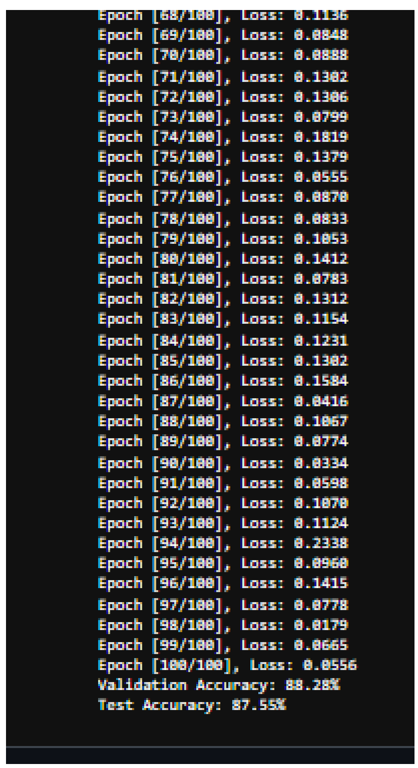
Problem 1a

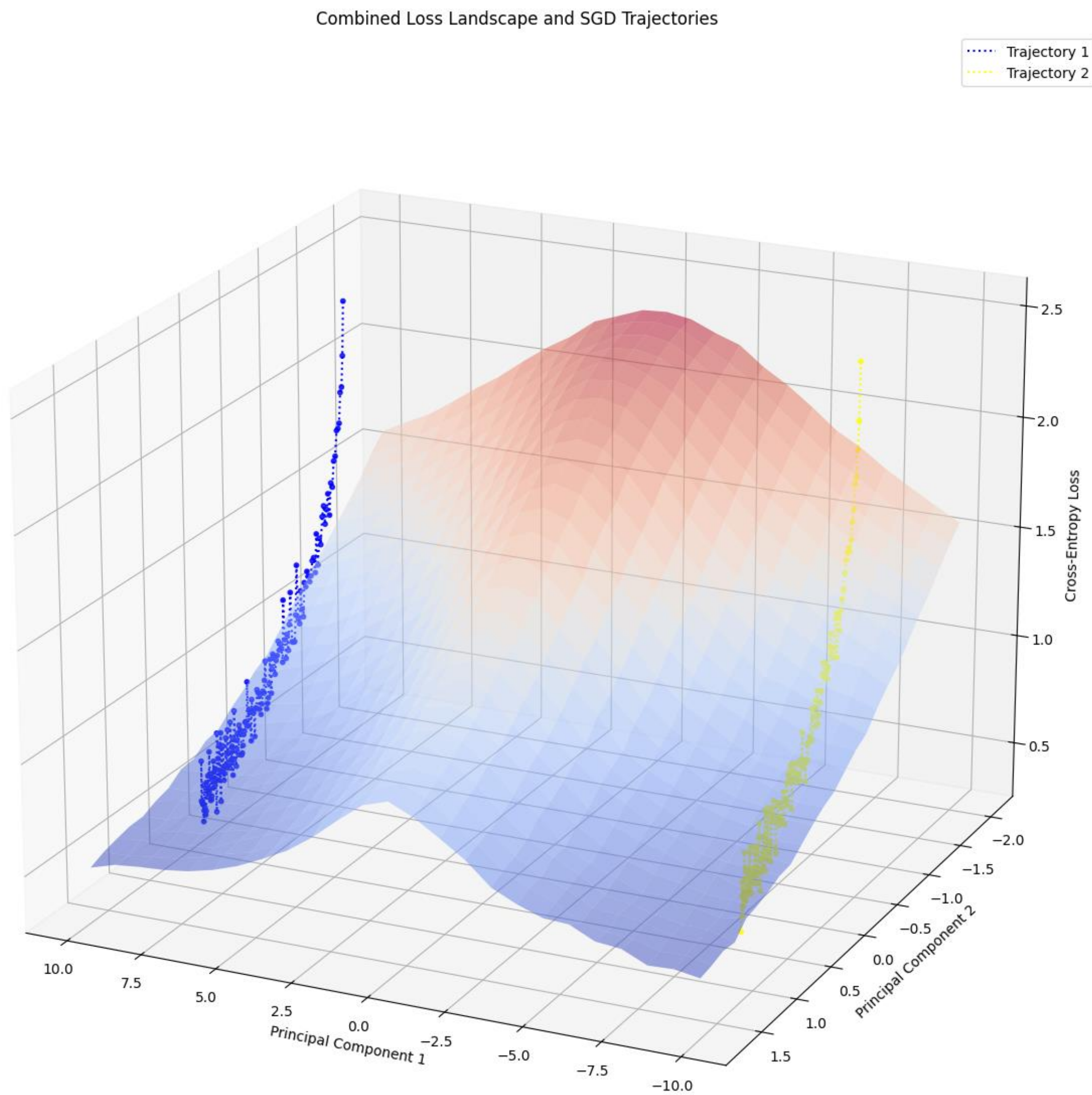
```
Shape of X [N, C, H, W]: torch.Size([64, 1, 28, 28])
Shape of y: torch.Size([64]) torch.int64
Using cpu device

Epoch [86/100], Loss: 0.1584
Epoch [87/100], Loss: 0.0416
Epoch [88/100], Loss: 0.1067
Epoch [89/100], Loss: 0.0774
Epoch [90/100], Loss: 0.0334
Epoch [91/100], Loss: 0.0598
Epoch [92/100], Loss: 0.1070
Epoch [93/100], Loss: 0.1124
Epoch [94/100], Loss: 0.2338
Epoch [95/100], Loss: 0.0960
Epoch [96/100], Loss: 0.1415
Epoch [97/100], Loss: 0.0778
Epoch [98/100], Loss: 0.0179
Epoch [99/100], Loss: 0.0665
Epoch [100/100], Loss: 0.0556
Validation Accuracy: 88.28%
Test Accuracy: 87.55%
```



```
Epoch [68/100], Loss: 0.1136
Epoch [69/100], Loss: 0.0848
Epoch [70/100], Loss: 0.0888
Epoch [71/100], Loss: 0.1302
Epoch [72/100], Loss: 0.1306
Epoch [73/100], Loss: 0.0799
Epoch [74/100], Loss: 0.1819
Epoch [75/100], Loss: 0.1379
Epoch [76/100], Loss: 0.0555
Epoch [77/100], Loss: 0.0870
Epoch [78/100], Loss: 0.0833
