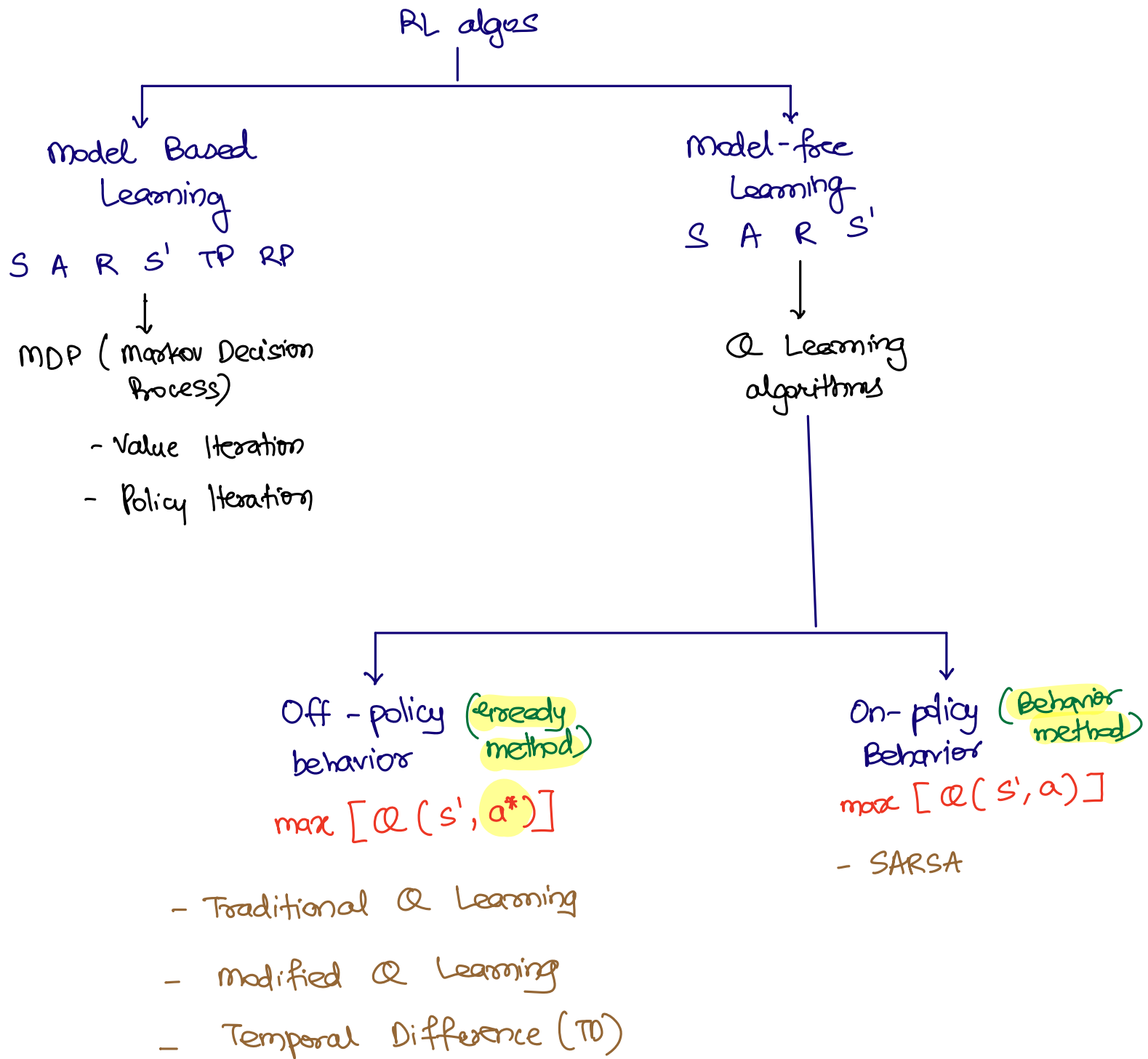


# Reinforcement Learning

Trainer: Prashant N



## Traditional Q learning\*

$$Q(s, A) = IR + (\gamma \max_{a^*} Q(s', a^*))$$

Annotations for Traditional Q learning:

- current state (points to  $s$ )
- current action (points to  $A$ )
- immediate reward (points to  $IR$ )
- discount factor (hyperparameters) (points to  $\gamma$ )
- next state (points to  $s'$ )
- all valid actions (points to  $a^*$ )

Q value of next state for all possible action and extract the max q value.

## Modified Q learning

$$Q(s, A) = Q(s, A) + \alpha [R(s, A) + \gamma \max_{a^*} Q'(s', a^*) - Q(s, A)]$$

Annotations for Modified Q learning:

- current Q value from Q table (points to  $Q(s, A)$ )
- learning rate (points to  $\alpha$ )
- IR (points to  $R(s, A)$ )
- discount factor (points to  $\gamma$ )
- max q value of next state from all possible actions (points to  $\max Q'(s', a^*)$ )
- current Q value from Q table (points to  $Q(s, A)$ )

q-target (points to  $R(s, A) + \gamma \max Q'(s', a^*)$ )

q-predict (points to  $Q(s, A)$ )

## Temporal Difference (TD)

$$Q(s, a) = Q(s, a) * (1 - \alpha) + \alpha * [R(s, a) + \gamma \max_{a^*} Q(s', a^*)]$$

Annotations for Temporal Difference (TD):

- current Q value from initialized Q Table (points to  $Q(s, a)$ )
- learning rate (points to  $\alpha$ )
- discount factor (points to  $\gamma$ )
- $R(s, a)$  (points to  $R(s, a)$ )