Intro to Deep Learning Assignment 3

Tsinghua University
Allan Li (李盼加乐), 2025403473

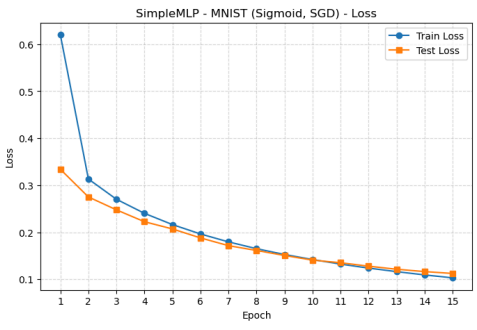
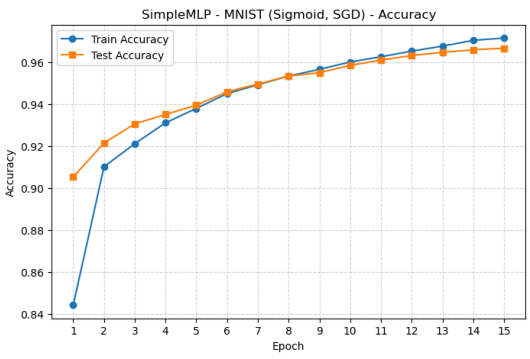
Part 1: Objective Questions

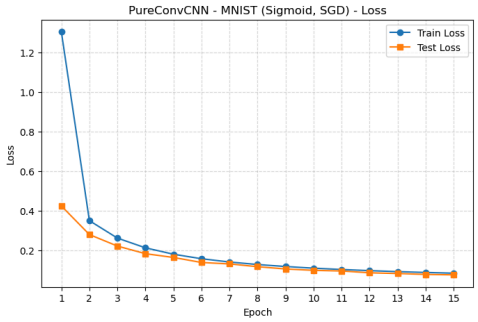
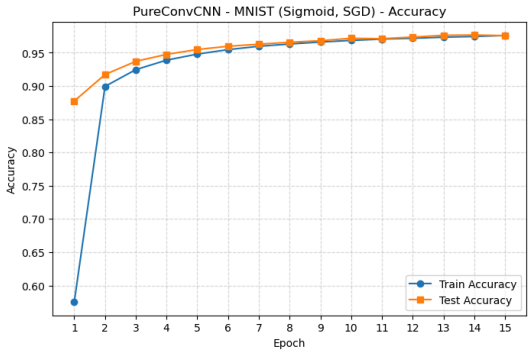
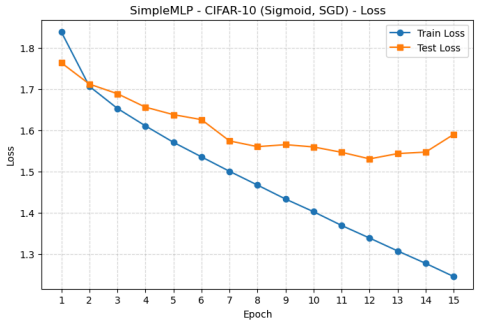
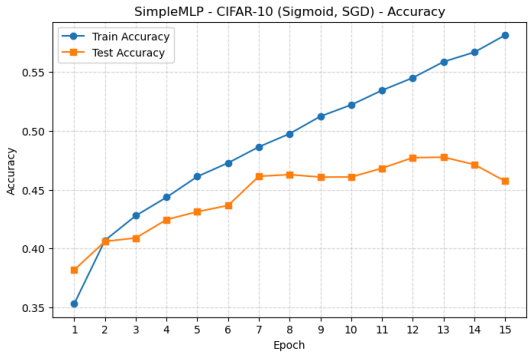
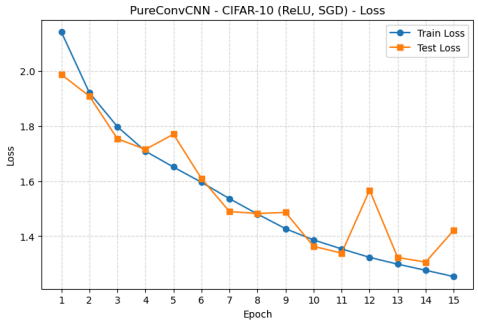
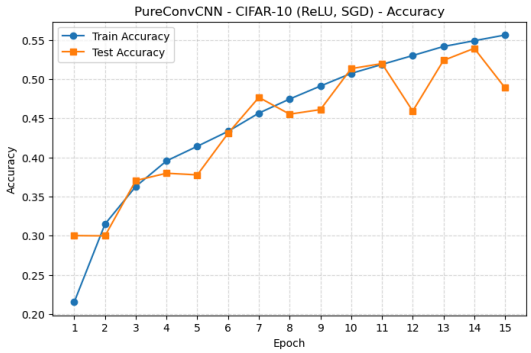
1. C, D, E
2. A
3. A, B, C
4. A
5. B
6. B

