Họ và tên: Mai Minh Quân

MSSV: 20225661

Nguồn tham khảo: Neural Networks and Deep Learning - Michael Nielsen

I.Nguyên nhân gây chậm tiến độ học trong mạng nơ-ron

1.Giới thiệu

Trong quá trình học của mạng nơ-ron, việc thay đổi trọng số và bias được điều chỉnh thông qua các đạo hàm riêng của hàm chi phí, vì vậy đạo hàm riêng của hàm chi phí đối với trọng số () và bias () quyết định tốc độ học của mạng nơ-ron. Khi nói "học chậm", thực chất là các đạo hàm này có giá trị rất nhỏ.

Dựa vào quy tắc cập nhật cho trọng số () và bias () trong thuật toán gradient descent:

Khi các đạo hàm riêng và rất nhỏ thì trọng số và bias sẽ cập nhật không đáng kể, làm chậm quá trình học.

Việc học chậm này không chỉ xảy ra trong các mô hình đơn giản mà còn trong các mạng nơ-ron phức tạp hơn. Đây là một vấn đề quan trọng cần giải quyết trong việc tối ưu hóa mạng nơ-ron.

2.Hàm Mean Squared Error

Xét hàm chi phí Mean Squared Error được định nghĩa là:

Trong đó là đầu ra của nơ-ron khi là đầu vào, và là đầu ra mong muốn.  
Tính toán các đạo hàm, với ta có:

Hàm sigmoid () khi đầu ra của nơ-ron gần với (hoặc ), đạo hàm trở nên rất nhỏ. Điều này dẫn đến các đạo hàm riêng và cũng rất nhỏ, quá trình cập nhật trọng số và bias trở nên chậm

II. Hàm cross entropy

1. Giới thiệu hàm cross entropy

Như đã thảo luận phía trên, việc học của mạng nơ-ron có thể bị chậm lại do các đạo hàm riêng của hàm chi phí trở nên quá nhỏ. Một cách để giải quyết vấn đề này là thay thế hàm mean squared error bằng một hàm chi phí khác, được gọi là hàm chi phí cross-entropy (cross-entropy). Hàm chi phí cross-entropy giúp cải thiện tốc độ học của mạng nơ-ron, tránh được sự chậm tiến độ học và mang lại những đặc điểm ưu việt cho quá trình huấn luyện.

2. Định nghĩa hàm chi phí Cross-Entropy

Hàm chi phí cross-entropy được định nghĩa là:

Trong đó:

* + là tổng số dữ liệu huấn luyện.
  + là đầu vào huấn luyện.
  + là đầu ra mong muốn.
  + là đầu ra thực tế của mạng nơ-ron.

3. Tính toán đạo hàm trong hàm chi phí Cross-Entropy

Để cập nhật trọng số trong quá trình học, ta cần tính đạo hàm của hàm chi phí cross-entropy đối với trọng số. Với đạo hàm này được tính như sau:

Quy đồng và đơn giản hóa, ta có:

Sau khi thay thế đạo hàm của hàm sigmoid , ta có thể rút gọn được biểu thức đạo hàm đối với trọng số:

Tương tự ta có đạo hàm đối với bias:

4. Ưu điểm của Cross-Entropy

- Giảm chậm tiến độ học:

Khi sử dụng hàm mean squared error, tốc độ học có thể bị chậm lại khi nơ-ron bị bão hòa khiến tiến về làm cho các đạo hàm riêng trở nên rất nhỏ. Với hàm chi phí cross-entropy, yếu tố trong các đạo hàm bị triệt tiêu, giúp tránh hiện tượng chậm tiến độ học này.

- Giảm dần chi phí khi nơ-ron học tốt hơn:  
Nếu đầu ra thực tế của nơ-ron gần với đầu ra mong muốn cho tất cả các đầu vào huấn luyện, hàm chi phí sẽ tiến gần về . Ví dụ, khi và , hoặc và , giá trị của hàm chi phí cross-entropy sẽ tiến về .

5.Khi nào nên sử dụng hàm chi phí Cross-Entropy thay vì hàm Mean Squared Error?

Hàm chi phí cross-entropy thường là lựa chọn tốt hơn khi các nơ-ron đầu ra sử dụng hàm sigmoid. Khi khởi tạo mạng nơ-ron, nếu trọng số và bias được thiết lập ngẫu nhiên, có thể xảy ra tình huống đầu ra của nơ-ron bị bão hòa (gần 1 hoặc 0), điều này làm chậm quá trình học nếu sử dụng hàm Mean Squared Error, vì đạo hàm của hàm chi phí sẽ rất nhỏ.

Trong khi đó, cross-entropy xử lý tốt các tình huống này, giúp mạng học nhanh và hiệu quả hơn. Vì vậy, cross-entropy là lựa chọn ưu tiên để tránh học chậm và cải thiện hiệu suất học trong mạng nơ-ron.