Comparative Analysis of Optimizers and Neural Network Architectures:

SGD vs. Adam and MLP vs. CNN

Kim Min-Seo

June 2, 2024

Abstract

본 연구는 최적화 알고리즘과 신경망 아키텍처의 성능을 비교하는 것을 목적으로 합니다. 먼저, MNIST 데이터셋을 사용하여 다층 퍼셉트론(MLP) 모델에서 확률적 경사 하강법(SGD)과 Adam 최적화 알고리즘의 성능을 비교하였습니다. 비교 결과, Adam 알고리즘이 SGD보다 빠르고 안정적으로 수렴하며 더 높은 성능을 보였습니다. 이후 Adam 알고리즘을 사용하여 MLP와 합성곱 신경망(CNN)의 성능을 비교한 결과, CNN이 검증 데이터셋에서 99% 이상의 정확도와 낮은 손실을 기록하며 MLP보다 우수한 성능을 나타냈습니다. 이러한 결과는 최적화 알고리즘과 신경망 아키텍처 선택이 모델 성능에 중요한 영향을 미친다는 것을 시사하며, 다양한 문제에 적합한 알고리즘과 아키텍처를 선택하는 것이 중요함을 강조합니다.

**Contents**

1. Introduction

1.1 Objectives of the Study

2. Literature Review

2.1 Overview of Stochastic Gradient Descent (SGD) and Adam Optimizer

2.2 Overview of Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN)

3. Methodology

3.1 Description of the MNIST Dataset

4. Experience

4.1. SGD vs. Adam Optimizer on MLP

4.1.1 Model Architecture of MLP

4.1.2 Training Procedure

4.1.3 Performance Metrics (Loss and Accuracy)

4.2. MLP vs. CNN using Adam Optimizer

4.2.1. Model Architectures of MLP and CNN

4.2.2. Training Procedure

4.2.3. Performance Metrics (Loss and Accuracy)

5. Result and Discussion

5.1 Comparative Analysis of SGD and Adam on MLP

5.2 Comparative Analysis of MLP and CNN using Adam

6. Conclusion

6.1 Summary of Findings

1. Introduction

1.1 Objectives of the Study

이 연구는 최적화 알고리즘과 CNN, MLP간의 비교를 목적으로 하고 있습니다.

우선 최적화 알고리즘을 비교합니다. 확률적 경사 하강법(SGD)과 Adam 최적화 알고리즘의 성능을 비교 분석합니다. 이 비교를 통해 두 알고리즘이 다층 퍼셉트론(MLP) 모델의 학습 과정에서 나타내는 차이점과 각 알고리즘의 장단점을 명확히 이해하고자 합니다.

다음은 신경망 아키텍처간의 비교입니다. 앞서 비교한 알고리즘 중 성능이 비교적 좋은 알고리즘을 사용하여 MLP와 합성곱 신경망(CNN)의 성능을 비교합니다. 이 비교를 통해 두 가지 신경망 아키텍처가 보여주는 성능 차이를 분석하고, 각 모델의 특성과 적용 가능성을 평가하고자 합니다.

2. Literature Review

2.1 Overview of Stochastic Gradient Descent (SGD) and Adam Optimizer

확률적 경사 하강법은 기계 학습 및 딥러닝 모델의 최적화에 널리 사용되는 알고리즘입니다. SGD는 전체 데이터셋을 사용하지 않고, 매 반복(iteration)마다 임의의 소규모 데이터셋(배치)를 선택하여 모델의 가중치를 업데이트합니다. 이는 계산 비용을 줄이고, 큰 데이터셋에서도 효율적으로 동작할 수 있게 합니다. 그러나 SGD는 학습 과정에서 진동과 불안정성을 겪을 수 있으며, 학습 속도가 느려질 수 있습니다.

Adam(Adaptive Moment Estimation)은 모멘텀과 RMSProp의 장점을 결합한 최적화 알고리즘입니다. Adam은 학습률을 각 가중치에 대해 개별적으로 조정하고, 과거의 기울기 정보를 활용하여 빠르고 안정적으로 학습을 진행합니다. Adam은 학습 과정에서의 진동을 줄이고, SGD보다 빠르게 수렴할 수 있습니다. 이러한 특징 덕분에 Adam은 다양한 딥러닝 모델에서 널리 사용되고 있습니다.

