**Chương 1: Các phương pháp Optimizer trong huấn luyện mô hình học máy**

1. Các phương pháp Optimizer

Trong huấn luyện mô hình học máy, các phương pháp Optimizer là các thuật toán dùng để cập nhật trọng số của mô hình sao cho giảm thiểu hàm mất mát. Dưới đây là một số phương pháp Optimizer. Mỗi phương pháp đều có ưu điểm và nhược điểm riêng.

1. Gradient Descent (GD)

Định nghĩa: Là phương pháp tìm giá trị nhỏ nhất của hàm mất mát bằng cách di chuyển theo hướng ngược với gradient của hàm mất mát trên toàn bộ tập dữ liệu.

Ưu điểm: Đơn giản ,dễ hiểu và dễ triển khai. Có thể hội tụ đến điểm cực tiểu toàn cục nếu chọn đúng learning rate. Phù hợp cho các bài toán có dữ liệu nhỏ hoặc không yêu cầu tính toán phức tạp.

Nhược điểm: Có thể bị mắc kẹt ở điểm cực tiểu của cục bộ, hội tụ chậm và không ổn định. Không phù hợp với bài toán dữ liệu lớn và nhiều chiều.

1. Stochastic Gradient Descent (SGD)

Định nghĩa: Là bản cải tiến của GD, trong đó mỗi bước cập nhật trọng số chỉ sử dụng một dữ liệu ngẫu nhiên thay vì toàn bộ tập dữ liệu.

Ưu điểm: Hiệu quả hơn so với GD với dữ liệu lớn, giúp tránh việc kẹt ở điểm tối ưu cục bộ và điểm yên ngựa.

Nhược điểm: Do sử dụng dữ liệu ngẫu nhiên nên có thể không hội tụ chính xác như GD, có thể dao động mạnh và không ổn định.

1. Mini-batch Gradient Descent

Định nghĩa: Là phiên bản kết hợp giữa GD và SGD, cập nhật trọng số dựa trên một lượng nhỏ điểm dữ liệu thay vì toàn bộ tập dữ liệu.

Ưu điểm: Có thể hội tụ nhanh hơn và ổn định hơn GD và SGD.

Nhược điểm: Cần chọn kích thước batch phù hợp.

1. Momentum

Định nghĩa: Là phương pháp sử dụng lịch sử các gradient trước đó để cập nhật trọng số.

Ưu điểm: Tăng tốc hội tụ và ổn định hơn, có thể vượt qua các vùng bằng phẳng và hẹp của hàm mất mát.

Nhược điểm: Cần chọn momentum phù hợp.

1. Adaptive Gradient (AdaGrad)

Định nghĩa: Là phương pháp thích ứng và điều chỉnh learning rate dựa trên lịch sử các gradient trước đó của nó.

Ưu điểm: Thích hợp với các bài toán có dữ liệu thưa hay có đặc trưng riêng biệt. Tự động điều chỉnh learning rate giúp hội tụ nhanh hơn.

Nhược điểm: Có thể bị giảm tốc độ học và dừng sớm quá trình hội tụ, khó xử lý với các bài toán có dữ liệu lớn và nhiều chiều.

1. Root Mean Sqaure Propagation (RMSProp)

Định nghĩa: Là phương pháp cải tiến của AdaGrad, sử dụng trung bình động của bình phương gradient để điều chỉnh learning rate cho từng trọng số.

Ưu điểm: Hội tụ nhanh hơn AdaGrad.

Nhược điểm: Cần điều chỉnh tham số để tránh learnung rate giảm quá nhanh

1. Adaptive Moment Estimation (Adam)

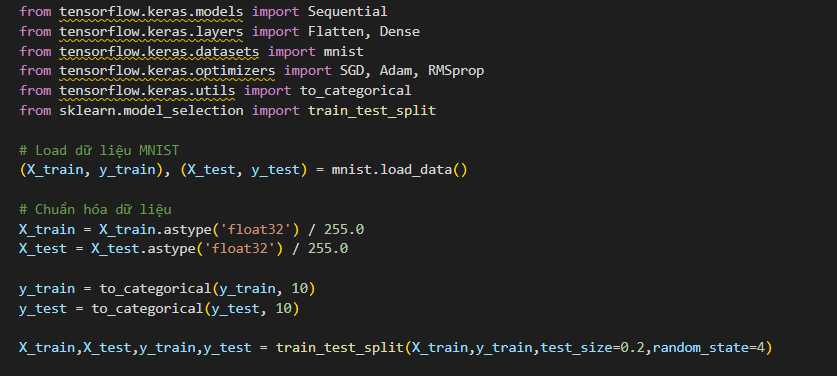
Định nghĩa: Là sự kết hợp giữa Momentum và RMSProp, sử dụng trung bình động của gradient và bình phương gradient để cập nhật trọng số.

Ưu điểm: Hội tụ nhanh và ổn định. Tự động điều chỉnh learning rate cho mỗi tham số.

