribution that results in good performance in environment.
* Action probabilities are changed by following the policy gradient, there fore reinforce is known as policy gradient algo
* Algo need 3 components:

+ A parameterized policy

+ An objective to be maximized

+ A method for updating the policy parameter

**1/ Policy:**

* Policy pi is a function mapping states to action probabilities, which is used to sample an action a.
* In reinforce, an agent learns a policy and uses this to act in an environment.
* A good policy maximizes the cumulative discounted rewards.
* We can represent policy with DNN.

**2/ Objective function:**

* We defined the objective that is maximized by the agent in Reinforce algo.
* An objective can be understood as the agent’s goal such as winning a game or getting the highest score possible.

**3/ Policy Gradient:**

* Policy and Objective are important to derive policy gradient algorithms.

+ Policy provides a way for an agent to act

+ objective provides a target to maximize