1. Jelaskan apa yang dimaksud dengan Reinforcement Learning!

Reinforcement Learning adalah salah satu cabang dari pembalajaran mesin. Ciri khas reinforcement learning adalah adanya sebuah himpunan *state ,* himpunan aksi A,  *policy* pemilihan aksi, serta *reward* perpindah dari suatu state ke state lain ketika menggunakan suatu aksi. Mesin ini belajar melalui trial dan error. Mesin akan mencoba gerakan tertentu dan menerima feedback berupa reward baik positif maupun negatif. Reward ini akan digunakan untuk mengoptimalkan gerakan mesin pada kasus-kasus selanjutnya.

1. Jelaskan bagaimana proses dari Q-learning bekerja!

Sebelum menjelaskan q-learning, adapun elemen-elemen yang harus didefinisikan terlebih dahulu.

S: Himpunan *State* suatu *environment*

A: Himpunan aksi

R: *Reward* dari perpindahan suatu *state* ke *state* lain menggunakan aksi A.

Q-Table: Tabel yang menyimpan nilai Q berukuran (|S|,|A|). Notasi Q(S,A) berarti pengaksesan nilai Q baris ke S kolom ke A, yang berarti mengakses nilai Q ketika mengambil aksi A pada himpunan S. Tabel ini dapat digunakan untuk mengukur seberapa bagus aksi A pada *state* S.

Policy: Kebijakan model dalam memilih tindakannya.

Pertama-tama, algoritma Q-learning akan menginisiasi Q-table dengan nilai 0, dengan suatu state awal s. Kemudian sebuah aksi diambil menggunakan *policy* yang digunakan dan dicatat reward atau punishment dari aksi yang diambil. Setelah itu, elemen q-table diperbarui dengan persamaan sebagai berikut

: State sebelum gerakan

Gerakan yang diambil

: State setelah gerakan

Nilai Q maksimum pada baris ke a tau nilai q pada gerakan terbaik pada state setelah gerakan

: Discount rate,

: Learning rate, skala kecepatan model untuk memperbarui modelnya

Proses tersebut diulang hingga kasus terminal, kemudian dimulai episode yang baru tanpa mengubah Q-table yang telah berubah pada episode sebelumnya. Algoritma tersebut diulang sesuai jumlah episode yang diinginkan.

Setelah semua episode selesai, terbentuklah Q-table yang dapat digunakan AI untuk memilih gerakan optimal yang sesuai statenya pada saat itu. Caranya dengan memilih aksi dengan Q value tertinggi dari statenya pada saat itu.

1. Jelaskan perbedaan algoritma Q-learning dan algoritma SARSA, serta kelemahan dan kekurangan masing-masing algoritma!

Q-Learning dan SARSA memiliki algoritma yang sama, hanya saja terdapat perbedaan cara dalam memperbarui nilai Q nya. SARSA memperbarui nilai Q nya sebagai berikut,

Sebelum memperbarui q-value, algoritma SARSA akan mengambil aksi baru sesuai policy yang digunakan. Aksi baru ini akan dipilih sebagai reward yang bisa diharapkan pada jika mengikuti policy. Di sisi lain, algoritma Q-learning tidak mengikuti policy, melainkan mengambil reward terbaik yang bisa diharapkan pada state baru tersebut. Maka perbedaannya terletak pada SARSA yang memprediksi reward dengan mengikuti policy, dengan Q-Learning yang memprediksi reward dengan cara greedy.

Kedua cara menghitung reward yang dapat diharapkan tersebut mempengaruhi sifat dari setiap model tersebut, antara lain

1. Q-Learning

Kelebihan:

* Karena sifat greedy Q-Learning, Q-Learning dapat belajar strategi yang optimal pada jumlah episode yang tepat.

Kekurangan:

* Sifat greedy Q-Learning dapat memberikan proses pembelajaran yang tidak stabil akibat overestimasi nilai Q.
* Q-learning butuh pembelajaran yang lebih lama pada lingkungan yang besar atau kompleks untuk eksplorasi.

1. SARSA

Kelebihan:

* SARSA lebih cepat menemukan solusi karena sifatnya yang disiplin mengikuti *policy.*

Kekurangan:

* SARSA cenderung memberikan solusi yang kurang optimal karena pembaruan Q value pada SARSA tetap mengikuti *policy*