Epoch [79/100], Loss: 0.1053
Epoch [80/100], Loss: 0.1412
Epoch [81/100], Loss: 0.0783
Epoch [82/100], Loss: 0.1312
Epoch [83/100], Loss: 0.1154
Epoch [84/100], Loss: 0.1231
Epoch [85/100], Loss: 0.1302
Epoch [86/100], Loss: 0.1584
Epoch [87/100], Loss: 0.0416
Epoch [88/100], Loss: 0.1067
Epoch [89/100], Loss: 0.0774
Epoch [90/100], Loss: 0.0334
Epoch [91/100], Loss: 0.0598
Epoch [92/100], Loss: 0.1070
Epoch [93/100], Loss: 0.1124
Epoch [94/100], Loss: 0.2338
Epoch [95/100], Loss: 0.0960
Epoch [96/100], Loss: 0.1415
Epoch [97/100], Loss: 0.0778
Epoch [98/100], Loss: 0.0179
Epoch [99/100], Loss: 0.0665
Epoch [100/100], Loss: 0.0556
Validation Accuracy: 88.28%
Test Accuracy: 87.55%
```

Problem 1b



Combined Loss Landscape and SGD Trajectories
Trajectory 1
Trajectory 2
Cross-Entropy Loss
Principal Component 1
Principal Component 2

Problem 1c

The observed discrepancy between the cross-entropy loss (fCE) values on the SGD trajectory points and those on the surface plot is due to the dimensionality reduction introduced by PCA.

