Reinforcement Learning
Milad Kiwan

Exercise 0

1.

Policy:

Is the agent's behaviour, in other words it determines how the agent would act in each situation and which action the agent could do. And it is very important because it should select an action from any state in such way we will obtain the most possible reward. So its main job is maximizing the reward in each state.

The policy could be deterministic which maps a situation or state to action, it is usually used when the agent have a complete knowledge of the environment, or Stochastic where in the case the policy specifies probabilities for each action given a particular state, and it is used specially when action space has many dimensions.

Value Function:

Is a prediction of the expected future reward, and it say us "how well we are doing in a particular situation (state) if the policy selects this action?" or "how much reward we can get if we follow this behaviour?". So intuitively we can see that it depends on the policy, but as well the policy uses it to select the action that gets most possible reward.

Model:

The model is used to imagine how might the environment behave and use it to select the action, and allow us to lean about the environment. To do that we have to define the tow parts of the model:

- Transition Model: learns to predict the dynamics of the environment, it say us in what state the environment is going be. So using this we could know what could happen the agent in the next. Transition model tell us the probability of the next state given the current one and the action.

- Reward Model: using the reward model we can also learn to predict the next state and send back a reward to indicate how well is our next state of environment. The reward model tell us the expectation of reward given the current one and the action.

Building the model is not a strict requirement.