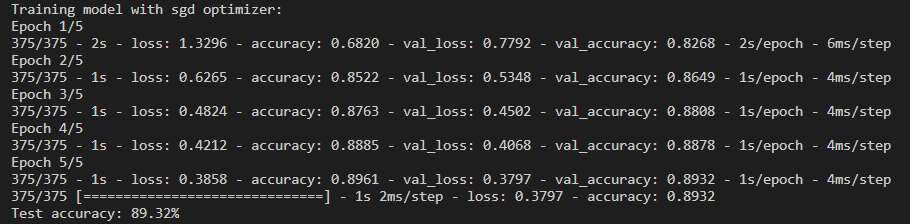
Nhược điểm: Cần điều chỉnh tham số phù hợp, có thể bị sai lệch khi gradient quá nhỏ và có thể bị kẹt ở điểm tối ưu cục bộ.

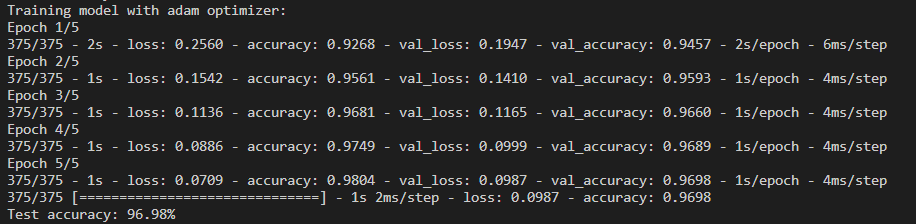
Nhìn chung, một số optimizer có khả năng điều chỉnh learning rate tự động, các biến thể có khả năng điều chinh momentum và learning rate tại mỗi bước cập nhật, mỗi optimizer có ưu điểm riêng trong việc hội tụ nhanh chóng và tránh khỏi bị kẹt ở các điểm cực tiểu cục bộ.

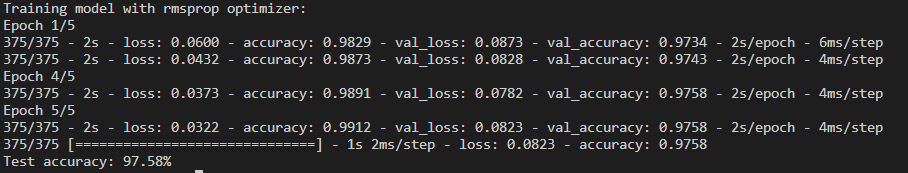
1. SGD, Adam, RMSProp







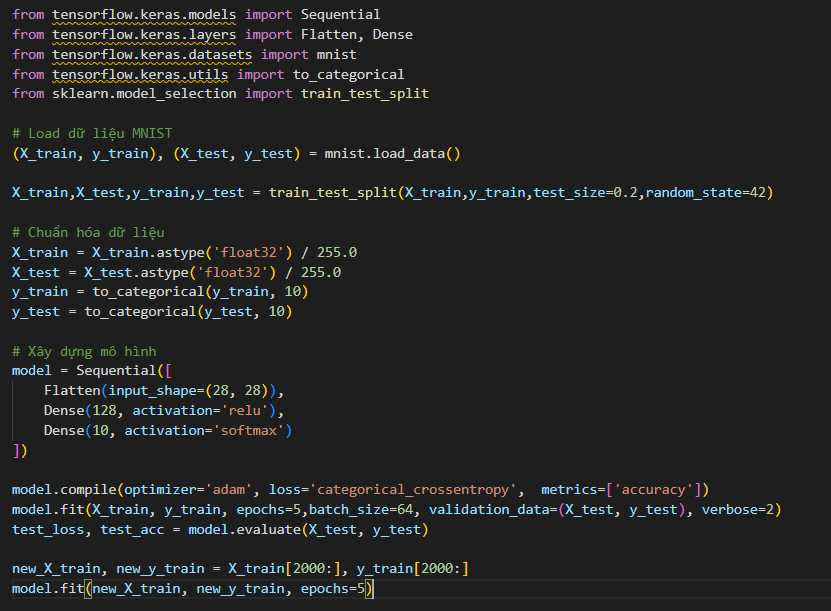


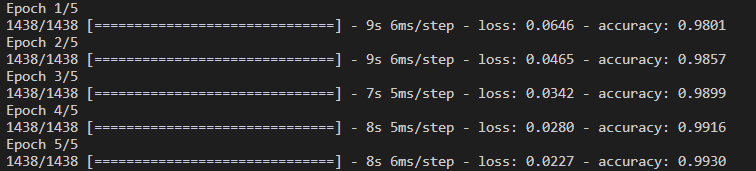


**Chương 2: Continual Learning và Test Production khi xây dựng giải pháp học máy giải quyết bài toán**

1. Continual Learning (CL)

Continual Learning (CL) là khả năng mô hình học và cập nhật liên tục các dữ liệu mới theo thời gian mà vẫn không quên đi những dữ liệu cũ. CL có thể gặp vấn đề như catastrophic forgetting – học dữ liệu mới có khả năng làm mất dữ liệu cũ.





1. Test Production

Test Production là quá trình kiểm tra và sản xuất dữ liệu kiểm tra nhằm đảm bảo mô hình học máy hoạt động hiệu quả và đáng tin cậy.

