Nama : Agung Wahyu Prayogo
Ervina Nurfa
Sandi Wiguna
Roy Fernando
Kelompok RL : 32
Program : Artificial Intelligence Mastery Program
Kampus : Universitas Bhayangkara Jakarta Raya

Minggu – 6
Selasa, 29 Maret 2022

Nomer 1
Konsep

- Kita butuh environment terlebih dahulu
- Observasi environtment tadi sehingga didapatkan aturan / policy yang telah di tentukan
- Ketika sudah dapat aturan, pasti ada episode yang didapatkan (harus sampai akhir)
- Setiap episode ada reward yang berbeda beda tergantung dari policy yang ada

First Visit

- Jika sudah jalankan First Visit, menjumlah semua reward tiap episode dengan catatan reward akan dihitung setelah agent mengunjungi state variable tertentu
- Setelah menghitung total sum, bagi dengan banyaknya episode

Every Visit

- Lalu lanjut metode Every Visit, bedanya disini kita menjumlahkan semua reward di **episode dengan catatan kamulatif** reward akan dihitung setiap agent mengunjungi state variable
- Lalu total sum (dengan aturan berbeda dari first fisit) bagi dengan banyaknya episode

Estimation & Control

- Karena tidak ada model di MC, Agent perlu explore semua kemungkinan yg ada, setiap Gerakan dihitung berupa estimasi, dari explore barusan Agent melakukan Evaluasi dan belajar dari kesalahan

Nomor 4

Perbedaan

DP

- Di DP untuk mencari value, dengan cara probabilitas dikali dengan reward lalu ditambah dengan state setelahnya
- Value di DP menggambarkan seberapa baik reward dari suatu action
- Untuk mengimplimentasikan algoritma DP, dibutuhkan pengetahuan yang lengkap sehingga membutuhkan waktu, tenaga, uang, yang besar untuk mendapatkan pengalaman yang cukup
- Lalu di DP tiap pergerakan dihitung probabilitas, tidak masuk akal jika agen bolak balik
- Di DP butuh Environment, State, Action, Reward, Probability
- Bootstrap, jadi value saat ini dibangun dari state berikutnya

MC

- Berbeda dari DP, Monte Carlo (MC) tidak mengambil pengetahuan lengkap
- MC belajar dari pengalaman, hanya saja butuh episode yang utuh (sampai selesai / akhir)
- Mencari value berdasarkan rata – rata returns. Dengan semakin banyak return, nilai rata- ratanya diharapkan konvergen nilai value
- Sayangnya MC ini metode untuk episodic, seperti catur yang terus berjalan hingga akhir, tidak seperti dadu
- Tidak bootstrap
- Nilai konvergen meningkat bila nilai return semakin mendekati angka tak hingga