CS747 Assignment 2

Manas Vashistha
17D070064

November 13, 2020

Task 1

1. The origin is assumed to be on the top left of the grid. X-axis is horizontally right and y-axis is vertically downwards.
2. States are numbered in column major order, total states = $n\_rows \times n\_cols$.
3. Actions are [0, 1, 2, 3] for [N, E, S, W] in baseline case and [0, 1, 2, 3, 4, 5, 6, 7] for [N, NE, E, SE, S, SW, W, NW] for King's moves. Reward for every transition is -1.
4. If a perimeter wall is encountered, the agent moves to the adjacent cell of the wall.

0	7	14	21	28	35	42	49	56	63
1	8	15	22	29	36	43	50	57	64
2	9	16	23	30	37	44	51	58	65
3	10	17	24	31	38	45	52	59	66
4	11	18	25	32	39	46	53	60	67
5	12	19	26	33	40	47	54	61	68
6	13	20	27	34	41	48	55	62	69

Task 2

1. $Q(s,\ a)$ is initialized with zeros for all s and a.
2. A valid action is chosen randomly with ϵ probability and the optimal action is chosen with probability $1-\epsilon$.
3. While choosing the optimal action, any one of the actions having the same maximum value is chosen randomly.
4. The update rule is given as $\hat{Q}^{t+1}(s^t,\ a^t) \leftarrow \hat{Q}^t(s^t,\ a^t) + \alpha[r^t + \gamma \hat{Q}^t(s^{t+1},\ a^{t+1}) - \hat{Q}^t(s^t,\ a^t)]$. This rule is used for SARSA(0) agent with $\alpha = 0.5$, $\gamma = 1$.