Part 2: MLP & CNNs for MNIST and CIFAR-10

Summary of Base Results

Experiment Description	Train Loss	Train Accuracy	Test Loss	Test Accuracy
MLP on MNIST	0.1030	97.15%	0.1123	96.67%
CNN on MNIST	0.8521	97.59%	0.0774	97.57%
MLP on CIFAR-10	1.2459	58.12%	1.5904	45.73%
CNN on CIFAR-10	1.2533	55.62%	1.4214	48.88%

Experiment Description	Loss Plot	Accuracy Plot
MLP on MNIST		

Experiment Description	Loss Plot	Accuracy Plot
CNN on MNIST		
MLP on CIFAR-10		
CNN on CIFAR-10		

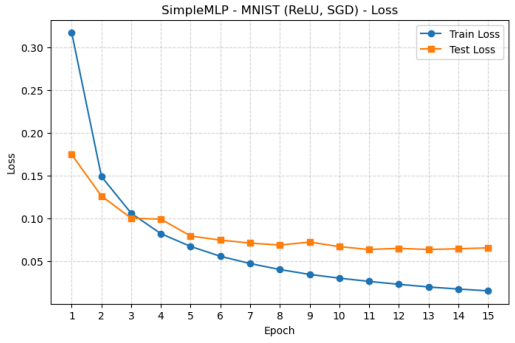
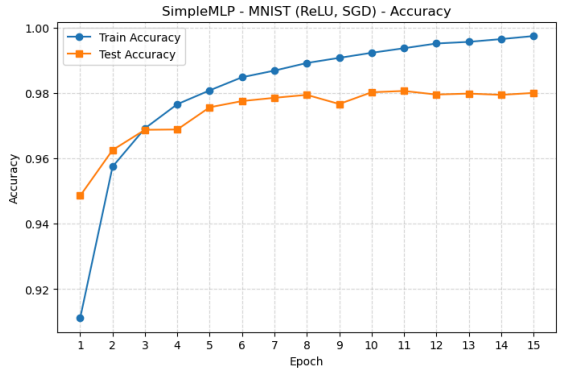
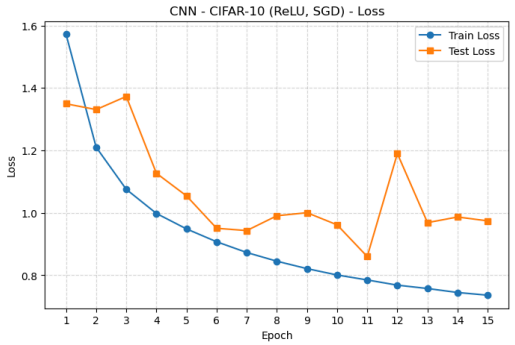
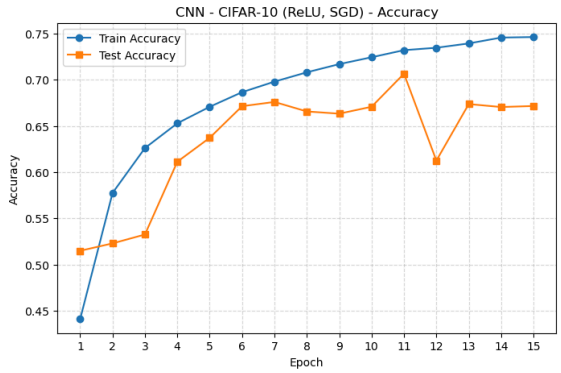
Generally speaking, the MLP and CNN were able to capture the complexity in the MNIST dataset, and performed similarly on the test data. However, on the CIFAR dataset, both models performed poorly, with accuracy rates hovering between 40-50%. This suggests significant underfitting and potentially an suboptimal set of model architecture for both the MLP and the CNN.

