



**TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA CÔNG NGHỆ THÔNG TIN**



BÁO CÁO GIỮA KỲ
HỌC PHẦN: MẠNG NƠ RON & HỌC SÂU

Sinh viên thực hiện	MSSV	Lớp sinh hoạt
Nguyễn Quang Sáng	102210305	21TCLC_KHDL

ĐÀ NẴNG, 04/2024

MỤC LỤC

Nội dung

MỤC LỤC	2
MỞ ĐẦU	3
CƠ SỞ LÝ THUYẾT	4
1. Giới thiệu.....	4
2. Bộ dữ liệu.....	4
3. Mô hình MLP.....	5
4. Các hàm kích hoạt.....	7
5. Chuẩn hóa dữ liệu.....	10
6. Khởi tạo trọng số.....	13
7. Kết luận và đánh giá.....	15

MỞ ĐẦU

Trong bối cảnh hiện đại, việc khám phá và tối ưu hóa các mô hình học sâu đã trở nên thiết yếu trong việc phát triển các ứng dụng trí tuệ nhân tạo. Mục tiêu của chúng tôi trong loạt thực nghiệm này là để đánh giá hiệu suất của mô hình Perceptron Đa Lớp (MLP) khi áp dụng các hàm kích hoạt khác nhau, các phương pháp chuẩn hóa dữ liệu, và các chiến lược khởi tạo trọng số. Cụ thể, em sẽ:

1. Hàm kích hoạt: Thực nghiệm hàm kích hoạt Tanh trên bộ dữ liệu Fashion-MNIST và Cifar10, và so sánh kết quả với khi sử dụng hàm Sigmoid, nhằm mục đích xác định hàm kích hoạt nào mang lại độ chính xác cao hơn cho mô hình.
2. Chuẩn hóa dữ liệu: Áp dụng các phương pháp chuẩn hóa dữ liệu khác nhau trên một mô hình cố định, để phân tích ảnh hưởng của chúng đối với khả năng học và hiệu suất tổng thể của mô hình.
3. Khởi tạo trọng số: Thực hiện thực nghiệm với ba loại khởi tạo trọng số - Gaussian với trung bình là 0 và độ lệch chuẩn là 0.01, Glorot, và He - để đánh giá tác động của chúng đối với quá trình học và độ chính xác của mô hình.

Thông qua thực nghiệm này, em mong muốn không chỉ xác định các yếu tố tối ưu cho mô hình MLP mà còn góp phần vào kho tàng kiến thức chung về học sâu, từ đó hỗ trợ cộng đồng nghiên cứu phát triển các mô hình học máy mạnh mẽ và hiệu quả hơn.

Nhân đây, em xin chân thành cảm ơn thầy **Phạm Minh Tuấn** đã tận tình hướng dẫn, chỉ bảo em hoàn thành tốt dự án lần này.

Em xin chân thành cảm ơn!

CƠ SỞ LÝ THUYẾT

1. Giới thiệu:

Báo cáo này trình bày kết quả thực nghiệm với mô hình mạng nơ-ron đa lớp (MLP) cho hai tập dữ liệu: Fashion-MNIST và Cifar10. Các thí nghiệm được thực hiện với các mục tiêu sau:

- So sánh hiệu quả của hàm kích hoạt Sigmoid và Tanh.
- Đánh giá tác động của các phương pháp chuẩn hóa dữ liệu khác nhau.
- So sánh hiệu quả của các phương pháp khởi tạo trọng số khác nhau.

2. Bộ dữ liệu

2.1. Bộ dữ liệu Fashion-MNIST

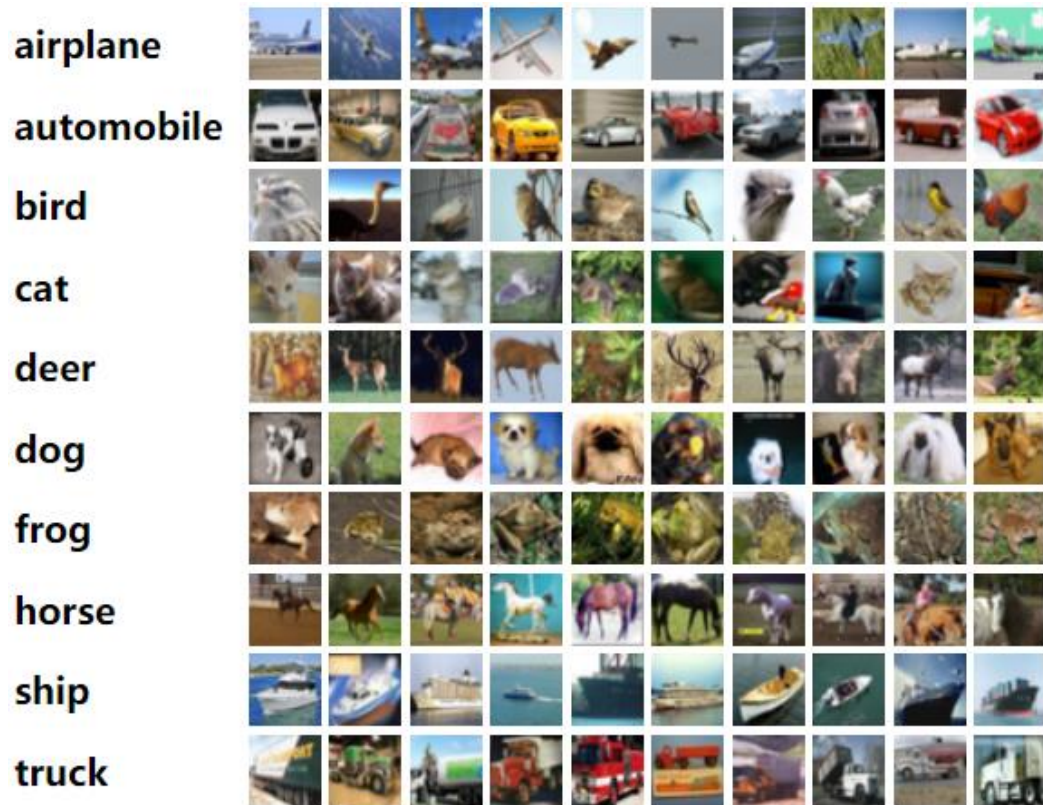
Fashion-MNIST là một bộ dữ liệu bao gồm các hình ảnh thời trang từ Zalando, được tạo ra nhằm mục đích cung cấp một phiên bản thay thế cho bộ dữ liệu MNIST gốc về các chữ số viết tay. Bộ dữ liệu này bao gồm một tập huấn luyện với 60,000 ví dụ và một tập kiểm tra với 10,000 ví dụ. Mỗi ví dụ là một hình ảnh xám cỡ 28x28 pixel, đi kèm với nhãn từ 10 lớp phân loại khác nhau.