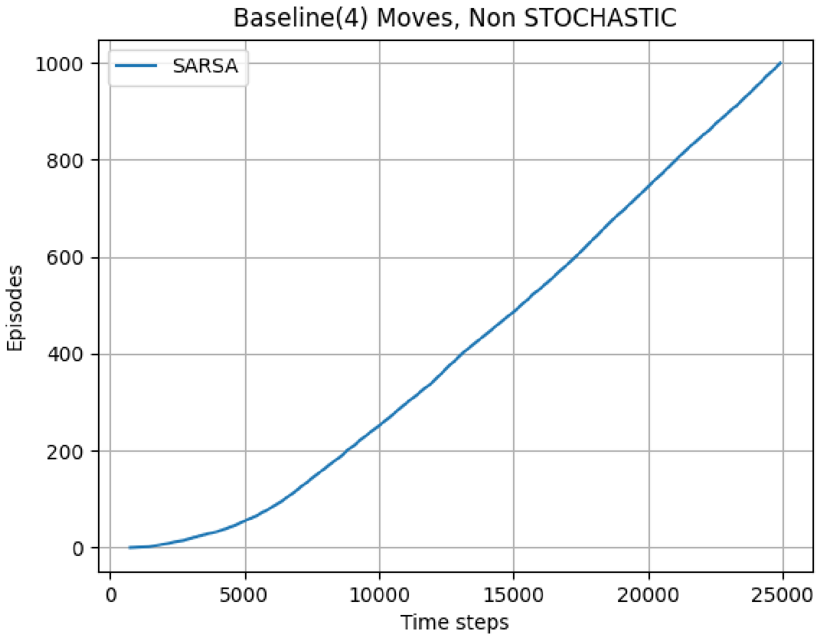

Figure 1: SARSA(0), Baseline(4) moves, Non-Stochastic

Task 3

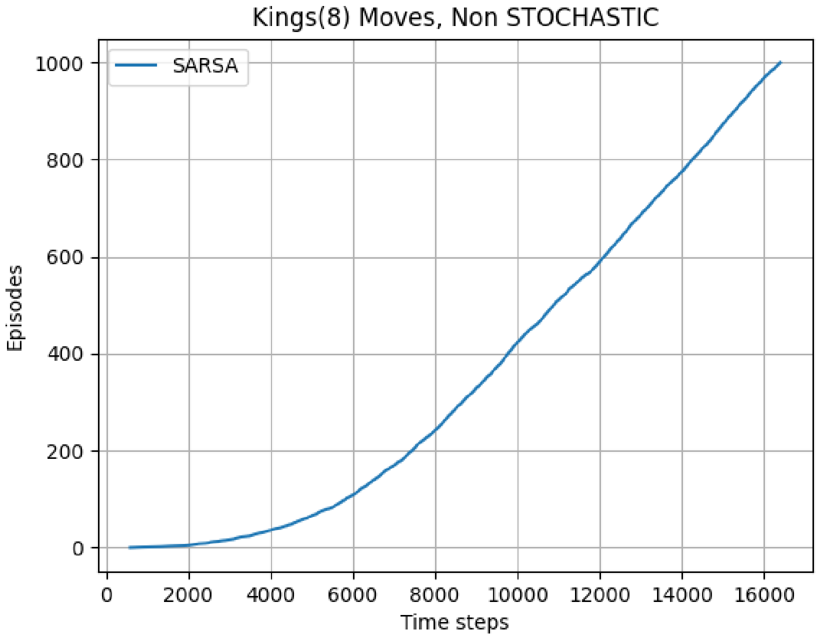

Figure 2: SARSA(0), Kings(8) moves, Non-Stochastic

Task 4

1. After taking an action, Y-coordinate is increased, decreased by 1 or remains unchanged with probability 1/3 each.

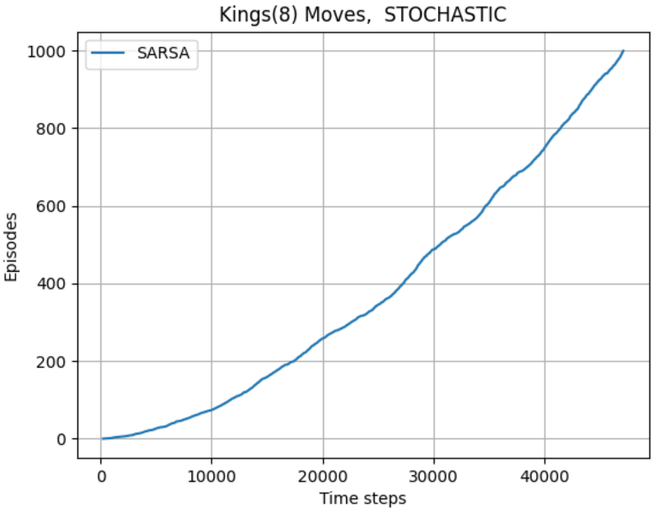

Figure 3: SARSA(0), Kings(8) moves, Stochastic

Task 5

1. The same parameters ($\alpha = 0.5,\ \gamma = 1$) are used for all the cases.

2. Update for Expected SARSA:
$\hat{Q}^{t+1}(s^t,\ a^t) \leftarrow \hat{Q}^t(s^t,\ a^t) + \alpha[r^t + \gamma \sum_{a \in A} \pi^t(s^{t+1},\ a)\hat{Q}^t(s^{t+1},\ a) - \hat{Q}^t(s^t,\ a^t)]$.

3. Update for Q-Learning:
$\hat{Q}^{t+1}(s^t,\ a^t) \leftarrow \hat{Q}^t(s^t,\ a^t) + \alpha[r^t + \gamma \max_{a \in A} \hat{Q}^t(s^{t+1},\ a) - \hat{Q}^t(s^t,\ a^t)]$.

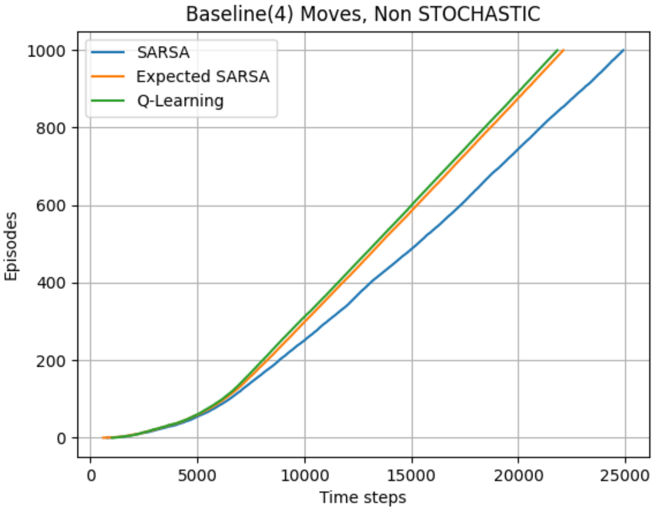

Figure 4: SARSA(0), Expected SARSA, Q-Learning, Baseline Case 4-moves, Non-Stochastic

Observations

1. All the curves have increasing slope which represents that the final state is achieved in less time steps as we progress.

2. The slope for Expected SARSA and Q-Learning is higher than that of SARSA signifying that they achieve the goal more quickly than SARSA. The order is $Q - Learning > ExpectedSARSA > SARSA$ where the greater implies quick achievement of the goal.

3. The number of expected moves in non-stochastic kings case is less than the number of expected moves in baseline case while this number is larger in the stochastic kings case.

4. In the non-stochastic kings case as we have more valid actions we can find a shorter path and hence less expected moves.

5. While in the stochastic case, a longer path is followed due to randomness as compared to the non-stochastic kings case and base line case.