Reinforcement Learning

Evelyn Yosiana / 13522083

1. Jelaskan cara kerja dari algoritma Q-Learning dan SARSA!
2. Bandingkan hasil dari kedua algoritma tersebut, bagaimana hasil perbandingannya? Jika ada perbedaan, jelaskan alasannya!

1. **Q-Learning** merupakan salah satu algoritma reinforcement learning off-policy (agent dapat memperbaharui nilai fungsi tanpa perlu mengikuti aturan dalam memilih aksi) dimana agent di-training untuk memaksimalkan total reward dengan mengeksplorasi berbagai tindakan. Langkah-langkah:
    - Inisialisasi Q table dengan 0.
    - Pilih aksi dengan probabilitas epsilon (jika dalam kode ini) untuk eksplorasi atau 1 - epsilon untuk eksploitasi data.
    - Lakukan aksi dengan bergerak ke state selanjutnya, kemudian perbarui Q table dengan Q value sebagai berikut.

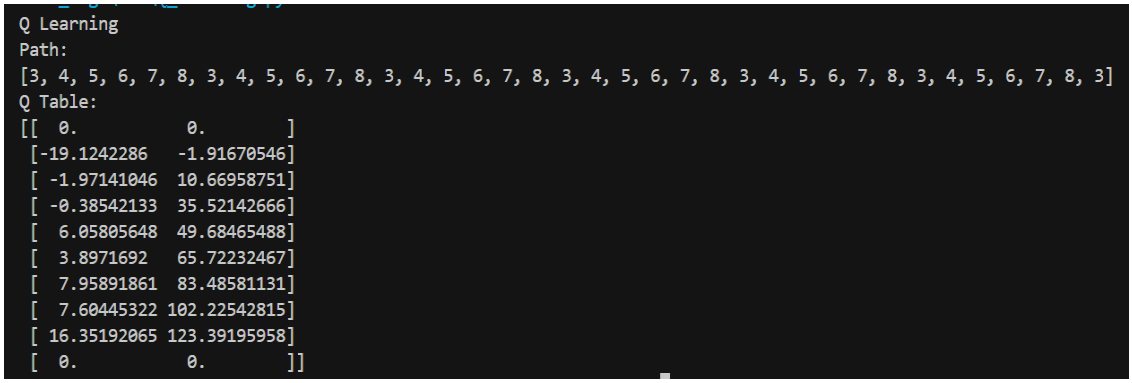
      $$Q_{(state,\ action)} = R_{(state,\ action)} + gamma\ (MAX(Q_{(state,\ all\ action)}))$$
    - Ulangi sejumlah episode yang diinginkan.

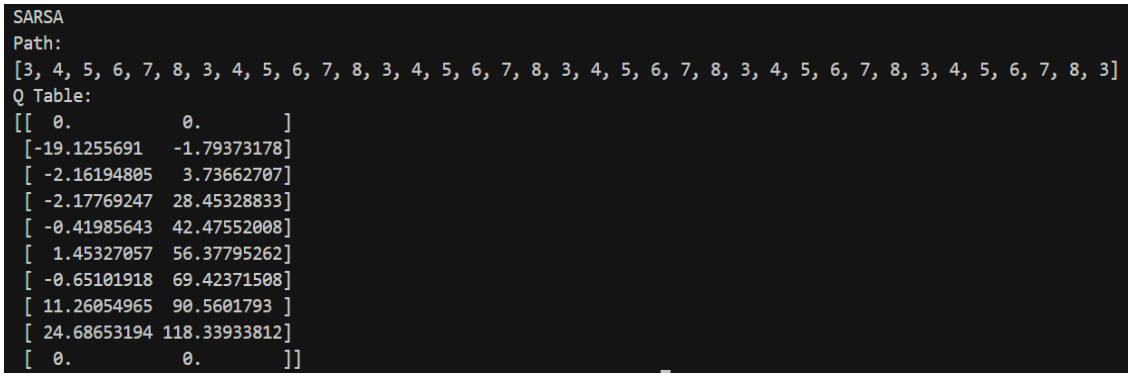
    State-Action-Reward-State-Action (SARSA) merupakan salah satu algoritma reinforcement learning on-policy (agent dapat memperbaharui nilai fungsi berdasarkan action yang sedang dilakukan) dimana agent di-training untuk memaksimalkan total reward dengan mengeksplorasi berbagai tindakan. Langkah-langkah:
    - Inisialisasi Q table dengan 0.
    - Pilih aksi dengan probabilitas epsilon (jika dalam kode ini) untuk eksplorasi atau 1 - epsilon untuk eksploitasi data.
    - Lakukan aksi dengan bergerak ke state selanjutnya, kemudian perbarui Q table dengan Q value sebagai berikut.

      $$Q_{(state,\ action)} = Q_{(state,\ action)} + alpha[r + gamma\ .\ Q\ (state',\ action') - Q_{(state,\ action)}$$
    - Ulangi sejumlah episode yang diinginkan.

    Jika Q table sudah didapatkan, perilaku dari agent akan sama, yaitu agent akan memilih Q yang paling optimal (bernilai paling besar) dalam mengambil tindakan selanjutnya (pindah ke state selanjutnya).

2. Perbandingan Q-learning dan SARSA dengan parameter yang sama: alpha=0.1, gamma=0.9, epsilon=0.1, episodes=10





Dengan parameter yang sama, algoritma Q-Learning dan SARSA memberikan hasil yang sama dalam konteks pathnya, namun memberikan hasil Q Table yang berbeda. Perbedaan ini terjadi karena dalam proses trainingnya, algoritma Q-Learning akan mengupdate nilai Q berdasarkan **estimasi reward maksimum** state berikutnya, sedangkan algoritma SARSA mengupdate nilai Q sesuai dengan **aksi aktual** yang dipilihnya. Dengan kata lain, algoritma Q-Learning lebih berani dalam mengambil resiko untuk mendapatkan solusi optimal dengan cepat, sedangkan algoritma SARSA cenderung lebih berhati-hati terhadap berbagai resiko sehingga hasilnya mungkin saja kurang optimal.