2.2 Overview of Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN)

다층 퍼셉트론 레이어(MLP)는 가장 기본적인 형태의 신경망 중 하나로, 입력층, 여러 개의 은닉층(hidden layer), 출력층으로 구성됩니다. 각 층의 뉴런은 이전 층의 모든 뉴런과 연결되어 있으며, 활성화 함수(activation function)를 통해 비선형성을 추가합니다. MLP는 이미지, 텍스트, 소리 등 다양한 데이터에 적용될 수 있지만, 이미지와 같은 고차원 데이터를 처리하는 데 한계가 있습니다.

합성곱 신경망(CNN)은 주로 이미지 인식 및 처리에 사용되는 신경망 아키텍처입니다. CNN은 합성곱 층(convolutional layer), 풀링 층(pooling layer), 완전 연결층(fully connected layer)으로 구성됩니다. 합성곱 층은 이미지의 국소적 패턴을 추출하고, 풀링 층은 공간적 차원을 축소하여 특징 맵(feature map)을 만듭니다. 이러한 구조 덕분에 CNN은 이미지의 공간적 계층 구조를 잘 학습할 수 있으며, 이미지 분류, 객체 검출 등 다양한 컴퓨터 비전 과제에서 뛰어난 성능을 발휘합니다.

3. Methodology

3.1 Description of the MNIST Dataset

MNIST 데이터셋은 손으로 쓴 숫자 0부터 9까지의 60,000개의 이미지로 구성된 이미지 분류 데이터셋입니다. 각 이미지는 28x28 픽셀 크기이며 0부터 255까지의 밝기 값을 가진 흑백 이미지입니다. 이 MNIST 데이터셋을 활용하여 아래 실험의 성능을 비교합니다.

4. Experience

4.1. SGD vs. Adam Optimizer on MLP

4.1.1 Model Architecture of MLP

본 실험에서는 MNIST 이미지 분류를 위해 두개의 은닉층을 갖는 MLP 모델을 사용했습니다. 입력층은 28x28 이미지를 1x784 벡터로 flatten합니다. 첫번째 은닉층은 784개의 입력 뉴런과 256개의 출력 뉴런을 가진 완전 연결층입니다. 이후 ReLU 활성화 함수를 사용하고, 30%의 dropout을 적용합니다. 두번째 은닉층은 256개의 입력 뉴런과 10개의 출력 뉴런을 가진 완전 연결층입니다. 10개의 출력 뉴런은 MNIST 데이터셋에 있는 10개의 클래스를 나타냅니다. 마지막으로 10개의 출력 뉴런으로 구성된 소프트맥스 함수를 사용하여 각 클래스에 대한 확률을 계산합니다.

4.1.2 Training Procedure

전체 train 데이터셋을 랜덤셔플한 후 0.85:0.15 비율로 train set과 validation set을 나눕니다. 이후 dataload에 입력하여 batch size인 32크기만큼의 mini-batch 형태의 데이터셋을 준비합니다. 학습에서는 CrossEntropyLoss를 손실 함수로 사용하며, 각 에포크(epoch)마다 모델을 학습하고 검증 세트를 사용하여 모델의 성능을 평가합니다.

4.1.3 Performance Metrics (Loss and Accuracy)

훈련 과정 동안 각 에포크마다 손실과 정확도를 기록하고, 학습 추이 및 절대적인 수치 등 다각도로 optimizer를 평가합니다. 이를 통해 SGD와 Adam 최적화 알고리즘을 각각 적용한 MLP 모델의 성능을 비교하고, 각 최적화 알고리즘이 학습 과정과 결과에 미치는 영향을 분석합니다.

4.2. MLP vs. CNN using Adam Optimizer

4.2.1. Model Architectures of MLP and CNN

MLP 모델은 이전 섹션에서 설명한 MLP 모델과 같은 형태의 모델이 사용됩니다.