- PCA compresses the high-dimensional parameter space into two principal components, capturing the largest sources of variance. However, this reduction discards other dimensions that may contain relevant information, resulting in a loss landscape that is an approximation rather than a precise representation of the original space.

Achieving an exact match between the "real" fCE values from the SGD trajectory and the surface plot would be computationally impractical:

- Constructing a precise surface plot would require evaluating the loss function over a dense grid across all dimensions of the parameter space, which grows exponentially with the number of parameters.
- Such exhaustive computations would demand immense computational resources and time, making it infeasible to visualize exact values in the high-dimensional space.

In summary, PCA allows us to approximate the general trends in the loss landscape within a 2D space, but at the cost of reduced fidelity for individual loss values. This trade-off is necessary to make the visualization computationally feasible while capturing the broad contours of the loss landscape.

Problem 1d.

Statement 1: If $p = 2p'$, then $\tilde{p} = 2\tilde{p}'$.

If we double the size of the neural network's parameters, the corresponding projections in the PCA space will also double. This is true because PCA is a linear transformation, and scaling the input data will scale the projections in a similar way.

Statement 2: If $\tilde{p} = 2\tilde{p}'$, then $p = 2p'$

If the projections in the PCA space are doubled, it doesn't necessarily mean the original parameters were doubled. This is because PCA can compress information, and the original data might have had other components that were not captured in the doubled projections.

Statement 3: If $\tilde{p} = 2\tilde{p}'$, then $\hat{p} = 2\hat{p}$

If the projections in the PCA space are doubled, the reconstructed data will also be doubled. This is because PCA is a reversible transformation, and scaling the projections will result in a scaled reconstruction.

Statement 4: $fCE(\hat{p}) \leq fCE(p)$

Doubling the neural network parameters doesn't necessarily reduce the cross-entropy loss. PCA can lead to information loss during compression, which might increase the loss.

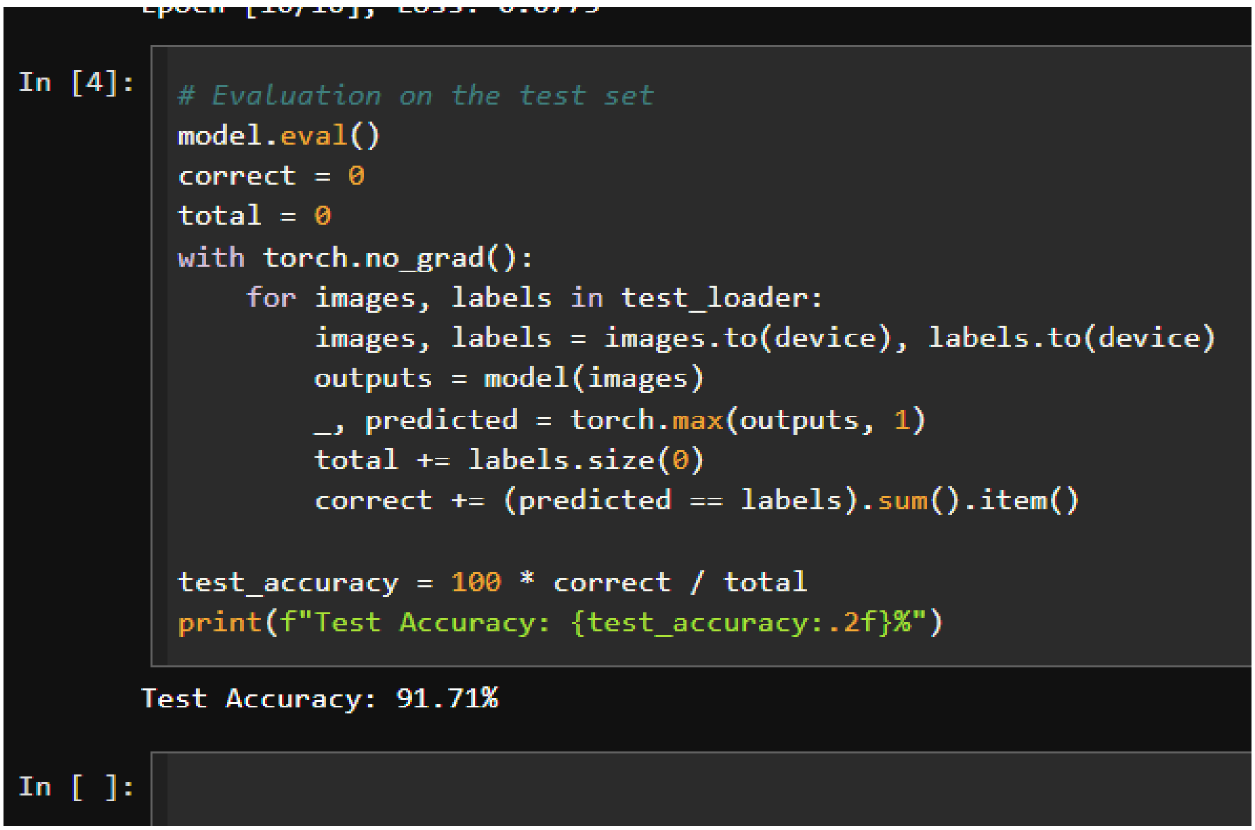
Therefore, statements 1 and 3 are true.

