TRÍ TUỆ NHÂN TẠO TÍNH TOÁN VÀ ỨNG DỤNG TRONG NGÀNH TÀI NGUYÊN VÀ MÔI TRƯỜNG

Bài số 2. Các hàm số và tham số cốt lõi của mạng nơ ron nhân tạo

TS Nguyễn Dư Khang

Tóm tắt:

Trong bài báo kỳ trước, tác giả đã trình bày một số khái niệm cơ bản về mạng nơ ron sinh học và mạng nơ ron nhân tạo, cơ chế hoạt động, hình thức huấn luyện và các chức năng cơ bản của mạng nơ ron nhân tạo. Trong bài báo này, tác giả sẽ trình bày về một số hàm số và tham số cốt lõi được sử dụng trong mạng nơ ron nhân tạo. Phương pháp trình bày sẽ giúp độc giả mới làm quen với lĩnh vực này dần tiếp cận với những kiến thức cơ bản để bắt đầu vào công tác huấn luyện mạng và ứng dụng vào các lĩnh vực quan tâm của mình trong ngành tài nguyên và môi trường.

Từ khoá: *Toán tử XOR, gradient, gradient descent, moment, quy tắc delta.*

1. Phép toán thao tác bit XOR

Khi bắt đầu học một ngôn ngữ mới, trong bài đầu tiên ta thường học những câu chào hỏi. Đối với mạng nơ ron nhân tạo cũng vậy, quá trình huấn luyện mạng bắt đầu từ toán tử thao tác bít (XOR).

Phép toán thao tác bit XOR lấy hai dãy bit có cùng độ dài và thực hiện phép toán logic bao hàm XOR trên mỗi cặp bit tương ứng. Kết quả ở mỗi vị trí là 1 chỉ khi bit đầu tiên là 1 hoặc nếu chỉ khi bit thứ hai là 1, nhưng sẽ là 0 nếu cả hai là 0 hoặc cả hai là 1. Ở đây ta thực hiện phép so sánh hai bit, kết quả là 1 nếu hai bit khác nhau và là 0 nếu hai bit giống nhau.

Bảng chân trị cho XOR:

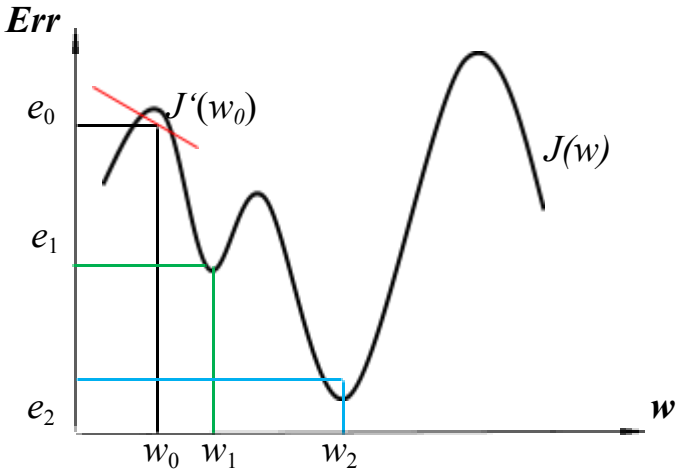
a	b	a XOR b
0	0	0
0	1	1
1	0	1
1	1	0

2. Gradien và gradient descent

Gradient là một véc tơ xác định độ cong của độ dốc và chỉ ra hướng của nó đối với bất kỳ điểm nào trên một bề mặt hoặc đồ thị (hình 1, màu đỏ).

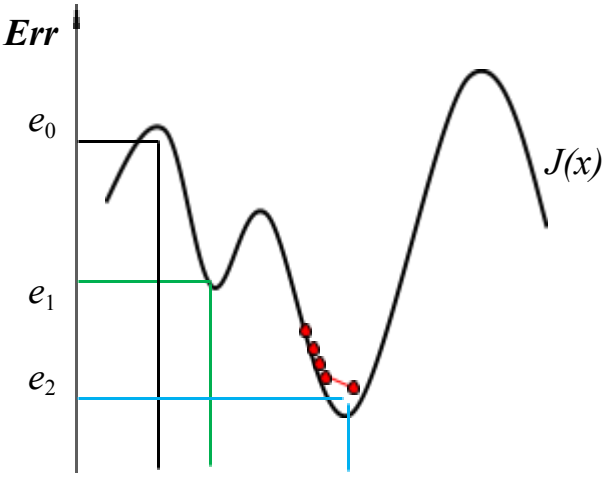
Gradient descent là phương pháp tìm cực trị (cực tiểu hoặc cực đại) của hàm số nhờ chuyển động của gradient dọc theo đồ thị.

