N.V Purna Chandar — Assignment - 1 — 20BAI10362

a) Contrast the Model-Based and Model free reinforcement learning.

Ans) Reinforcement learning (RL) is a machine learning approach in which an agent learns to interact with an environment by taking actions and receiving feedback in the form of reward or penalties.

There are two main approaches to RL: model-based and model-free.

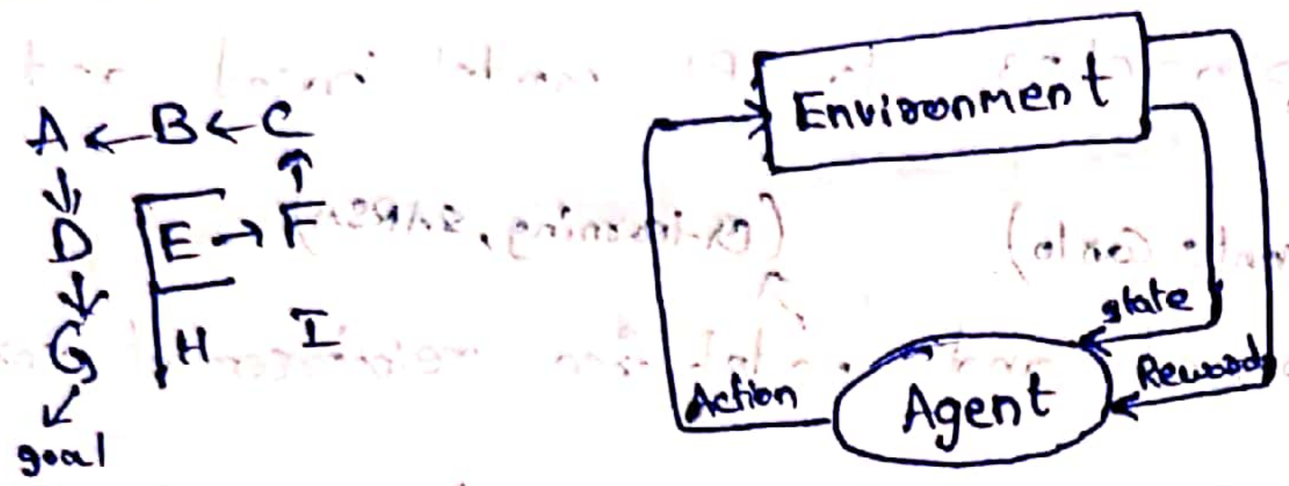
(Dyna-Q; Monte Carlo) (Q-learning, SARSA)

Key difference between model-based and model-free reinforcement learning

	Model-Based RL	Model-Free RL
Knowledge representation	The agent learns an explicit model of the environment	The agent does not build an explicit model of the environment.
Planning	The agent plans its actions based on its internal model of the environment.	The agent uses trial-and-error to learn a policy with out planning explicity.
Computational Complexity	Can be computationally expensive since the agent needs to maintain a model of the environment.	Can be more computationally efficient since it does not require the agent to maintain explicit model
Generalization	Can generalize better to new situations since it has an explicit model of the environment.	May struggle to generalize to new situations because it learns a policy based on specific experiences.
Sample efficiency	Can require fewer samples to learn a good policy because it can use its internal model to simulate the different scenarios.	May require more samples to learn a good policy because it learn directly from experiences.

Reinforcement learning :- (RL)

RL is a type of feedback based machine learning Approach in which an agent learns to make decision by interacting with an environment. For each correct action, the agent gets positive feedback, and for each incorrect action, the agent gets negative feedback or penalty



-> RL is an example of semi-supervised learning technique and is used to model sequential decision-making Process.

-> The agent interacts with environment and identifies the possible actions he can perform.

-> The primary goal of an agent in RL is to perform actions by looking at the environment and get the maximum positive reward.

-> In RL, the agent learns automatically using feedbacks without any labeled data unlike supervised learning, so the agent is bound to learn by experience.

-> RL is used to solve sequential, and the goal is long-term, such as game-playing robotics, etc.. autonomous driving, Healthcare, finance, Inventory management.

Key elements in RL :-

Agent: is the decision-maker that interacts with the environment by taking action and receiving reward signal.

Environment: is the external world that includes current state, set of possible actions and reward function.

Reward function: assigns a scalar value to each state-action pair, which indicates the desirability of taking that action in that state.

RL uses Reward signal to update agents policy, which leads to better decision making over time.

Diff b/w supervised, unsupervised and Reinforcement learning:-

Supervised learning:-

This involves learning from labelled data i.e learning is done in presence of supervisor, where the algorithm receives inputs and corresponding outputs, and learns to map input to output. The main goal is to predict the target output variable based on input variable.

eg:- Image classification, speech recognition and object detection.

Unsupervised learning:-

This involves learning from unlabelled data i.e done in absence of supervisor, where the algorithm receives only inputs and no corresponding output, and learns to identify patterns and relationships within data. The main goal is to discover the hidden patterns, structures and relationships within data.

eg:- clustering, association, anomaly detection & dimensionality reduction.

Reinforcement learning:-

This involves learning through interaction with the environment, where the algorithm receives reward (or) punishment for certain actions, and learns to take actions that maximize total reward. The main goal is to learn a policy that maximizes the cumulative reward over time.

eg:- Game playing, robotics, autonomus driving.

Advantages:-

i) Adaptability: learns from experience and adapt to changes in the environment, suitable for applications where the env is dynamic and complex.

ii) Ability to handle complex problems:

iii) Self-learning: work on unlabelled data and learns from experiences.

iv) Exploration: explor the env and learn for better decision making.