For the remainder of this report, the experiments were conducted using two scenarios: MLP & MNIST, and CNN & CIFAR-10, since these were generally the best performing models in each scenario (and for diversification of results during analysis).

Summary of Experiments

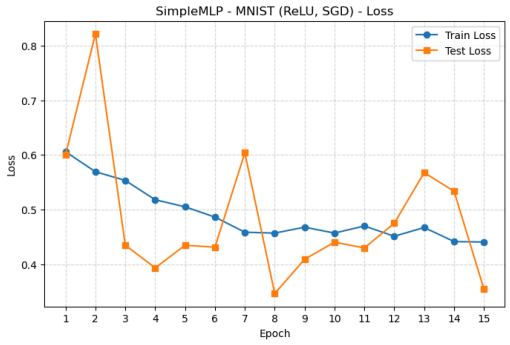
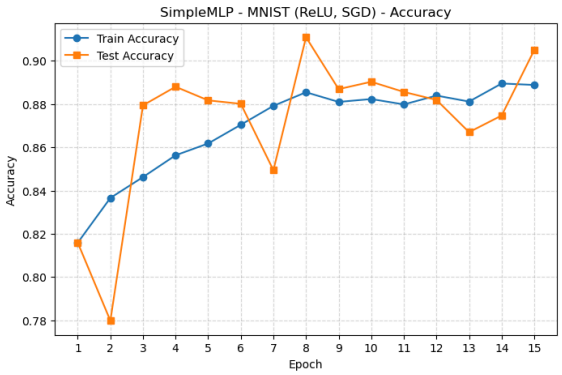
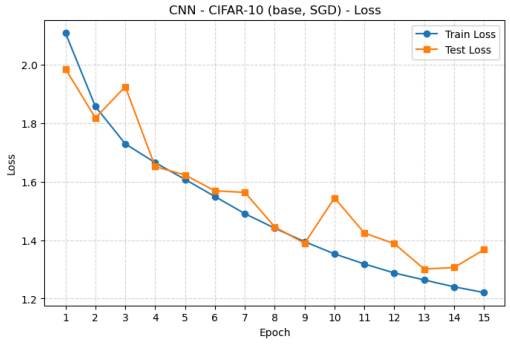
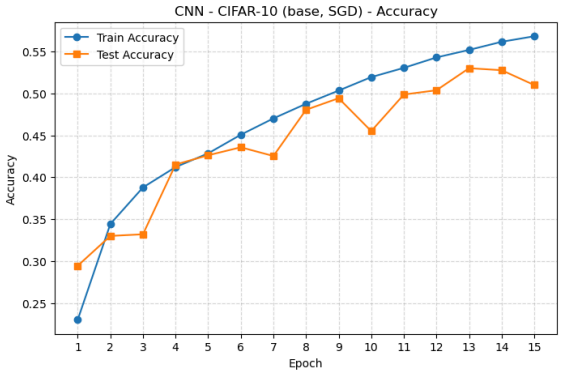
Experiment 1: Changing Activation Function

Experiment Description	Train Loss	Train Accuracy	Test Loss	Test Accuracy
MLP on MNIST (sigmoid)	0.1030	97.15%	0.1123	96.67%
CNN on CIFAR-10 (sigmoid)	1.2533	55.62%	1.4214	48.88%
MLP on MNIST (ReLU)	**0.0153**	**99.75%**	**0.0656**	**98.01%**
CNN on CIFAR-10 (ReLU)	**0.7360**	**74.63%**	**0.9741**	**67.16%**

