

1 of 10

Một trong những nhược điểm của Gradient Descent tiêu chuẩn là gì?

- ☐ A Không tính đến đạo hàm
- ☐ B Đường đi bị zig-zag trong các thung lũng dốc
- ☐ C Không thể hội tụ về cực tiểu
- ☐ D Luôn bị kẹt ở cực đại

SUBMIT ANSWER

3 of 10

Lợi ích chính của Momentum trong tối ưu hóa là:

- ☐ A Làm mất ảnh hưởng của đạo hàm
- ☐ B Loại bỏ noise trong gradient
- ☐ C Điều chỉnh tốc độ học động
- ☐ D Duy trì và tích lũy hướng đi "tốt" để tăng tốc hội tụ

SUBMIT ANSWER

2 of 10

Cập nhật tham số trong Gradient Descent với momentum được thực hiện như thế nào?

- ☐ A Sử dụng giá trị đạo hàm hiện tại
- ☐ B Sử dụng trung bình của toàn bộ tập huấn luyện
- ☐ C Sử dụng trung bình lũy thừa mũ của các đạo hàm trước đó
- ☐ D Sử dụng độ lớn tuyệt đối của gradient

SUBMIT ANSWER

4 of 10

Điểm khác biệt chính giữa Gradient Descent tiêu chuẩn và Normalized Gradient Descent là gì?

- ☐ A Normalized Gradient Descent sử dụng mini-batch thay vì toàn bộ dữ liệu
- ☐ B Normalized Gradient Descent chia gradient cho độ dài (norm) của nó
- ☐ C Normalized Gradient Descent sử dụng đạo hàm cấp hai
- ☐ D Normalized Gradient Descent thêm nhiễu vào gradient

SUBMIT ANSWER

5 of 10

Tại sao ta muốn chuẩn hóa gradient trong Normalized Gradient Descent?

- A Để cập nhật luôn có cùng độ lớn, tránh gradient quá nhỏ hoặc quá lớn
- B Để gradient có thể hội tụ về 0 nhanh hơn
- C Để giảm chi phí tính toán đạo hàm
- D Để tránh đạo hàm bị triệt tiêu ở điểm yên ngựa

SUBMIT ANSWER

7 of 10

Trong không gian có điều kiện rất lệch (ill-conditioned), CW-NGD có thể ưu việt hơn NGD vì:

- A Cho phép bước đi độc lập theo từng chiều, phù hợp với địa hình có hình elip méo
- B Giảm chi phí tính toán
- C Loại bỏ được đạo hàm bậc hai
- D Luôn hội tụ về nghiệm toàn cục

SUBMIT ANSWER

6 of 10

Điểm khác biệt chính giữa Normalized Gradient Descent (NGD) và Component-wise Normalized Gradient Descent (CW-NGD) là

- A CW-NGD sử dụng đạo hàm cấp hai, NGD thì không
- B NGD chuẩn hóa toàn bộ vector gradient, CW-NGD chuẩn hóa từng thành phần riêng lẻ
- C NGD dùng momentum, CW-NGD không dùng
- D CW-NGD yêu cầu bước học thay đổi theo thời gian

SUBMIT ANSWER

8 of 10

Cập nhật tham số trong CW-NGD tương đương với:

- A Di chuyển theo hướng gradient chuẩn hóa toàn phần
- B Di chuyển với bước hằng số theo mỗi chiều độc lập, chỉ giữ lại dấu gradient
- C Di chuyển theo đạo hàm bậc hai
- D Cập nhật sử dụng momentum

9 of 10

Điểm khác biệt chính của Adam so với Gradient Descent tiêu chuẩn là gì?



- A** Adam sử dụng đạo hàm bậc hai để cập nhật
- B** Adam chuẩn hóa gradient về độ dài đơn vị
- C** Adam duy trì các trung bình lũy thừa mũ của gradient và bình phương gradient
- D** Adam chỉ dùng dấu của gradient để cập nhật

SUBMIT ANSWER

10 of 10

So với Normalized Gradient Descent, Adam có ưu điểm nào nổi bật?



- A** Luôn cập nhật theo độ dài bằng nhau
- B** Không bị ảnh hưởng bởi các thành phần gradient lớn
- C** Dễ dàng mở rộng cho mạng nơ-ron sâu
- D** Giữ lại thông tin về độ lớn gradient và điều chỉnh tốc độ học tương ứng

SUBMIT ANSWER