CNN 모델은 두개의 합성곱층을 거칩니다. 첫번째 합성곱층은 입력 채널 1, 출력 채널 32, 커널 크기 3, 패딩 1입니다. 두번째 합성곱층은 입력 채널 32, 출력 채널 64, 커널 크기 3, 패딩 1입니다. 각각의 합성곱층 이후에는 ReLu 활성화함수, 커널 크기2 및 패딩크기 2의 맥스풀링레이어, 25%의 dropout이 적용됩니다. 이후 7x7x64 크기의 입력을 256개의 노드로 변환하는 완전 연결층 1을 지난 후 ReLu 활성화함수, 50%의 dropout이 적용됩니다. 마지막으로 256개의 입력을 10개의 출력 노드로 변환하는 완전 연결층 2를 거쳐 10개의 출력이 출력됩니다.

4.2.2. Training Procedure

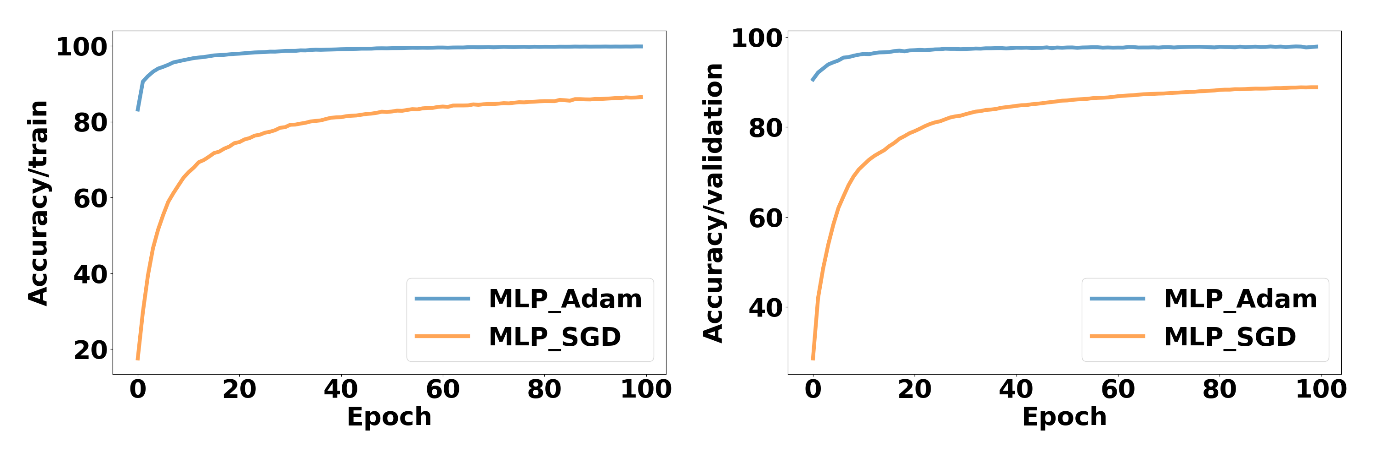
이전 실험과 동일한 데이터 로드 및 전처리, 손실함수, 훈련과정이 적용됩니다.

