

Worksheet2 (Output)

```
Dataset Preview:
  label pixel_0 pixel_1 pixel_2 pixel_3 pixel_4 pixel_5 pixel_6 \
0      5      0      0      0      0      0      0      0
1      0      0      0      0      0      0      0      0
2      4      0      0      0      0      0      0      0
3      1      0      0      0      0      0      0      0
4      9      0      0      0      0      0      0      0

  pixel_7 pixel_8 ... pixel_774 pixel_775 pixel_776 pixel_777 \
0      0      0 ...          0          0          0          0
1      0      0 ...          0          0          0          0
2      0      0 ...          0          0          0          0
3      0      0 ...          0          0          0          0
4      0      0 ...          0          0          0          0

  pixel_778 pixel_779 pixel_780 pixel_781 pixel_782 pixel_783
0          0          0          0          0          0          0
1          0          0          0          0          0          0
2          0          0          0          0          0          0
3          0          0          0          0          0          0
4          0          0          0          0          0          0

[5 rows x 785 columns]

Dataset Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60000 entries, 0 to 59999
Columns: 785 entries, label to pixel_783
dtypes: int64(785)
memory usage: 359.3 MB
None
```

```
Unique Classes: [0]
Encoded Labels: [0]
One-Hot Encoded Labels:
[[1.]
 [1.]
 [1.]
 [1.]
 [1.]]
```

Shapes:

X_train: (48000, 783) y_train: (48000, 1)

X_test: (12000, 783) y_test: (12000, 1)

Softmax function passed the test case!

Predicted class labels: [1 1 0]

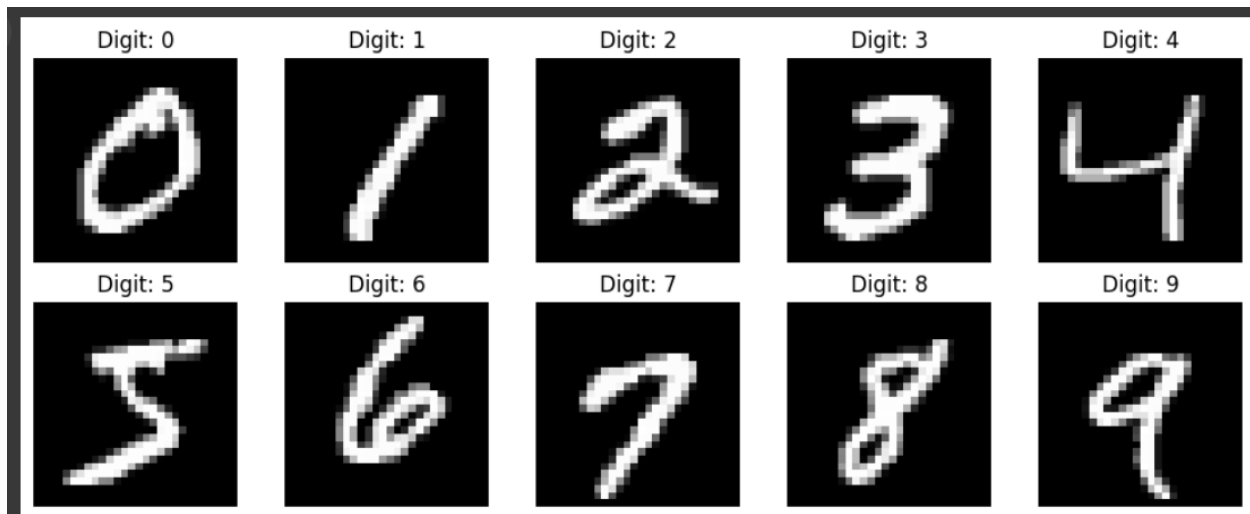
Cross-Entropy Loss (Correct Predictions): 0.1435

