Họ và tên: Mai Minh Quân

MSSV: 20225661

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

1. Giới Thiệu về Regularization

Một cách để giảm thiểu overfitting trong mạng nơ-ron là sử dụng các kỹ thuật Regularization. Trong khi việc tăng kích thước dữ liệu huấn luyện là một giải pháp, thì việc giảm kích thước mạng hoặc sử dụng các kỹ thuật regularization có thể giúp giảm overfitting ngay cả khi mạng và dữ liệu huấn luyện đã được cố định.

1. L2 Regularization (Weight Decay)

Một trong những kỹ thuật regularization phổ biến là L2 Regularization, còn được gọi là Weight decay. Ý tưởng của L2 regularization là thêm một đại lượng bổ sung vào hàm chi phí, được gọi là , nhằm ngăn chặn các trọng số trở nên quá lớn. Điều này được thực hiện bằng cách thêm vào tổng bình phương của tất cả các trọng số trong mạng.

Hàm chi phí Cross-entropy có Regularization được tính như sau:

Trong đó:

* + là tham số regularization, điều khiển mức độ ảnh hưởng của Regularization.
  + là kích thước của tập huấn luyện.
  + là tổng các bình phương trọng số trong mạng, giúp hạn chế giá trị của các trọng số.

Regularization có thể được áp dụng cho các hàm chi phí khác ngoài Cross-entropy, như hàm Mean Squared Error. Công thức regularization cho hàm Mean Squared Error được viết như sau:

Trong cả hai trường hợp, ta có thể viết lại hàm chi phí với Regularization là:

Trong đó là hàm chi phí chưa áp dụng Regularization.

1. Lợi ích của Regularization

Việc áp dụng regularization giúp giảm overfitting bằng cách ngăn các trọng số trở nên quá lớn và giúp cho mô hình học được các đặc điểm tổng quát thay vì học quá chi tiết từ dữ liệu huấn luyện.

Regularization giúp giảm bớt số lượng nơ-ron trong các lớp ẩn, hoặc loại bỏ ảnh hưởng của một số nơ-ron, từ đó làm giảm độ phức tạp và tăng tính tuyến tính của mô hình, giúp mô hình tổng quát tốt hơn.

Khi áp dụng L2 regularization, trọng số sẽ được rescale bằng cách nhân với một đại lượng . Điều này có tên gọi là weight decay, giúp làm giảm giá trị của các trọng số theo thời gian trong quá trình huấn luyện.

Ví dụ: Trong ví dụ hồi quy, nếu mô hình quá phức tạp với nhiều tham số, nó sẽ khớp chính xác với từng điểm dữ liệu trong tập huấn luyện. Tuy nhiên, điều này không giúp mô hình tổng quát tốt khi gặp dữ liệu mới. Thay vào đó, một mô hình đơn giản, với ít tham số hơn, sẽ tạo ra một đường cong mượt mà hơn và tổng quát tốt hơn.

1. Áp dụng Regularization trong Gradient Descent

Khi áp dụng L2 regularization trong thuật toán gradient descent, ta sẽ tính toán các đạo hàm từng trọng số và cập nhật chúng theo công thức:

Trong đó, là hàm chi phí không có regularization. Đối với trọng số, ta sẽ cập nhật theo công thức:

Đây chính là công thức cập nhật với weight decay.

Đối với bias, ta sẽ cập nhật theo công thức:

1. Stochastic Gradient Descent với Regularization

Trong quá trình huấn luyện mạng nơ-ron có sử dụng Regularization (L2 Regularization), ta sẽ áp dụng Stochastic Gradient Descent (SGD) để cập nhật trọng số dựa trên chi phí đã được regularized. Công thức cho việc cập nhật trọng số và bias trong trường hợp này có sự thay đổi so với phương pháp không regularization.

Trọng số (w) được cập nhật như sau:

Trong đó:

* là learning rate (tỷ lệ học).
* là tham số regularization.
* là số lượng ví dụ trong tập huấn luyện.
* là kích thước của mini-batch.
* là hàm chi phí chưa được áp dụng Regularization cho mỗi ví dụ huấn luyện trong mini-batch.

Quy tắc cập nhật cho biases không thay đổi khi có regularization, công thức được giữ nguyên:

Tương tự như với trọng số, là hàm chi phí chưa được áp dụng Regularization cho mỗi ví dụ huấn luyện trong mini-batch.