Hình 1: Các lớp trong bộ dữ liệu Fashion-MNIST

2.2. Bộ dữ liệu CIFAR-10

Bộ dữ liệu CIFAR-10, được phát triển bởi Alex Krizhevsky, Vinod Nair, và Geoffrey Hinton, là một tập hợp các hình ảnh màu nhỏ với kích thước 32x32 pixel. Bộ dữ liệu này bao gồm 60,000 hình ảnh thuộc 10 lớp phân loại khác nhau, với mỗi lớp chứa 6,000 hình ảnh¹. Trong số đó, có 50,000 hình ảnh được sử dụng cho việc huấn luyện và 10,000 hình ảnh dành cho việc kiểm tra.



Hình 2: Các lớp trong bộ dữ liệu CIFAR-10

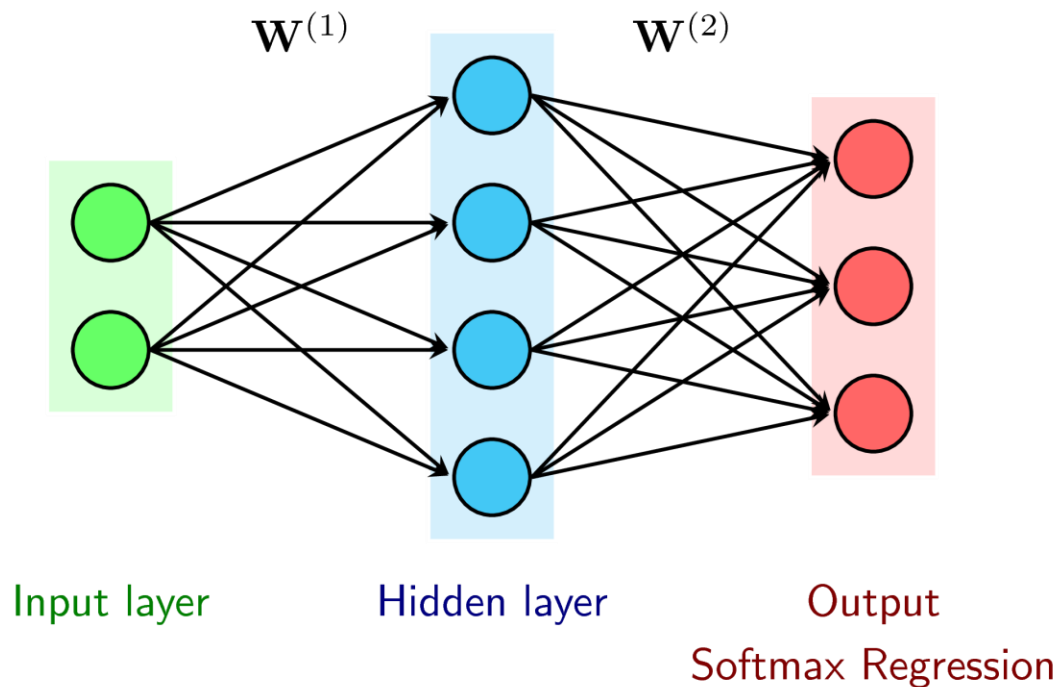
3. Mô hình MLP:

Mô hình Perceptron Đa Lớp (MLP) là một kiến trúc mạng nơ-ron nhân tạo phổ biến trong học máy. MLP là một mạng nơ-ron tiến hóa, bao gồm nhiều lớp nơ-ron liên kết với nhau, bao gồm một lớp đầu vào, một hoặc nhiều lớp ẩn, và một lớp đầu ra. Mỗi nơ-ron trong một lớp nhận đầu vào từ các nơ-ron của lớp trước đó, thực hiện một phép biến đổi tuyến tính (như tổng trọng số của đầu vào và bias), sau đó áp dụng một hàm kích hoạt phi tuyến để tạo ra đầu ra cho lớp tiếp theo.