Xét đồ thị hàm số (hình 1), trục hoành là giá trị trọng số *w* của nơ ron, trục tung là sai số *Err* gây ra bởi trọng số trên. Như vậy, đồ thị hàm số *J*(*w*), được gọi là ***hàm mất mát***, là sự phụ thuộc giữa sai số (*Err*) và trọng số (*w*) được chọn. Trên đồ thị tồn tại các điểm cực tiểu (gọi là *local minimum*), trong đó tồn tại điểm mà tại đó hàm số đạt giá trị nhỏ nhất (gọi là *global minimum*). Global minimum là một trường hợp đặc biệt của local minimum. Chúng ta quan tâm đến global minimum, tức là điểm (w_2, e_2).



Hình 1. Gradient descent

Như ta đã biết, giá trị cực trị của hàm số tại các điểm có đạo hàm bằng 0. Đối với global minimum, ta có đạo hàm (hay gradient) $J'(w_2) = 0$ và $J(w_2) < J(w)$ với mọi *w*. Gradient chuyển động dọc theo đồ thị từ điểm xuất phát w_0 theo từng bước với ***tốc độ*** α (đối với mạng nơ ron, gọi là tốc độ huấn luyện). Nếu tốc độ chậm, sẽ mất nhiều thời gia để huấn luyện mạng, nếu tốc độ quá lớn, sẽ xảy ra trường hợp bỏ qua điểm cần quan tâm là w_2, như hình 2.



w_0 w_1 w_2

Hình 2. Gradient descent với tốc độ quá lớn.

Nguyên tắc để xác định gia số hiệu chỉnh trọng số w như sau:

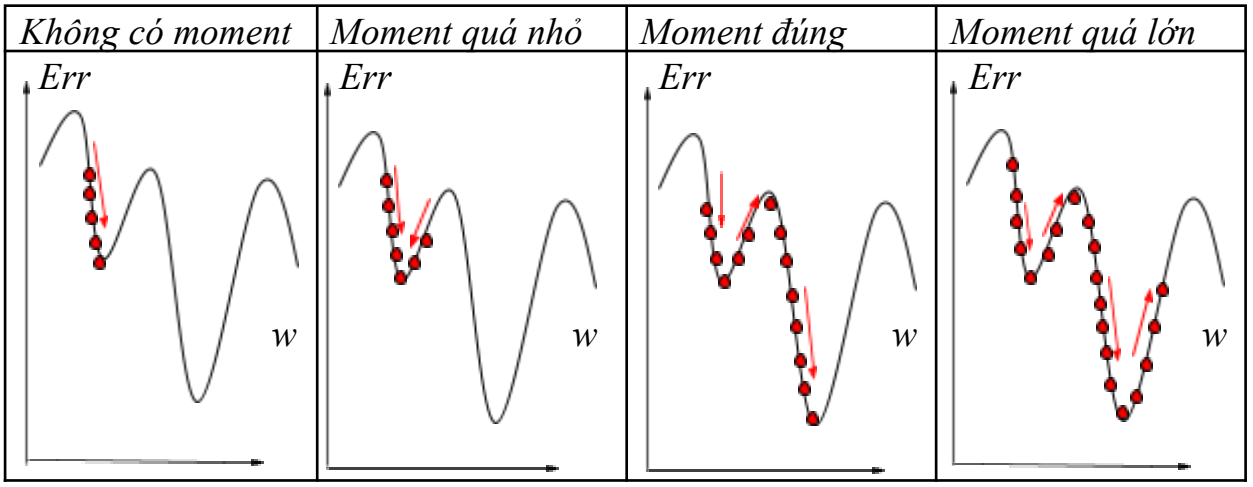
- Cho một điểm khởi tạo $w = w_0$.
- Xác định gradient $\text{Grad}_k = J'(w_k)$, trình bày sau đây ở mục quy tắc delta.
- Gia số hiệu chỉnh trọng số w xác định theo công thức:

$$\Delta w_{k,k+1} = w_{k+1} - w_k = -\alpha.\text{Grad}_k, \quad (1)$$

trong đó, k – bước chuyển động của gradient. Ở đây dấu âm (-) là thể hiện gradient đi về hướng cực tiểu (chứ không phải cực đại).

