1. TD Learning kết hợp các ưu điểm của Dynamic Programming (DP) và Monte Carlo (MC)

- **Bootstrapping (mượn từ DP):**
  TD cập nhật giá trị của trạng thái hiện tại dựa trên phần thưởng quan sát được và ước lượng giá trị của trạng thái kế tiếp, tức là "cập nhật từ một ước lượng khác" (update a guess from a guess).

- **Không cần mô hình môi trường (khác DP):**
  TD không yêu cầu có mô hình chuyển trạng thái, chỉ cần dữ liệu kinh nghiệm trực tiếp từ môi trường.

- **Cập nhật online và từng bước (khác MC):**
  TD cập nhật giá trị ngay tại mỗi bước đi, không cần đợi đến cuối episode như Monte Carlo.

---

2. Tabular TD(0) - thuật toán cơ bản của TD

- Chỉ cần lưu trữ trạng thái trước đó (previous state) để cập nhật giá trị.

- Thuật toán rất đơn giản nhưng hiệu quả.

---

3. Ưu điểm so với MC và DP

Phương pháp	Cần mô hình?	Cập nhật online?	Cập nhật ngay từng bước?	Tốc độ hội tụ
Dynamic Programming (DP)	Có	Không	Không	Tốt
Monte Carlo (MC)	Không	Không	Không (đợi cuối episode)	Chậm

Temporal Difference (TD)	Không	Có	Có	Nhanh hơn MC

---

4. Ví dụ minh họa

- Dự đoán thời gian về nhà từng bước trên đường đi.
- TD giúp điều chỉnh dự đoán liên tục khi có thêm thông tin mới.
- Monte Carlo phải đợi tới lúc về nhà mới biết kết quả để cập nhật.

---

5. Thí nghiệm so sánh TD và Monte Carlo

- TD cập nhật sau mỗi bước chuyển trạng thái.
- Monte Carlo cập nhật chỉ sau khi kết thúc episode.
- TD hội tụ nhanh hơn và có sai số cuối cùng thấp hơn.

---

6. Phía trước: Xây dựng các thuật toán điều khiển (Control) dựa trên TD

- Trong các module tiếp theo, bạn sẽ học cách TD không chỉ để dự đoán giá trị mà còn để học các chính sách tối ưu trong bài toán điều khiển.
-