Các đặc điểm chính của MLP bao gồm:

- **Khả năng học các mối quan hệ phi tuyến:** Nhờ vào hàm kích hoạt phi tuyến, MLP có thể học được các mối quan hệ phức tạp giữa đầu vào và đầu ra.
- **Cấu trúc lớp:** MLP có ít nhất ba lớp gồm lớp đầu vào, lớp ẩn, và lớp đầu ra. Số lượng lớp ẩn và nơ-ron trong mỗi lớp có thể thay đổi tùy thuộc vào vấn đề cụ thể.

- **Học có giám sát:** MLP thường được huấn luyện sử dụng phương pháp học có giám sát, nghĩa là mô hình được huấn luyện trên một tập dữ liệu có nhãn.
- **Lan truyền tiến và lan truyền ngược:** Trong quá trình huấn luyện, thông tin được lan truyền từ lớp đầu vào qua các lớp ẩn và đến lớp đầu ra (lan truyền tiến), và sau đó lỗi được tính toán và lan truyền ngược trở lại để cập nhật trọng số.



Hình 3: Mô hình MLP

MLP được sử dụng rộng rãi trong các bài toán phân loại, hồi quy, và nhận dạng mẫu. Nó là một công cụ mạnh mẽ và linh hoạt, có thể được tinh chỉnh để giải quyết nhiều loại vấn đề khác nhau trong học máy và thị giác máy tính.

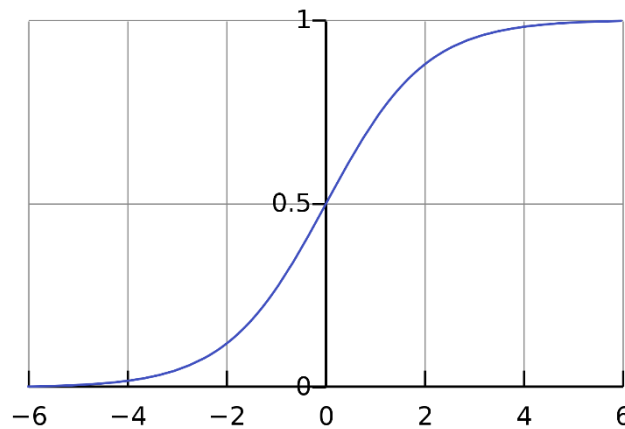
4. Các hàm kích hoạt:

Hàm kích hoạt là một hàm được sử dụng trong mạng nơ-ron nhân tạo (ANN) để biến đổi đầu vào của một nơ-ron thành đầu ra của nó. Nó đóng vai trò quan trọng trong việc xác định hành vi của nơ-ron và ảnh hưởng đến hiệu quả của toàn bộ mạng nơ-ron.

4.1. Hàm Sigmoid

Hàm Sigmoid (hay còn được gọi là hàm logistic hoặc hàm squash) là một hàm toán học ánh xạ các số thực vào các giá trị nằm trong khoảng từ 0 đến 1. Biểu diễn đồ thị của nó (Hình 2) là một đường cong hình chữ S, do đó có tên là “sigmoid”. Với x là giá trị đầu vào, giá trị của hàm Sigmoid được tính qua công thức:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

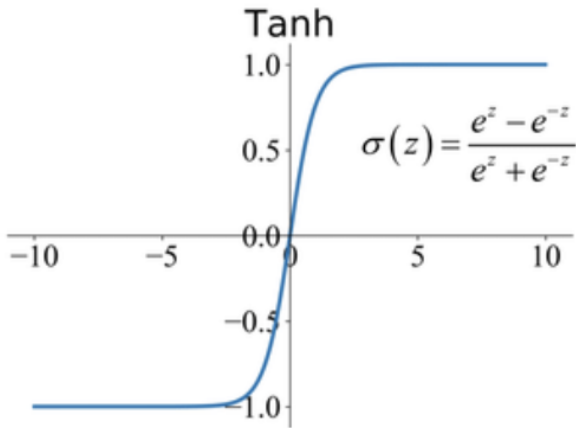


Hình 4: Công thức và đồ thị biểu diễn hàm Sigmoid

4.2. Hàm Tanh

Hàm Tanh (hay còn gọi là hàm tan hyperbolic) là một hàm toán học có đồ thị (Hình 3) là một đường cong hình chữ S, tương tự như hàm Sigmoid. Tuy nhiên, hàm Tanh có giá trị nằm trong khoảng từ -1 đến 1. Với x là giá trị đầu vào, giá trị của hàm Tanh được tính qua công thức:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



Hình 5: Công thức và đồ thị biểu diễn hàm Tanh

4.3. So sánh hàm Sigmoid và Tanh

- **Phạm vi đầu ra:** Tanh nén đầu vào vào phạm vi từ -1 đến 1, trong khi Sigmoid nén vào phạm vi từ 0 đến 1. Điều này có nghĩa là Tanh có đầu ra trung bình gần với 0 hơn, điều này thường giúp việc huấn luyện mạng nơ-ron nhanh hơn do hiệu ứng chuẩn hóa.
- **Đạo hàm:** Cả hai hàm đều có đạo hàm xác định trên toàn bộ miền số thực, nhưng đạo hàm của Tanh có thể có giá trị âm, trong khi đạo hàm của Sigmoid luôn dương.
- **Sự bão hòa:** Cả hai hàm đều có vấn đề về sự bão hòa khi đầu vào có giá trị lớn hoặc nhỏ, dẫn đến việc gradient biến mất, một vấn đề có thể làm chậm quá trình học hoặc khiến mạng không học được.