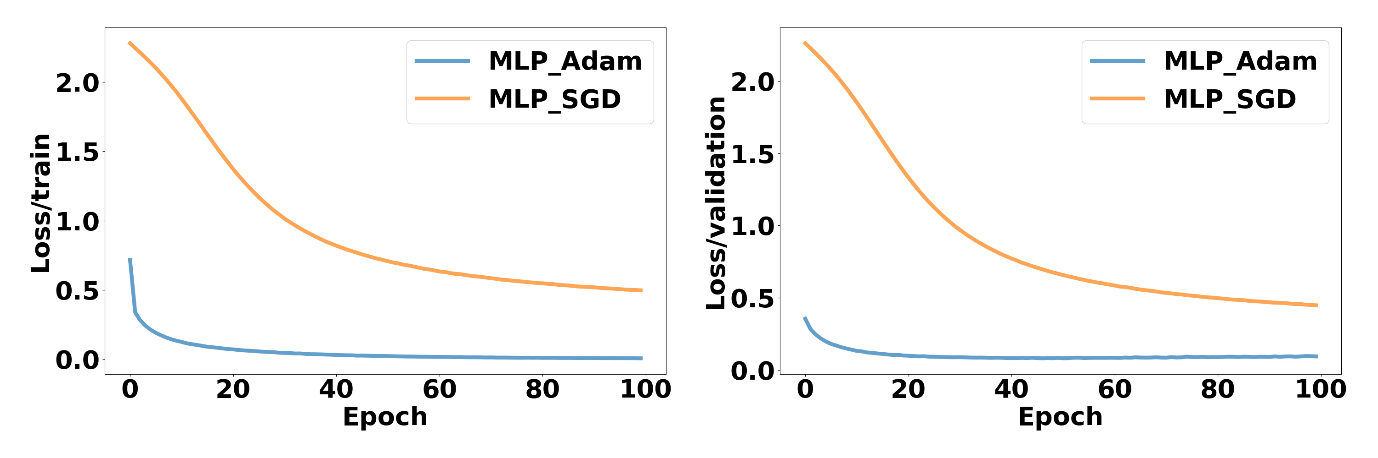
4.2.3. Performance Metrics (Loss and Accuracy)

이전 실험과 동일하게 훈련 과정 동안 각 에포크마다 손실과 정확도를 기록하고, 학습 추이 및 절대적인 수치 등 다각도로 mlp과 cnn을 비교분성합니다.

5. Result and Discussion

5.1 Comparative Analysis of SGD and Adam on MLP





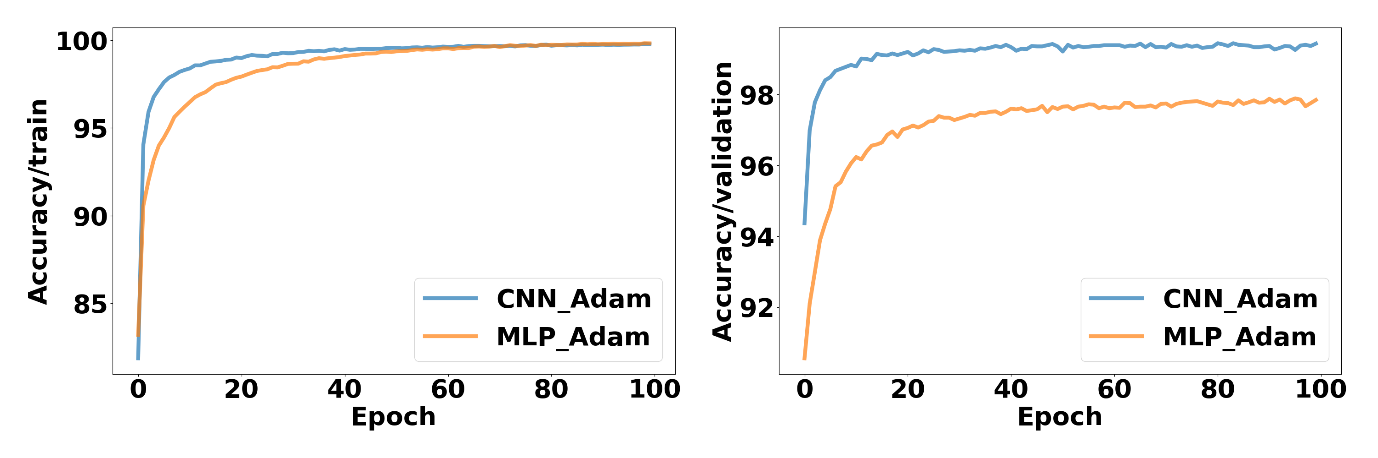
Adam 최적화 알고리즘을 사용한 MLP 모델은 초기 학습 단계에서부터 정확도가 빠르게 상승하는 경향을 보였습니다. 반면에, SGD를 사용한 모델은 동일한 에포크 수에서 상대적으로 느리게 정확도가 상승하는 패턴을 나타냈습니다. 이는 Adam이 학습률을 각 파라미터에 대해 개별적으로 조정하기 때문에 학습 과정에서 더 빠르고 안정적으로 최적화가 이루어졌기 때문으로 해석할 수 있습니다.