Cross-Entropy Loss (Incorrect Predictions): 2.9957

Gradient w.r.t. W: $\begin{bmatrix} 0.1031051 & 0.01805685 & -0.12116196 \\ -0.13600547 & 0.00679023 & 0.12921524 \end{bmatrix}$

Gradient w.r.t. b: $[-0.03290036 \quad 0.02484708 \quad 0.00805328]$

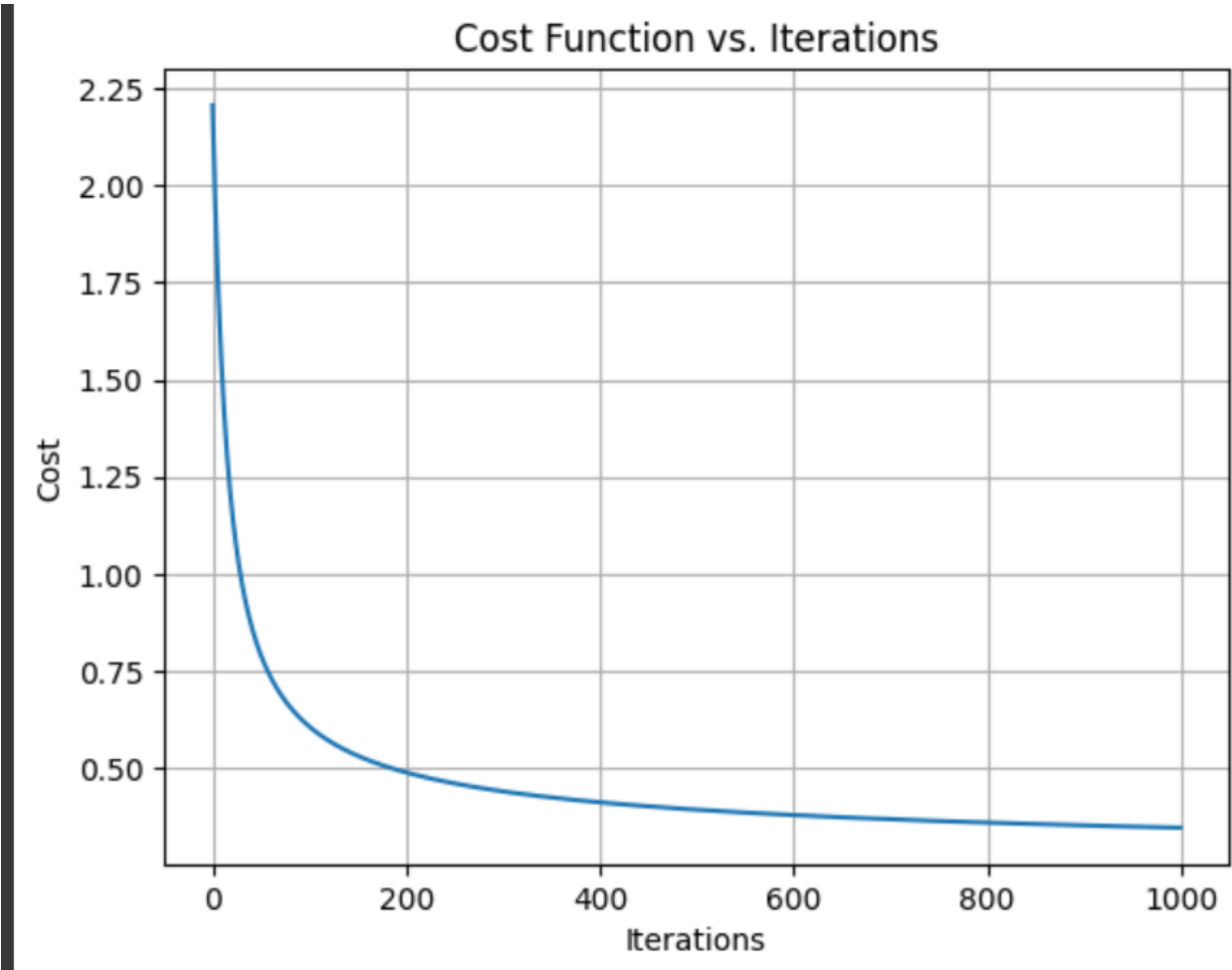
Test passed!