4.4. Tiến hành thực nghiệm:

4.4.1. Quá trình huấn luyện:

- **Chuẩn bị dữ liệu:** Dữ liệu từ tập CIFAR-10 và Fashion-MNIST được tải xuống và chia thành tập huấn luyện (train_images, train_labels) Chuẩn bị dữ liệu: Dữ liệu từ tập CIFAR-10 và Fashion-MNIST được tải xuống và chia thành tập huấn luyện (train_images, train_labels) và tập kiểm tra (test_images, test_labels). Dữ liệu hình ảnh sau đó được chuẩn hóa để có giá trị từ 0 đến 1 bằng cách chia cho 255.0.
- **Xây dựng mô hình:** Mô hình được xây dựng dưới dạng một mạng nơ-ron hoàn toàn kết nối (Fully Connected Neural Network) với 2 hoặc 5 lớp ẩn, mỗi lớp có 128 nơ-ron. Hàm kích hoạt Tanh được sử dụng ở mỗi lớp ẩn. Lớp cuối cùng sử dụng hàm kích hoạt softmax để chuyển đổi đầu ra thành các xác suất.
- **Biên dịch mô hình:** Mô hình sau đó được biên dịch với thuật toán tối ưu hóa 'adam', hàm mất mát 'sparse_categorical_crossentropy' (phù hợp cho bài toán phân loại nhiều lớp với nhãn dưới dạng số nguyên), và 'accuracy' là chỉ số đánh giá mô hình.

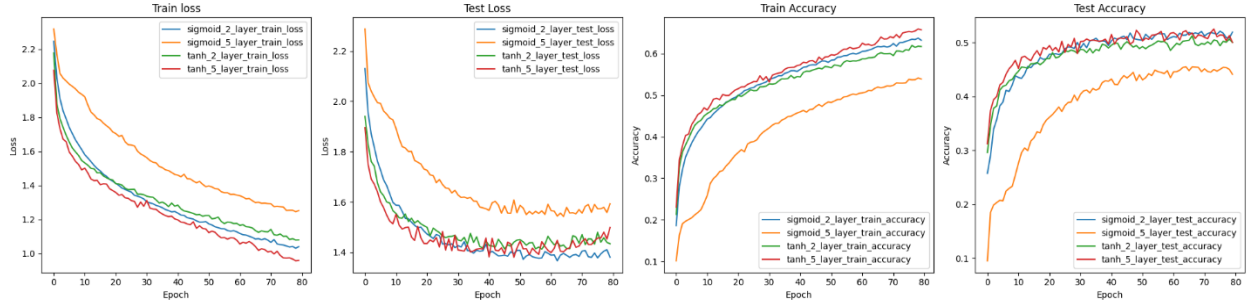
- **Huấn luyện mô hình:** Mô hình được huấn luyện trên tập huấn luyện với kích thước batch là 1024 và số epoch là 80. Trong mỗi epoch, mô hình sẽ học từ tất cả các mẫu trong tập huấn luyện. Tập kiểm tra được sử dụng như một tập validation để theo dõi độ chính xác của mô hình sau mỗi epoch.

4.4.2. Kết quả thực nghiệm:

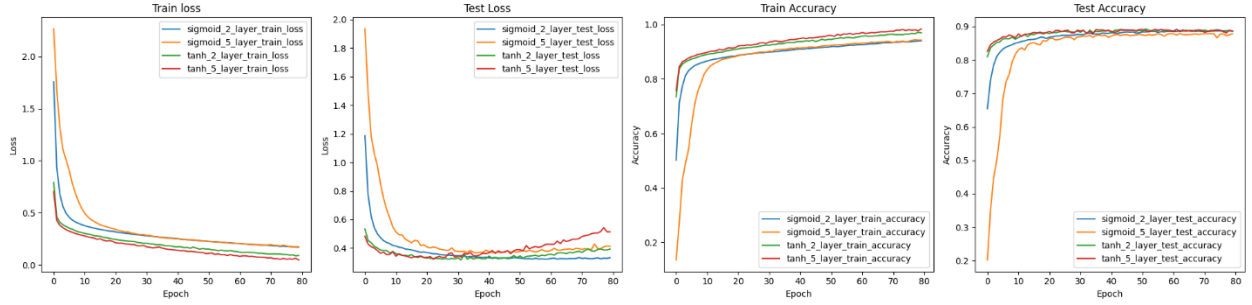
Bộ dữ liệu	Hàm kích hoạt	Số lượng lớp ẩn	Độ chính xác trên tập huấn luyện	Độ chính xác trên tập kiểm thử
CIFAR10	Sigmoid	2	0.6324	0.5194
		5	0.5391	0.4411
	Tanh	2	0.6174	0.5003
		5	0.6576	0.5004
Fashion MNIST	Sigmoid	2	0.9401	0.8856
		5	0.9430	0.8783
	Tanh	2	0.9696	0.8867
		5	0.9840	0.8860

