Họ và tên: Mai Minh Quân

MSSV: 20225661

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

1. Giới thiệu về Overfitting

Overfitting là hiện tượng khi mô hình học quá chi tiết và cụ thể từ dữ liệu huấn luyện mà không thể tổng quát hóa tốt đối với dữ liệu kiểm tra. Khi xảy ra overfitting, mô hình có thể đạt độ chính xác rất cao trên dữ liệu huấn luyện nhưng lại không hiệu quả khi áp dụng trên dữ liệu mới, chưa thấy.

Trong mạng nơ-ron, overfitting thường xảy ra khi mô hình có quá nhiều tham số và được huấn luyện với một bộ dữ liệu huấn luyện nhỏ. Mô hình có thể học thuộc các đặc điểm ngẫu nhiên trong dữ liệu huấn luyện, từ đó không còn khả năng tổng quát cho các dữ liệu mới.

2. Dấu hiệu của Overfitting

Một dấu hiệu rõ ràng của overfitting là khi độ chính xác trên dữ liệu huấn luyện (training data) tăng rất cao, có thể đạt đến 100%, trong khi độ chính xác trên dữ liệu kiểm tra (validation data) lại không cải thiện hoặc chỉ đạt mức rất thấp.

Nếu độ chính xác trên dữ liệu huấn luyện (training data) tiếp tục tăng mà không có sự cải thiện tương ứng trên dữ liệu kiểm tra, mô hình có thể đang học những chi tiết ngẫu nhiên không có ích.

Ví dụ, khi sử dụng 1.000 hình ảnh huấn luyện, mô hình có thể đạt độ chính xác 100% trên dữ liệu huấn luyện, nhưng chỉ đạt khoảng 82.27% trên dữ liệu kiểm tra, cho thấy mô hình chỉ học thuộc lòng dữ liệu huấn luyện mà không tổng quát hóa được.

3. Phát hiện Overfitting

Một cách để phát hiện overfitting là theo dõi độ chính xác trên dữ liệu kiểm tra (validation data) trong suốt quá trình huấn luyện. Nếu độ chính xác trên dữ liệu kiểm tra không còn cải thiện hoặc thậm chí giảm sau một thời gian, đó là dấu hiệu cho thấy mô hình đang overfit.

4. Giải quyết Overfitting bằng cách tăng kích thước dữ liệu huấn luyện

Một trong những phương pháp hiệu quả để giảm thiểu overfitting là tăng kích thước dữ liệu huấn luyện. Khi mô hình được huấn luyện trên một tập dữ liệu đủ lớn và đa dạng, nó có thể tổng quát tốt hơn và giảm khả năng overfitting.

Mặc dù tăng kích thước dữ liệu huấn luyện là một chiến lược hiệu quả để giảm thiểu overfitting, nhưng trong thực tế, việc thu thập dữ liệu huấn luyện đủ lớn có thể rất tốn kém và khó thực hiện. Do đó, không phải lúc nào cũng khả thi để thu thập một lượng dữ liệu lớn.

5. Các chiến lược khác để giảm thiểu Overfitting

Ngoài việc tăng kích thước dữ liệu huấn luyện, một số kỹ thuật cũng có thể giúp giảm thiểu overfitting và cải thiện khả năng tổng quát của mô hình như:

* Regularization (điều chỉnh trọng số của mạng)
* Early stopping (ngừng huấn luyện khi độ chính xác trên dữ liệu kiểm tra không còn cải thiện)

Hãy sửa lại phần mô tả của regularization trong đoạn mô tả sau

Ngoài việc tăng kích thước dữ liệu huấn luyện, một số kỹ thuật cũng có thể giúp giảm thiểu overfitting và cải thiện khả năng tổng quát của mô hình như:

• Regularization (kỹ thuật điều chỉnh trọng số của mạng để hạn chế độ phức tạp của mô hình, giúp tránh việc mô hình học quá chi tiết vào dữ liệu huấn luyện và cải thiện khả năng tổng quát trên dữ liệu mới)

• Early stopping (ngừng huấn luyện khi độ chính xác trên dữ liệu kiểm tra không còn cải thiện)