Experiment	Loss Plot	Accuracy Plot
ReLU MLP		
ReLU CNN		

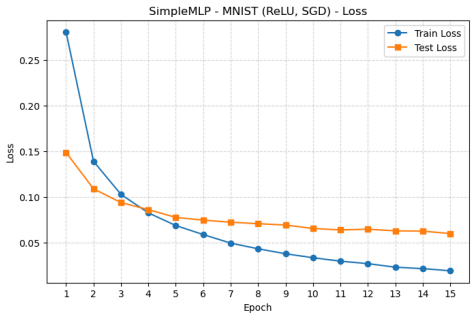
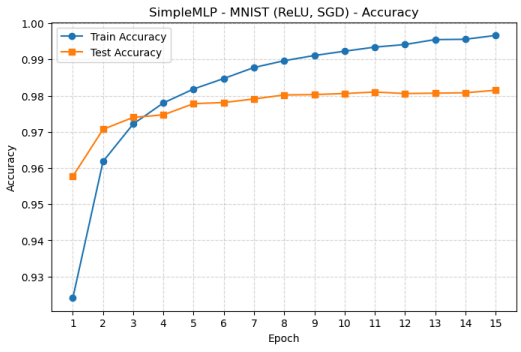
Experiment 2: Changing Optimizer

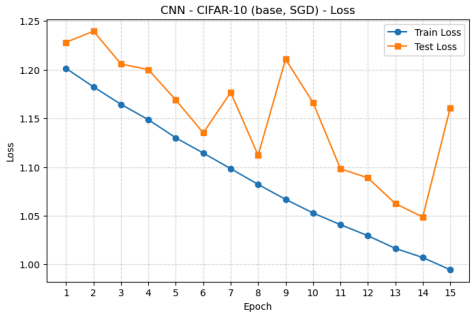
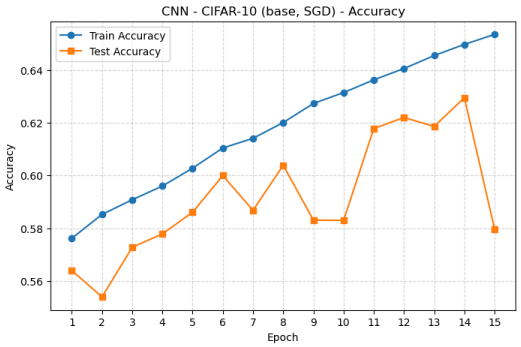
Experiment Description	Train Loss	Train Accuracy	Test Loss	Test Accuracy
MLP on MNIST (SGD)	0.1030	97.15%	0.1123	96.67%
CNN on CIFAR-10 (SGD)	1.2533	55.62%	1.4214	48.88%
MLP on MNIST (SGD)	**0.4406**	**88.88%**	**0.3543**	**90.50%**
CNN on CIFAR-10 (SGD)	**1.2209**	**56.81%**	**1.3674**	**50.99%**

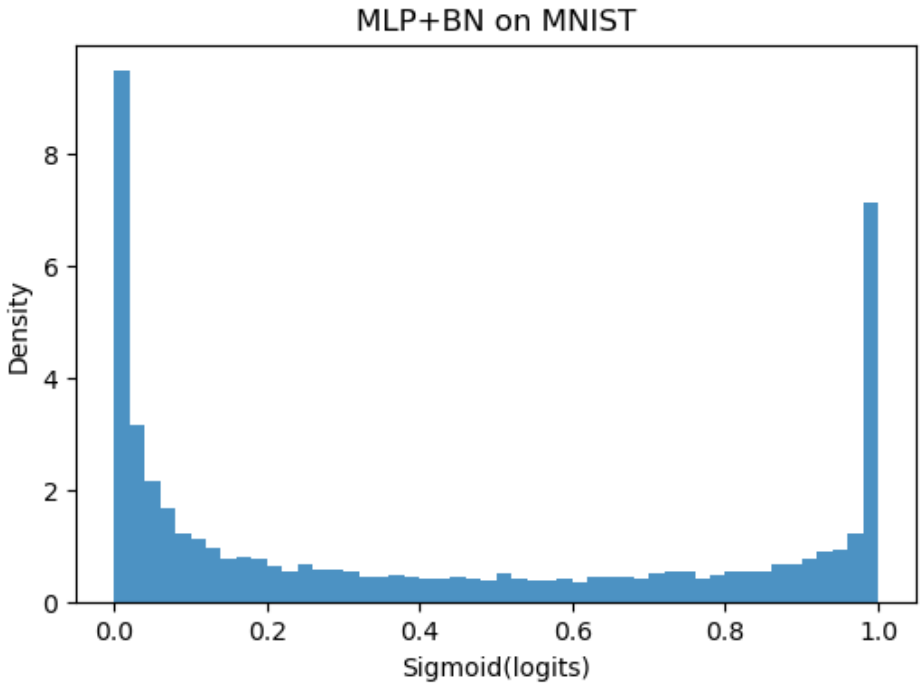
Experiment	Loss Plot	Accuracy Plot
Adam MLP		
Adam CNN		