Bảng 1: Kết quả huấn luyện bằng các hàm kích hoạt

- **Hiệu suất của hàm kích hoạt Tanh:** Mô hình MLP với hàm kích hoạt Tanh và 5 lớp ẩn cho thấy hiệu suất tốt trên cả hai bộ dữ liệu, đặc biệt là trên tập kiểm thử CIFAR10 và Fashion MNIST với độ chính xác lần lượt là **0.6576** và **0.9840**. Điều này cho thấy Tanh có thể là một lựa chọn tốt khi cần cải thiện khả năng phân loại của mô hình.
- **Ảnh hưởng của số lượng lớp ẩn:** Tăng số lượng lớp ẩn thường giúp cải thiện khả năng phân loại, nhưng kết quả này không đồng đều trên tất cả các cấu hình. Đối với CIFAR10, việc tăng số lớp ẩn khi sử dụng hàm Sigmoid lại dẫn đến sự giảm độ chính xác, điều này có thể do việc mô hình trở nên quá phức tạp và overfitting.
- **So sánh giữa Sigmoid và Tanh:** Trên tập Fashion MNIST, mô hình sử dụng Tanh cho kết quả tốt hơn so với Sigmoid, với sự cải thiện gần 0.03, 0.04 tương ứng với 2 và 5 lớp ẩn về độ chính xác. Điều này cũng được phản ánh qua đồ thị hàm mất mát và độ chính xác, nơi Tanh cho thấy sự ổn định và hiệu suất tốt hơn trong các epoch đầu.
- **Biểu đồ hàm mất mát và độ chính xác:** Các biểu đồ cho thấy rằng mô hình với 2 lớp ẩn có đồ thị hàm mất mát và độ chính xác tương tự nhau khi sử dụng cả hai hàm Sigmoid và Tanh. Tuy nhiên, mô hình với 5 lớp ẩn sử dụng Tanh cho thấy sự cải thiện đáng kể sau khoảng 8 epoch, mặc dù sự tăng này không ổn định.



Hình 6: Đồ thị loss và accuracy trên bộ dữ liệu CIFAR-10



Hình 7: Đồ thị loss và accuracy trên bộ dữ liệu Fashion-MNIST

5. Chuẩn hóa dữ liệu:

Chuẩn hóa dữ liệu là một quá trình quan trọng trong việc xử lý và phân tích dữ liệu, đặc biệt là trong lĩnh vực học máy và khoa học dữ liệu. Mục tiêu chính của chuẩn hóa dữ liệu là đảm bảo rằng dữ liệu đầu vào cho các mô hình học máy được định dạng một cách nhất quán, giúp cải thiện hiệu suất và độ chính xác của mô hình. Hiện nay, có rất nhiều phương pháp chuẩn hóa phổ biến, trong đó em sẽ sử dụng 3 phương pháp để tiến hành thực nghiệm này

5.1. Các phương pháp chuẩn hóa:

5.1.1. Min-Max Scaling:

Min-Max Scaling, hay còn gọi là chuẩn hóa min-max, là một phương pháp chuẩn hóa dữ liệu phổ biến trong học máy. Phương pháp này biến đổi dữ liệu về dạng có giá trị nhỏ nhất bằng 0 và giá trị lớn nhất bằng 1. Trong đó x là giá trị khi ban đầu, x_{scaled} là giá trị sau khi chuẩn hóa, min là giá trị nhỏ nhất trong tập dữ liệu, ngược lại max là giá trị lớn nhất trong tập dữ liệu

$$x_{scaled} = \frac{x - min}{max - min}$$

Ưu điểm của Min-Max Scaling là đơn giản và dễ thực hiện, hiệu quả với dữ liệu có phân phối đều. Nhược điểm là nhạy cảm với các giá trị ngoại lai (outlier),...

5.1.2. Z-score Normalization (Standardization)

Chuẩn hóa Z-score là phương pháp chuẩn hóa dữ liệu phổ biến trong học máy. Phương pháp này biến đổi dữ liệu về phân bố có giá trị trung bình bằng 0 và độ lệch chuẩn bằng 1. Với dữ liệu đầu vào là x , giá trị trung bình và độ lệch chuẩn của tập dữ liệu là μ, σ , giá trị của dữ liệu sau khi được chuẩn hóa x_{scaled} được tính qua công thức:

$$x_{scaled} = \frac{x - \mu}{\sigma}$$

Trong thực nghiệm, để tránh mẫu số bằng 0 khi bộ dữ liệu này đồng nhất, với ε là một hằng số rất nhỏ, công thức trên được thay đổi thành:

$$x_{scaled} = \frac{x - \mu}{\sigma + \varepsilon}$$

5.1.3. Mean Normalization

Mean Normalization, hay còn gọi là chuẩn hóa trung bình, là một phương pháp chuẩn hóa dữ liệu đơn giản trong học máy. Phương pháp này biến đổi dữ liệu về dạng có giá trị trung bình bằng 0. Phương pháp đơn giản này nhìn quá có vẻ giống nhưng thực tế nó khác với Min-Max Scaling và Standardization.

