1. **State utility theory**

Utility functions are a product of Utility Theory which is one of the disciplines that helps to address the challenges of building knowledge under uncertainty. It is often combined with the probabilistic theory to create decision-theoretic agents. Utility Theory lays out the foundation to create and evaluate Utility Functions. Typically, Utility Theory uses the notion of Expected Utility (EU) as a value that represents the average utility of all possible outcomes of a state, weighted by the probability that the outcome occurs.

1. **Types of Reinforcement learning**

**Positive Reinforcement:**

The positive reinforcement learning means adding something to increase the tendency that expected behavior would occur again. It impacts positively on the behavior of the agent and increases the strength of the behavior.

This type of reinforcement can sustain the changes for a long time, but too much positive reinforcement may lead to an overload of states that can reduce the consequences.

**Negative Reinforcement:**

The negative reinforcement learning is opposite to the positive reinforcement as it increases the tendency that the specific behavior will occur again by avoiding the negative condition.

It can be more effective than the positive reinforcement depending on situation and behavior, but it provides reinforcement only to meet minimum behavior.

1. **List the terms used in reinforcement learning**

* Agent(): An entity that can perceive/explore the environment and act upon it.
* Environment(): A situation in which an agent is present or surrounded by. In RL, we assume the stochastic environment, which means it is random in nature.
* Action(): Actions are the moves taken by an agent within the environment.
* State(): State is a situation returned by the environment after each action taken by the agent.
* Reward(): A feedback returned to the agent from the environment to evaluate the action of the agent.
* Policy(): Policy is a strategy applied by the agent for the next action based on the current state.
* Value(): It is expected long-term retuned with the discount factor and opposite to the short-term reward.
* Q-value(): It is mostly similar to the value, but it takes one additional parameter as a current action (a).

**4. Define stationary process**

In [mathematics](https://en.wikipedia.org/wiki/Mathematics) and [statistics](https://en.wikipedia.org/wiki/Statistics), a stationary process (or a strict/strictly stationary process or strong/strongly stationary process) is a [stochastic process](https://en.wikipedia.org/wiki/Stochastic_process) whose unconditional [joint probability distribution](https://en.wikipedia.org/wiki/Joint_probability_distribution) does not change when shifted in time.[[1]](https://en.wikipedia.org/wiki/Stationary_process#cite_note-1) Consequently, parameters such as [mean](https://en.wikipedia.org/wiki/Mean) and [variance](https://en.wikipedia.org/wiki/Variance) also do not change over time.

**5. State optimal policy**

In a finite Markov Decision Process (MDP), the optimal policy is defined as a policy that maximizes the value of all states at the same time¹. In other words, if an optimal policy exists, then the policy that maximizes the value of state s is the same as the policy that maximizes the value of state s'.

**6. List the components of MDP**

MDP contains a tuple of four elements (S, A, Pa, Ra):

* A set of finite States S
* A set of finite Actions A
* Rewards received after transitioning from state S to state S', due to action a.
* Probability Pa.

**7. what is reinforcement learning?**

* Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
* In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike [supervised learning.](https://www.javatpoint.com/supervised-machine-learning)
* Since there is no labeled data, so the agent is bound to learn by its experience only.

**8.** **What Is interference in temporal model?**

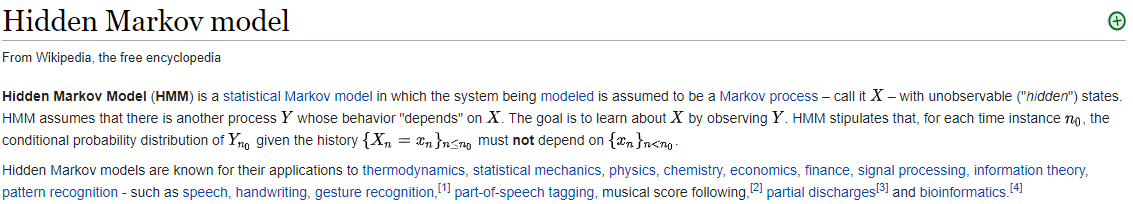
**(Dorkatle... idi okkati own ga rayandi)**

**9.State policy iteration and policy evaluation**

Policy iteration starts with a policy and iteratively improves it. It starts with an arbitrary policy *π0* (an approximation to the optimal policy works best) and carries out the following steps starting from *i=0*.

* Policy evaluation: determine *Vπi(S)*. The definition of *Vπ* is a set of *|S|* linear equations in *|S|* unknowns. The unknowns are the values of *Vπi(S)*. There is an equation for each state. These equations can be solved by a linear equation solution method (such as Gaussian elimination) or they can be solved iteratively.

**10. What is Hidden Markov model?**

****