Problem 2

Training :

```
Epoch [1/10], Loss: 0.4331
Epoch [2/10], Loss: 0.2769
Epoch [3/10], Loss: 0.2288
Epoch [4/10], Loss: 0.1984
Epoch [5/10], Loss: 0.1720
Epoch [6/10], Loss: 0.1494
Epoch [7/10], Loss: 0.1296
Epoch [8/10], Loss: 0.1085
Epoch [9/10], Loss: 0.0913
Epoch [10/10], Loss: 0.0773
```

`Accuracy: Test Accuracy: 91.71%`



```
In [4]:
# Evaluation on the test set
model.eval()
correct = 0
total = 0
with torch.no_grad():
    for images, labels in test_loader:
        images, labels = images.to(device), labels.to(device)
        outputs = model(images)
        _, predicted = torch.max(outputs, 1)
        total += labels.size(0)
        correct += (predicted == labels).sum().item()

test_accuracy = 100 * correct / total
print(f"Test Accuracy: {test_accuracy:.2f}%")

Test Accuracy: 91.71%

In [ ]:
```

Problem 3

[8]

```
# Train Only Final Layer
print("Training Only Final Layer")
train_model(model, train_loader, criterion, optimizer, epochs=5)
final_layer_accuracy = evaluate_model(model, test_loader)
print(f"Final Layer Only Accuracy: {final_layer_accuracy}%")
```