$$x_{scaled} = \frac{x - \mu}{max - min}$$

5.2. Tiến hành chuẩn hóa:

5.2.1. Quá trình chuẩn hóa:

- **Chuẩn bị dữ liệu:** Dữ liệu từ tập CIFAR-10 được tải xuống và chia thành tập huấn luyện (train_images, train_labels) Chuẩn bị dữ liệu: Dữ liệu từ tập CIFAR-10 được tải xuống và chia thành tập huấn luyện (train_images, train_labels) và tập kiểm tra (test_images, test_labels). Dữ liệu hình ảnh sau đó được chuẩn hóa bằng ba phương pháp khác nhau: Z-score normalization, mean normalization, và min-max scaling. Mỗi phương pháp chuẩn hóa này

giúp cải thiện hiệu suất của mô hình bằng cách đưa các đặc trưng về cùng một khoảng giá trị, giúp quá trình huấn luyện diễn ra mượt mà và ổn định hơn.

- **Xây dựng mô hình:** Mô hình được xây dựng dưới dạng một mạng nơ-ron hoàn toàn kết nối (Fully Connected Neural Network) với 3 lớp ẩn, mỗi lớp có 128 nơ-ron. Hàm kích hoạt Sigmoid hoặc Tanh được sử dụng ở mỗi lớp ẩn. Lớp cuối cùng sử dụng hàm kích hoạt softmax để chuyển đổi đầu ra thành các xác suất.
- **Biên dịch mô hình:** Mô hình sau đó được biên dịch với thuật toán tối ưu hóa 'adam', hàm mất mát 'sparse_categorical_crossentropy' (phù hợp cho bài toán phân loại nhiều lớp với nhãn dưới dạng số nguyên), và 'accuracy' là chỉ số đánh giá mô hình.
- **Huấn luyện mô hình:** Mô hình được huấn luyện trên tập huấn luyện với kích thước batch là 1024 và số epoch là 50. Trong mỗi epoch, mô hình sẽ học từ tất cả các mẫu trong tập huấn luyện. Tập kiểm tra được sử dụng như một tập validation để theo dõi độ chính xác của mô hình sau mỗi epoch.

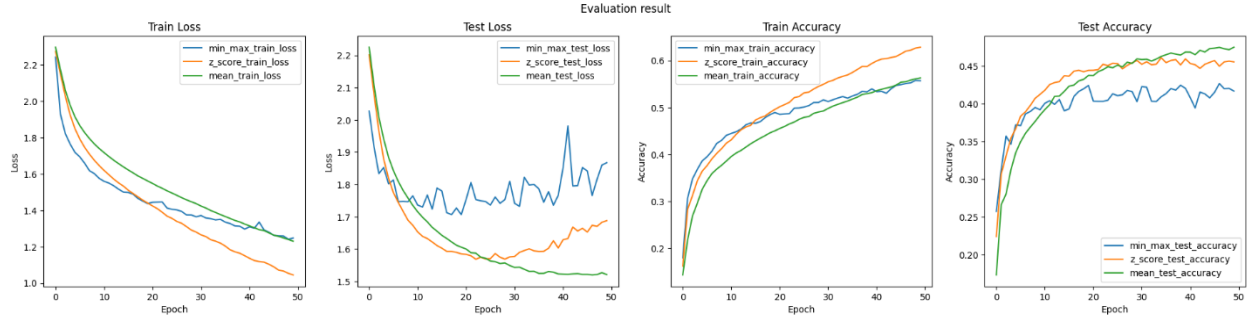
5.2.2. Kết quả huấn luyện

Phương pháp chuẩn hóa	Độ chính xác trên tập huấn luyện	Độ chính xác trên tập kiểm thử
Min-Max Scaling	0.5571	0.4166
Z-score Normalization	0.6284	0.4519
Mean Normalization	0.5630	0.4746

Bảng 2: Kết quả huấn luyện bằng các phương pháp chuẩn hóa

Dựa vào kết quả số liệu từ Bảng 2, chúng ta có thể thấy rằng có sự so sánh giữa ba phương pháp chuẩn hóa dữ liệu khác nhau: Min-Max Scaling, Z-score Normalization và Mean Normalization. Trong đó, Z-score Normalization cho thấy hiệu suất tốt nhất với độ chính xác là **0.6284** trên tập huấn luyện và **0.4519** trên tập kiểm thử. Phương pháp Min-Max Scaling có độ chính xác là **0.5571** trên tập huấn luyện và **0.4166** trên tập kiểm thử, trong khi Mean Normalization có độ chính xác là **0.5630** trên tập huấn luyện và **0.4746** trên tập kiểm thử. Dựa trên hình

Nhìn chung, có vẻ như Z-score Normalization là phương pháp ưu việt nhất, cung cấp một sự cân bằng tốt giữa hiệu suất trên tập huấn luyện và tập kiểm thử, điều này quan trọng để tránh hiện tượng overfitting. Tuy nhiên, Mean Normalization cũng cho thấy kết quả khá tốt trên tập kiểm thử, điều này có thể hữu ích trong một số trường hợp cụ thể. Để đưa ra quyết định cuối cùng, cần xem xét thêm các yếu tố như loại dữ liệu đang được xử lý và mục tiêu cụ thể của mô hình học máy.



