

Câu hỏi: Với kiến trúc Softmax Regression mà có nhiều lớp thì có giải được bài toán phi tuyến không? Tại sao?

Trả lời: Câu trả lời là không. Dù có nhiều lớp nhưng các node xử lý đều là hàm tuyến tính, output của lớp sau chỉ là tổ hợp tuyến tính của lớp trước, vậy nên hiệu quả chỉ như một lớp tuyến tính.

Giả sử xét trên một kiến trúc Softmax Regression có 2 lớp với input là X .

Ở lớp thứ nhất, ta có:

$$Z^1 = (W^1)^T X$$

Nếu sử dụng hàm kích hoạt là một hàm tuyến tính thì dễ thấy rằng, thu được kết quả a^1 là output của lớp thứ nhất:

$$a^1 = A * Z^1 + B \text{ (A, B là số thực)}$$

Tiếp theo, ở lớp thứ hai, input là a^1 , ta có:

$$Z^2 = (W^2)^T a^1$$

Qua hàm kích hoạt tuyến tính thứ 2, thu được:

$$a^2 = A' * Z^2 + B'$$

$$a^2 = A' * (W^2)^T * A * (W^1)^T X + A' * B + B'$$

$$a^2 = A'' * (W')^T X + B''$$

a^2 thực chất vẫn chỉ là một hàm tuyến tính của input X ban đầu, vì vậy việc xếp chồng nhiều lớp tuyến tính lên nhau cũng như chỉ một lớp tuyến tính, vì thế nên nếu chỉ xếp chồng các lớp như trên mà không có sự xuất hiện của các hàm phi tuyến thì sẽ không thể nào giải được bài toán phi tuyến ở trên.