3. **Moment** (momentum)

Trên đường chuyển động, gradient gặp local minimum tại điểm (w_1, e_1), chưa phải là điểm chúng ta quan tâm. Để vượt qua điểm này, gradient cần ***moment*** β như là một lực đẩy. Sẽ xảy ra một số trường hợp nếu không chọn đúng giá trị moment như hình 3.



Hình 3. Gradient descent với các trường hợp moment.

Tại điểm (w_1, e_1) ở hình 1, ta có $J'(w_1) = 0$, chưa đạt được kết quả mong muốn (không phải global minimum).

Sử dụng moment, công thức (1) có dạng:

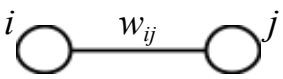
$$\Delta w_{k,k+1} = w_{k+1} - w_k = -\alpha.\text{Grad}_k + \beta.\Delta w_{k-1,k}, \quad (2)$$

trong đó, $\Delta w_{k-1,k}$ – là gia số hiệu chỉnh lần trước.

4. **Quy tắc delta** (delta rule)

Quy tắc delta được sử dụng trong gradient descent để xác định gia số hiệu chỉnh trọng số, hướng đến giá trị cực tiểu của sai số huấn luyện mạng nơ ron.

Xét hai nơ ron *i* và *j*.



Gọi đầu ra của nơ ron *i* là O_i, khi đó đầu vào của nơ ron *j* là $I_j = O_i.w_{ij}$.

Hàm sai số bình phương có dạng [1]:

$$J(w_{ij}) = (y_j - O_j)^2/2, \quad (3)$$

trong đó, y_j – là mục tiêu đầu ra và O_j – là đầu ra thực tế của nơ ron *j*.

Ta có: $O_j = F(I_j) = F(O_i.w_{ij})$,

trong đó, F - là hàm kích hoạt.

Xác định đạo hàm (là gradient của khớp nối giữa hai nơ ron) $J'(w_{ij})$:

$$\frac{\partial J(w_{ij})}{\partial w_{ij}} = \frac{\partial J(w_{ij})}{\partial O_j}\frac{\partial O_j}{\partial I_j}\frac{\partial I_j}{\partial w_{ij}} = \mathrm{Grad}_{ij}, \quad (4)$$

trong đó:

$$\frac{\partial I_j}{\partial w_{ij}} = \frac{\partial (O_i.w_{ij})}{\partial w_{ij}} = O_i; \quad (5)$$

Ký hiệu delta (δ) của nơ ron *j*:

$$\delta_j = \frac{\partial J(w_{ij})}{\partial O_j}.\frac{\partial O_j}{\partial I_j} = \frac{\partial J(w_{ij})}{\partial O_j}.F'(I_j). \quad (6)$$

Khi đó: $J'(w_{ij}) = \delta_j . O_i$. (7)

Nếu hàm kích hoạt F là hàm sigmoid, ta có [1]:

$$F'(I_j) = F(I_j)(1 - F(I_j)) = O_j.(1- O_j). \quad (8)$$

Nếu *j* là nơ ron lớp ra, từ công thức (3), ta có:

$$\frac{\partial J(w_{ij})}{\partial O_j} = O_j - y_j. \quad (9)$$

Như vậy, $\delta_j = (O_j - y_j).O_j.(1- O_j)$. (10)

Đối với trường hợp j là nơ ron của lớp ẩn. Gọi L - là số nơ ron của lớp kế tiếp nơ ron j – lớp gần hơn với lớp ra của mạng. Delta của nơ ron j xác định theo công thức [2]:

$$\delta_j = F'(I_j).\sum_{l=1}^{L}(\delta_l . w_{jl}) = O_j.(1- O_j).\sum_{l=1}^{L}(\delta_l . w_{jl}), \quad (11)$$

trong đó, δ_l – là delta của nơ ron l trong lớp nơ ron kế tiếp nơ ron j, w_{jl} – là trọng số của khớp nối giữa hai nơ ron j và l.

Trong các bài báo tiếp theo, tác giả sẽ giới thiệu chi tiết các thuật toán huấn luyện mạng nơ ron nhân tạo cùng với các ví dụ cụ thể, sau đó là các thuật toán ứng dụng mạng nơ ron nhân tạo để giải quyết một số bài toán trong ngành tài nguyên và môi trường.