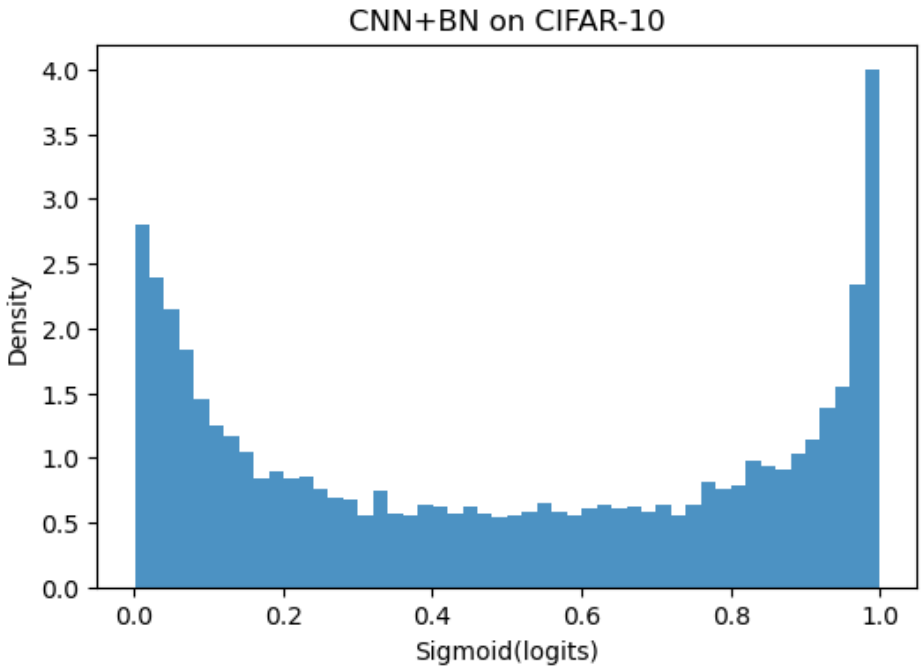
Experiment 3: Adding and Observing Batch Normalization

Experiment Description	Train Loss	Train Accuracy	Test Loss	Test Accuracy
MLP on MNIST (SGD)	0.1030	97.15%	0.1123	96.67%
CNN on CIFAR-10 (SGD)	1.2533	55.62%	1.4214	48.88%
MLP on MNIST (ReLU)	0.0192	99.67%	0.0599	98.15%
CNN on CIFAR-10 (ReLU)	0.9945	65.35%	1.1605	57.97%

Experiment	Loss Plot	Accuracy Plot
Batch Normalization MLP		

Experiment	Loss Plot	Accuracy Plot
Batch Normalization CNN		

Experiment	Distribution Values of Second Sigmoid Function
Batch Normalization MLP	

Experiment	Distribution Values of Second Sigmoid Function
Batch Normalization CNN	

Experiment 4: CNN vs MLP Analysis

Parameters

On average, CNNs were able to express the data using much fewer parameters than the MLP. Comparing the parameter count of the base models for CNNs and MLPs analyzing the CIFER-10 dataset, the parameter count for the CNN was 20,778, compared to 789,258 for the MLP.

Training Speed

Comparing the two, CNN based models took significantly longer time to train. Comparing the base models, the CNN model took 683.76 seconds, compared to 569.03 seconds from the MLP model on the CIFER-10 dataset.

Convergence Speed

On both datasets, it seems that the MLP was able to converge much faster than the CNN. Despite the fact that both models seemed to underfit the CIFER-10 dataset, it seems as if the CNN was converging slightly slower than the MLP model, and would definitely benefit further from more training time and a larger amount of epochs than the MLP.

Raw statistics of performance metrics for each model in each scenerio can be found in the raw Jupyter Notebook code block outputs.