**Họ và tên: Lê Hải Đăng**

**MSSV: 20110243**

**Bài tập 2**: Nếu cài đặt Cross-Entropy method để huấn luyện agent leo dốc trong **MountainCar-v0** environment, thì thời gian huấn luyện đến khi agent thực hiện được nhiệm vụ sẽ có thể mất nhiều giờ. Thời gian này **dài hơn rất nhiều so với Bài tập 1** (chỉ dưới vài chục phút) dù sử dụng cùng reward bounds và network architectures.   
**Hãy giải thích lý do**.

Theo em có 02 lý do chính:

+ Trong **MountainCar-v0** cách di chuyển của xe sẽ tương ứng với 03 actions: qua trái, đứng yên và qua phải. Mình không thể tăng/giảm tốc độ của xe nên trong một action xe chỉ di chuyển thêm một khoảng cách tương ứng nên việc leo dốc sẽ khó khăn hơn.

+ Theo mô tả trong enviroment, một episode chỉ dừng lại khi xe tìm được goal hoặc chiều dài episode lớn hơn 200. Do đó, đa phần episode sẽ không đạt được điều đó và xe dừng lại khi chưa tìm được goal.