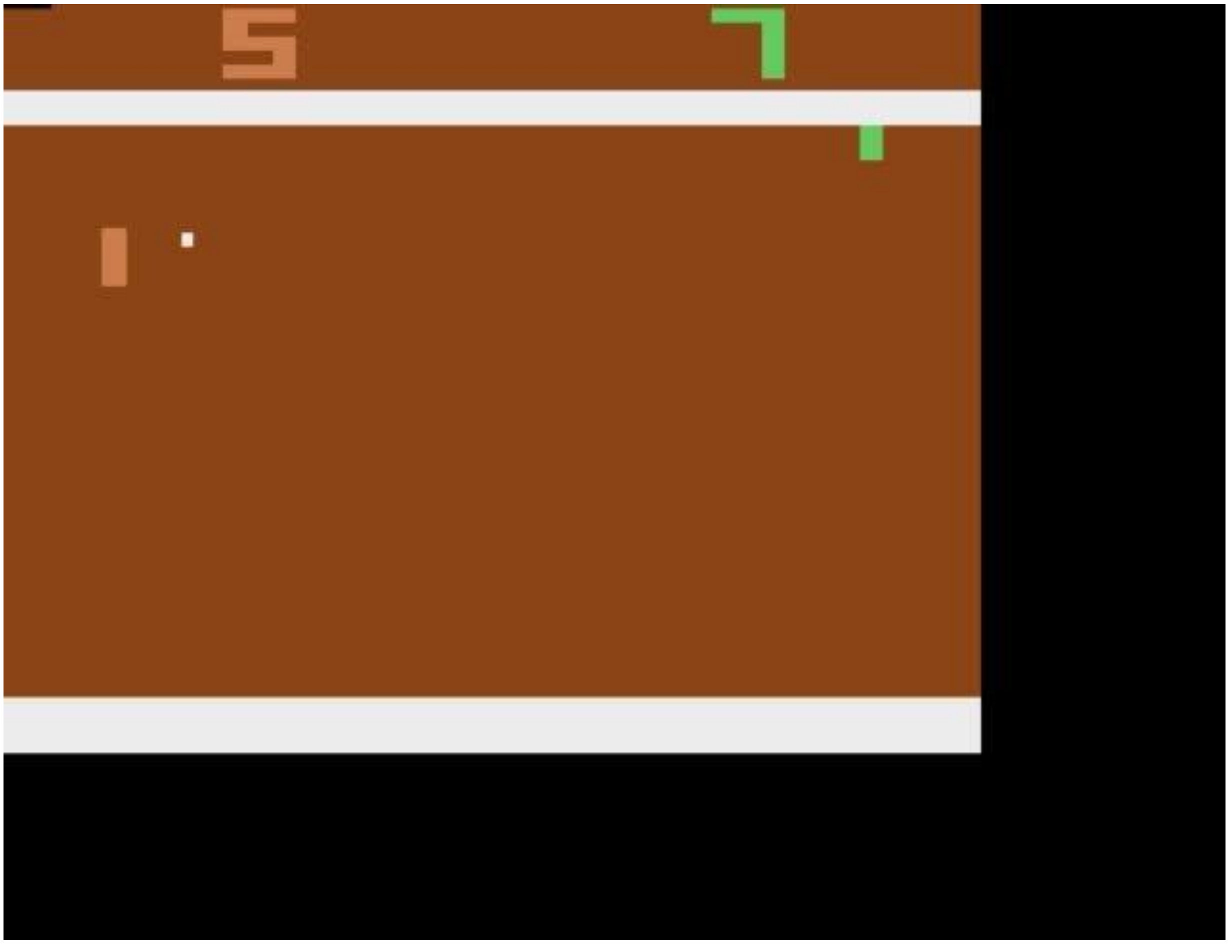
```
Training Only Final Layer
Epoch 1/5, Loss: 0.3680
Epoch 2/5, Loss: 0.1947
Epoch 3/5, Loss: 0.1535
Epoch 4/5, Loss: 0.1231
Epoch 5/5, Loss: 0.0975
Test Accuracy: 93.36%
Final Layer Only Accuracy: 93.36%
```

```
# Fine-Tune the Whole Model
print("Fine-Tuning the Entire Model")
train_model(model, train_loader, criterion, optimizer, epochs=5)
fine_tune_accuracy = evaluate_model(model, test_loader)
print(f"Fine-Tune Accuracy: {fine_tune_accuracy}%")
```

```
Fine-Tuning the Entire Model
Epoch 1/5, Loss: 0.0742
Epoch 2/5, Loss: 0.0558
Epoch 3/5, Loss: 0.0384
Epoch 4/5, Loss: 0.0284
Epoch 5/5, Loss: 0.0206
Test Accuracy: 93.96%
Fine-Tune Accuracy: 93.96%
```

Problem 4

The Arcade Learning Environment 2024-11-07 19-12-46.mp4



URL->
https://wpi0-my.sharepoint.com/:v:/g/personal/aggole_wpi_edu/Ea9aJXzr6zxMuneJuvbPTZUBs41q_4URFVAsO94-wbmKkQ?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJTdHJlYW1XZWJBcHAiLCJyZWZlcnJhbFZpZXciOiJTaGFyZURpYWxvZy1MaW5rIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXcifX0%3D&e=kO74Dn