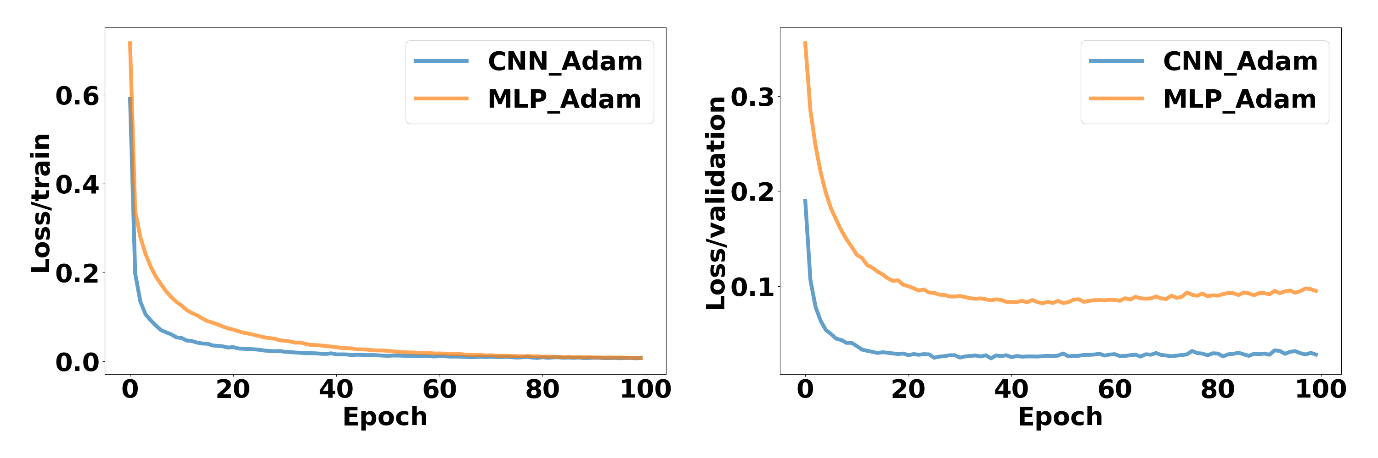
손실 지표에서도 Adam 최적화 알고리즘을 사용한 모델이 더 빠르게 감소하는 양상을 보였습니다. SGD를 사용한 모델은 손실이 점진적으로 감소하는 반면, Adam을 사용한 모델은 초기 몇 에포크 동안 손실이 급격히 감소하였고, 이후에도 지속적으로 낮은 손실 값을 유지하였습니다. 이는 Adam이 모멘텀과 RMSProp의 장점을 결합하여, 학습 과정에서의 진동을 줄이고 보다 안정적으로 수렴할 수 있기 때문입니다.

SGD와 Adam의 학습 속도 차이 역시 매우 두드러졌습니다. SGD는 학습 속도가 느리고, 손실 감소 및 정확도 증가가 점진적으로 이루어졌습니다. 이는 SGD가 단일 학습률을 사용하며, 학습 과정에서 진동이 발생할 수 있기 때문입니다. 반면에, Adam은 초기 학습 단계에서 빠르게 손실이 감소하고 정확도가 상승하였으며, 지속적으로 높은 성능을 유지하였습니다. Adam은 학습률을 동적으로 조정하고, 과거의 기울기 정보를 활용하여 빠르게 수렴할 수 있었습니다.

결론적으로, MNIST 데이터셋을 사용한 MLP 모델의 학습에서 Adam 최적화 알고리즘이 SGD에 비해 더 효율적이고 빠르게 성능을 향상시킬 수 있음을 확인하였습니다. 이는 Adam이 보다 복잡한 신경망 구조와 큰 데이터셋을 처리하는 데 있어서도 우수한 성능을 발휘할 가능성을 시사합니다. 학습 과정에서 손실과 정확도의 변화를 통해 Adam이 SGD보다 더 효과적인 최적화 알고리즘임을 명확히 알 수 있었습니다.