Hình 8: Đồ thị loss và accuracy khi sử dụng các phương pháp chuẩn hóa

6. Khởi tạo trọng số

6.1. Tầm quan trọng của việc khởi tạo trọng số:

Khởi tạo trọng số là một bước quan trọng trong quá trình huấn luyện mô hình học máy. Nó ảnh hưởng trực tiếp đến hiệu quả và tốc độ hội tụ của mô hình. Khởi tạo trọng số ngẫu nhiên có thể dẫn đến mô hình học sai lệch hoặc không hội tụ. Việc khởi tạo giá trị ban đầu phù hợp cho trọng số giúp mô hình học chính xác hơn và hội tụ nhanh hơn. Khởi tạo trọng số tốt giúp mô hình học nhanh hơn và đạt được độ chính xác cao hơn trong thời gian ngắn hơn. Điều này đặc biệt quan trọng cho các mô hình phức tạp với nhiều dữ liệu.

Trong bài này, em sẽ sử dụng các phương pháp khởi tạo trọng số như sau:

6.1.1. Khởi tạo trọng số Gaussian

Khởi tạo trọng số Gaussian là phương pháp khởi tạo trọng số phổ biến trong các mô hình học sâu. Phương pháp này gán giá trị ngẫu nhiên cho các trọng số từ một phân phối chuẩn (Gaussian distribution). Với $\mathcal{N}(\mu, \sigma)$ là phân phối chuẩn có giá trị trung bình và độ lệch chuẩn lần lượt là μ và σ , trọng số w được khởi tạo qua công thức:

$$w = \mathcal{N}(\mu, \sigma)$$

6.1.2. Khởi tạo trọng số Glorot

Khởi tạo trọng số Glorot, hay còn gọi là khởi tạo trọng số Xavier Glorot, là phương pháp khởi tạo trọng số phổ biến trong các mô hình học sâu. Phương pháp này giúp đảm bảo rằng gradient của hàm mất mát được truyền ngược qua mạng nơ-ron một cách hiệu quả. Với $\mathcal{U}(-a, a)$ là phân phối đồng nhất trong khoảng $[-a, a]$, n_{in} và n_{out} lần lượt là số lượng nơ-

ron trong lớp đầu vào và đầu ra, khi đó trọng số w được khởi tạo qua công thức:

$$w = \mathcal{U}\left(-\frac{\sqrt{6}}{n_{in} + n_{out}}, \frac{\sqrt{6}}{n_{in} + n_{out}}\right)$$

6.1.3. Khởi tạo trọng số He

Khởi tạo trọng số He, hay còn gọi là khởi tạo trọng số Kaiming He, là phương pháp khởi tạo trọng số phổ biến trong các mô hình học sâu sử dụng ReLU hoặc leaky ReLU làm hàm kích hoạt. Phương pháp này giúp giảm thiểu vấn đề biến mất gradient. Với $\mathcal{N}(\mu, \sigma)$ là phân phối chuẩn có giá trị trung bình và độ lệch chuẩn lần lượt là μ và σ , n_{in} là số lượng nơ-ron trong lớp đầu vào, khi đó trọng số w được khởi tạo qua công thức:

$$w = \mathcal{N}\left(\mu, \frac{\sqrt{2}}{n_{in}}\right)$$

6.2. Tiến hành khởi tạo trọng số:

6.2.1. Quá trình khởi tạo trọng số:

- **Chuẩn bị dữ liệu:** Dữ liệu từ tập CIFAR-10 được tải xuống và chia thành tập huấn luyện (train_images, train_labels) Chuẩn bị dữ liệu: Dữ liệu từ tập CIFAR-10 được tải xuống và chia thành tập huấn luyện (train_images, train_labels) và tập kiểm tra (test_images, test_labels). Dữ liệu đầu vào được chuẩn hóa về giá trị từ 0 đến 1.
- **Xây dựng mô hình:** Mô hình được xây dựng dưới dạng một mạng nơ-ron hoàn toàn kết nối (Fully Connected Neural Network) với 3 lớp ẩn, mỗi lớp có 128 nơ-ron. Hàm kích hoạt Sigmoid được sử dụng ở mỗi lớp ẩn. Lớp cuối cùng sử dụng hàm kích hoạt softmax để chuyển đổi đầu ra thành các xác suất.
- **Biên dịch mô hình:** Mô hình sau đó được biên dịch với thuật toán tối ưu hóa 'adam', hàm mất mát 'sparse_categorical_crossentropy' (phù hợp cho bài toán phân loại nhiều lớp với nhãn dưới dạng số nguyên), và 'accuracy' là chỉ số đánh giá mô hình.
- **Huấn luyện mô hình:** Mô hình được huấn luyện trên tập huấn luyện với kích thước batch là 1024 và số epoch là 50. Trong mỗi epoch, mô hình sẽ học từ tất cả các mẫu trong tập huấn luyện. Tập kiểm tra được sử dụng như một tập validation để theo dõi độ chính xác của mô hình sau mỗi epoch.