Disadvantages:-

i) Time-consuming : can't provide real-time decisions in complex env.

ii) lack of interpretability : diff to interpret, making it hard to identify reason behind thier actions.

iii) Reward engineering : Designing a suitable RF is challenging.

iv) High computational requirments :

v) Ethical concerns.

Application of RL :-

i) Robotics

ii) Game playing

iii) Autonomous vehicles

iv) NLP

v) Finance

vi) Healthcare

vii) Industrial control

viii) Education.

Exploration :- Process of agent exploring the env to learn more about it and discover potentally better actions. This is done as in many env the optimal actions may not obtain right away so, the agent explor the env to know more about env & increase rewards.

Strategies used :- Random exploration, Epsilon-greedy, Boltzman exploration.

Exploitation : Process of agent taking the actions that it belives will lead to highest rewards based on its current knowledge of env. This is done since, once the agent gathered some info about the env, it can use that info to make better decisions.

Strategies used : Greedy ; softmax ;

-) The exploration-exploitation Trade-off should be balanced, as it may waste time by exploring all paths (or) miss out best ways by exploiting some paths.

Strategies used : Epsilon greedy ; Thompson sampling ; UCB (upper confidence boundary)

Q learning algorithm :-

$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s',a')$$

s = current state

a = current action

r = immediate reward.

γ = Discount factor

s' = next state

a' = next available actions.

For each s,a initialize the table entry $\hat{Q}(s,a)$ to zero.

observe the current state.

Do forever:
- select an action and execute it
- Receive immediate reward r
- observe the new state s'
- update the table entry for $\hat{Q}(s,a)$ as follows:

$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s',a')$$

- $s \leftarrow s'$ (set next state as current state & repeate).

Bellman equation :-

$$V(s) = \max\left[R(s,a) + \gamma V(s')\right]$$

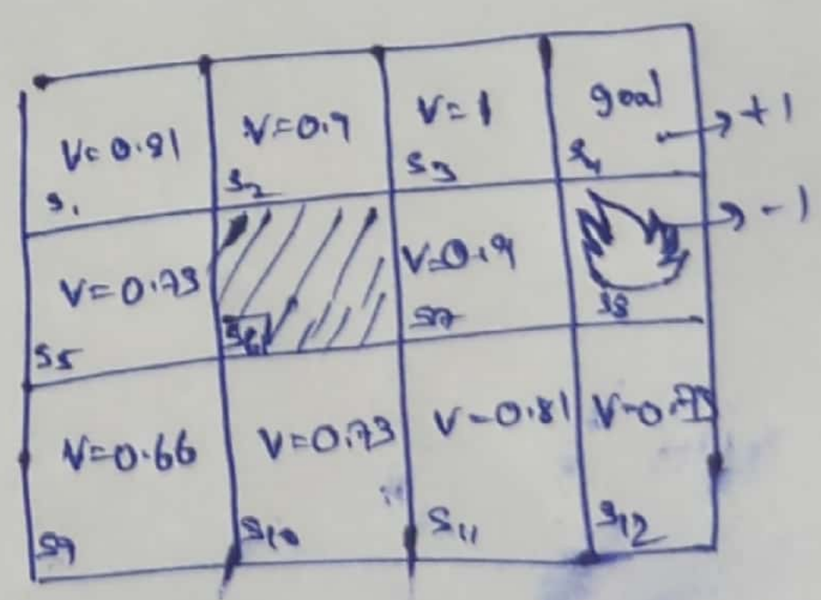
a = action performed by agent ; s = stat occured by performing the action

R = reward obtained ; γ = discount factor.

$V(s)$ = value calculated at a particular point. ; $V(s')$ = value at previous state.

$R(s,a)$ = Reward at a particular state s by performing an action a

this $R(s,a)$ will be '0' for all new blocks except one beside goal state.



For block S_3 :- given $(\gamma = 0.9)$

$$\therefore V(S_3) = \max\left[R(s,a) + \gamma V(s')\right]$$

$$\therefore V(S_3) = \max\left[1 + 0.9(0)\right] = 1$$

for block S_2 :- $V(S_2) = \max\left[R(s,a) + \gamma V(s')\right]$

$$= \max\left[0 + 0.9(1)\right] = 0.9$$

block S_1 :- $V(S_1) = \max\left[0 + 0.9(0.9)\right] = 0.81.$

So, on ...

Bag ① Bag ②

4W 6B 4W 3B

$\Rightarrow P(E_1) = 1/2 = P(E_2)$.

• Let A is event of drawing black ball.

$\Rightarrow P\left(\frac{A}{E_1}\right) = \frac{6}{10} = \frac{3}{5}$; $P\left(\frac{A}{E_2}\right) = \frac{3}{7}$

now Probability of drawing black ball out of bag 1 from two bags

$$P\left(\frac{E_1}{A}\right) = \frac{P\left(\frac{A}{E_1}\right) \cdot P(E_1)}{P\left(\frac{A}{E_1}\right) \cdot P(E_1) + P\left(\frac{A}{E_2}\right) \cdot P(E_2)} = \frac{\left(\frac{3}{5}\right)\left(\frac{1}{2}\right)}{\left(\frac{3}{5}\right)\left(\frac{1}{2}\right) + \left(\frac{3}{7}\right)\left(\frac{1}{2}\right)} = \frac{\frac{3}{10}}{\frac{3}{10} + \frac{3}{14}}$$

$$\Rightarrow \frac{3}{10} \times \frac{70}{36_{12}} \Rightarrow \frac{7}{12}$$