5.2 Comparative Analysis of MLP and CNN using Adam





본 섹션에서는 Adam 최적화 알고리즘을 사용하여 MLP와 CNN 모델의 성능을 비교 분석합니다. 성능 평가 지표로는 정확도(Accuracy)와 손실(Loss)을 사용하였으며, 총 100 에포크 동안의 성능 변화를 관찰하였습니다.

훈련 데이터셋에서 CNN 모델은 초기 학습 단계에서부터 정확도가 빠르게 상승하는 경향을 보였습니다. 반면, MLP 모델은 CNN에 비해 상대적으로 느리게 정확도가 상승하였으나, 결국 두 모델 모두 100 에포크 내에 거의 99% 이상의 높은 정확도를 기록하였습니다. 손실 지표에서도 CNN 모델이 초기에는 더 빠르게 감소하였으나, MLP 모델 역시 꾸준히 손실이 줄어들어 최종적으로 두 모델 모두 로스값이 약 0.005 수준까지 손실이 감소하였습니다.

검증 데이터셋에서는 두 모델 간의 성능 차이가 더 명확하게 나타났습니다. MLP 모델의 경우 검증 정확도가 97%를 상회하는 반면, CNN 모델은 99% 이상의 성능을 보여주었습니다. 손실 지표에서도 CNN 모델은 검증 손실이 0.002까지 감소한 반면, MLP 모델은 0.008 수준까지 밖에 감소하지 않았습니다. 이는 CNN이 MLP보다 일반화 성능에서 더 우수함을 시사합니다.

종합적으로, CNN모델이 MLP모델에 비해 훈련 및 검증 데이터셋 모두에서 높은 정확도와 낮은 손실을 기록하였으며, 특히 검증 데이터셋에서의 성능이 우수하였습니다. 이는 CNN이 합성곱 층을 통해 이미지의 공간적 특징을 효과적으로 학습할 수 있기 때문입니다.

6. Conclusion

6.1 Summary of Findings

SGD와 Adam 최적화 알고리즘을 비교한 결과, Adam 알고리즘이 MLP 모델의 학습에서 더 빠르고 안정적인 성능을 보여주었습니다. Adam을 사용한 MLP 모델은 초기 학습 단계에서부터 정확도가 빠르게 상승하고 손실이 급격히 감소하는 경향을 보였으며, 전체 학습 과정에서도 꾸준히 높은 성능을 유지하였습니다. 반면, SGD를 사용한 MLP 모델은 정확도와 손실이 점진적으로 개선되는 모습을 보였으나, Adam에 비해 학습 속도가 느렸습니다.

다음으로, Adam 최적화 알고리즘을 사용하여 MLP와 CNN 모델의 성능을 비교한 결과, CNN 모델이 MLP 모델보다 더 우수한 성능을 나타냈습니다. 훈련 데이터셋에서 두 모델 모두 높은 정확도와 낮은 손실을 기록하였으나, 검증 데이터셋에서는 CNN 모델이 더 높은 정확도와 낮은 손실을 보여주었습니다. 이는 CNN이 이미지의 공간적 특징을 효과적으로 학습할 수 있는 구조를 가지고 있기 때문입니다. 특히, CNN 모델은 검증 데이터셋에서 99% 이상의 정확도와 0.002 이하의 손실을 기록하며, MLP 모델에 비해 일반화 성능에서 더 우수함을 확인할 수 있었습니다.

결론적으로, 본 연구를 통해 Adam 최적화 알고리즘이 MLP 모델의 학습에서 더 효과적이며, CNN 모델이 이미지 분류 문제에서 MLP 모델보다 더 우수한 성능을 발휘함을 확인하였습니다. 이는 최적화 알고리즘과 신경망 아키텍처의 선택이 모델 성능에 중요한 영향을 미친다는 점을 시사하며, 다양한 데이터셋과 문제 유형에 적합한 알고리즘과 아키텍처를 선택하는 것이 중요함을 강조합니다.