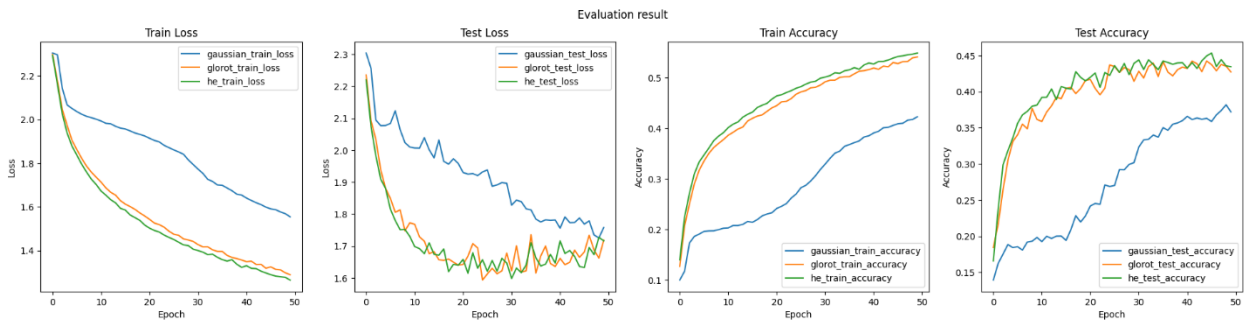
6.2.2. Kết quả khởi tạo trọng số:

Kĩ thuật khởi tạo	Độ chính xác trên tập huấn luyện	Độ chính xác trên tập kiểm thử
Khởi tạo trọng số Gaussian	0.4225	0.3719
Khởi tạo trọng số Glorot	0.5407	0.4272
Khởi tạo trọng số He	0.5484	0.4345

Bảng 3: Kết quả huấn luyện khi sử dụng các kĩ thuật khởi tạo trọng số

Dựa trên số liệu từ bảng 3, chúng ta có thể thấy rằng có sự so sánh giữa ba kỹ thuật khởi tạo trọng số khác nhau: Gaussian, Glorot và He. Các kỹ thuật này được đánh giá dựa trên độ chính xác của chúng trên tập huấn luyện và tập kiểm thử. Kỹ thuật khởi tạo trọng số He cho thấy hiệu suất cao nhất với độ chính xác là **0.5484** trên tập huấn luyện và **0.4345** trên tập kiểm thử. Glorot đứng thứ hai với độ chính xác **0.5407** trên tập huấn luyện và **0.4272** trên tập kiểm thử, trong khi Gaussian có hiệu suất thấp nhất với độ chính xác **0.4225** trên tập huấn luyện và **0.3719** trên tập kiểm thử.

Nhìn chung, kỹ thuật khởi tạo trọng số He có vẻ là lựa chọn tốt nhất trong ba phương pháp được so sánh, với hiệu suất ổn định trên cả hai tập dữ liệu. Điều này có thể giúp mô hình học máy có khả năng tổng quát hóa tốt hơn khi áp dụng vào dữ liệu thực tế. Tuy nhiên, cần lưu ý rằng việc lựa chọn kỹ thuật khởi tạo trọng số phù hợp cũng phụ thuộc vào kiến trúc mạng cụ thể và loại vấn đề mà mô hình đang giải quyết.



Hình 9: Đồ thị loss và accuracy khi sử dụng các kĩ thuật khởi tạo trọng số

7. Kết luận và đánh giá:

Trong việc xây dựng mô hình MLP (Multilayer Perceptron), việc lựa chọn hàm kích hoạt phù hợp là rất quan trọng. Hàm Tanh thường được ưu tiên hơn hàm Sigmoid do hai lý do chính. Đầu tiên, hàm Sigmoid có thể gây ra hiện tượng Gradient Vanishing, nơi giá trị đạo hàm trở nên rất nhỏ, làm chậm quá trình học của mạng nơ-ron. Thứ hai, hàm Sigmoid không tập trung dữ liệu quanh giá trị 0, điều này có thể gây ra sự bất cân xứng

trong việc truyền tín hiệu qua mạng và ảnh hưởng đến hiệu suất học. Trong khi đó, hàm Tanh có giá trị trung bình bằng 0 và đạo hàm lớn hơn ở các điểm xa giá trị trung bình, giúp tăng tốc độ hội tụ và cải thiện độ chính xác của mô hình.

Khi áp dụng các phương pháp chuẩn hóa dữ liệu, chuẩn hóa Mean và chuẩn hóa Z-score thường cho kết quả tốt hơn so với Min-Max Scaler, đặc biệt là trên tập dữ liệu có độ sáng và độ tương phản đa dạng như CIFAR10. Chuẩn hóa Z-score giữ nguyên thông tin thống kê của dữ liệu, giúp mạng nơ-ron học tập hiệu quả hơn.

Cuối cùng, trong việc khởi tạo trọng số cho mô hình học sâu, kỹ thuật He và Glorot thường được ưu tiên hơn Gaussian. Cả hai kỹ thuật này giúp giảm thiểu vấn đề Gradient Vanishing và Exploding Gradient, từ đó giúp mạng nơ-ron hội tụ nhanh hơn và đạt được độ chính xác cao hơn. Đặc biệt, kỹ thuật khởi tạo trọng số He rất phù hợp với các mạng nơ-ron sử dụng hàm kích hoạt ReLU hoặc Leaky ReLU, làm cho nó trở thành lựa chọn tốt nhất trong nhiều trường hợp.