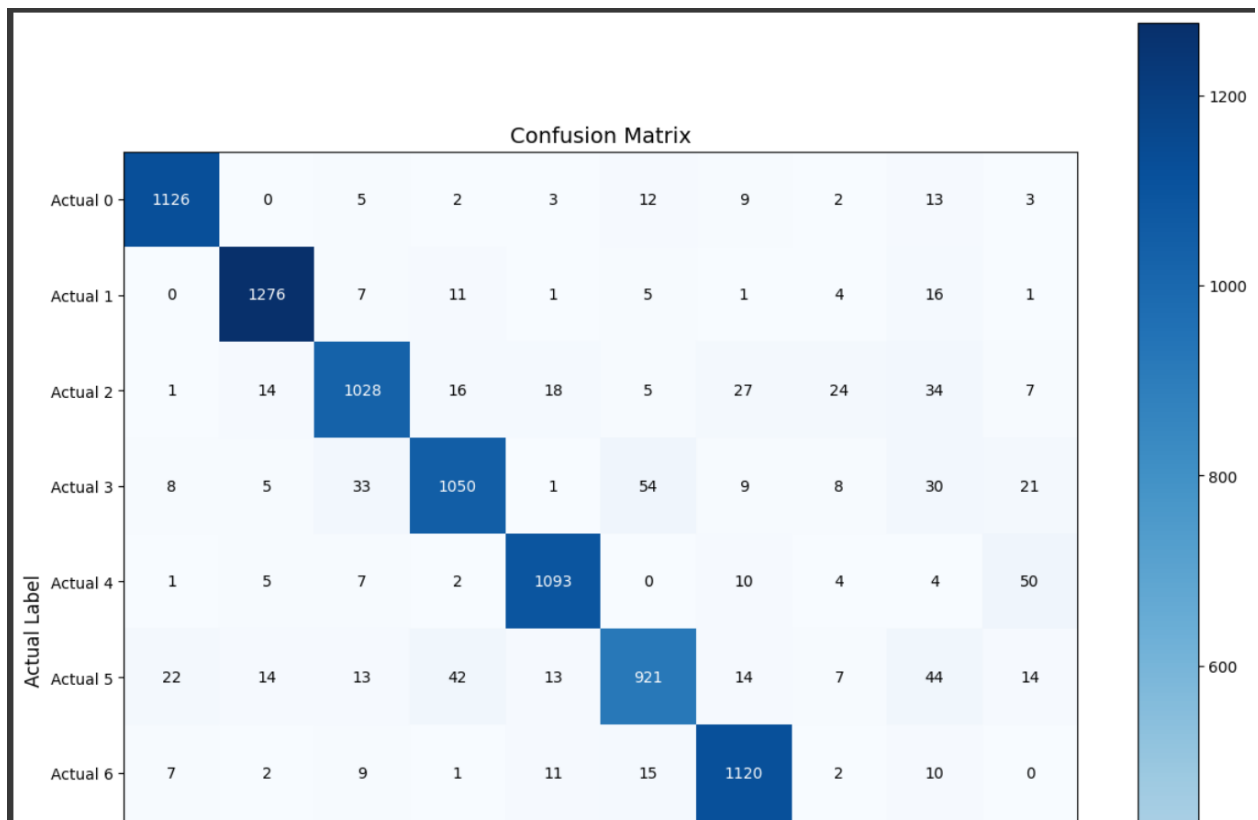


Move forward: Dimension of Feature Matrix X and label vector y matched.

```
Training data shape: (48000, 784)
Test data shape: (12000, 784)
```

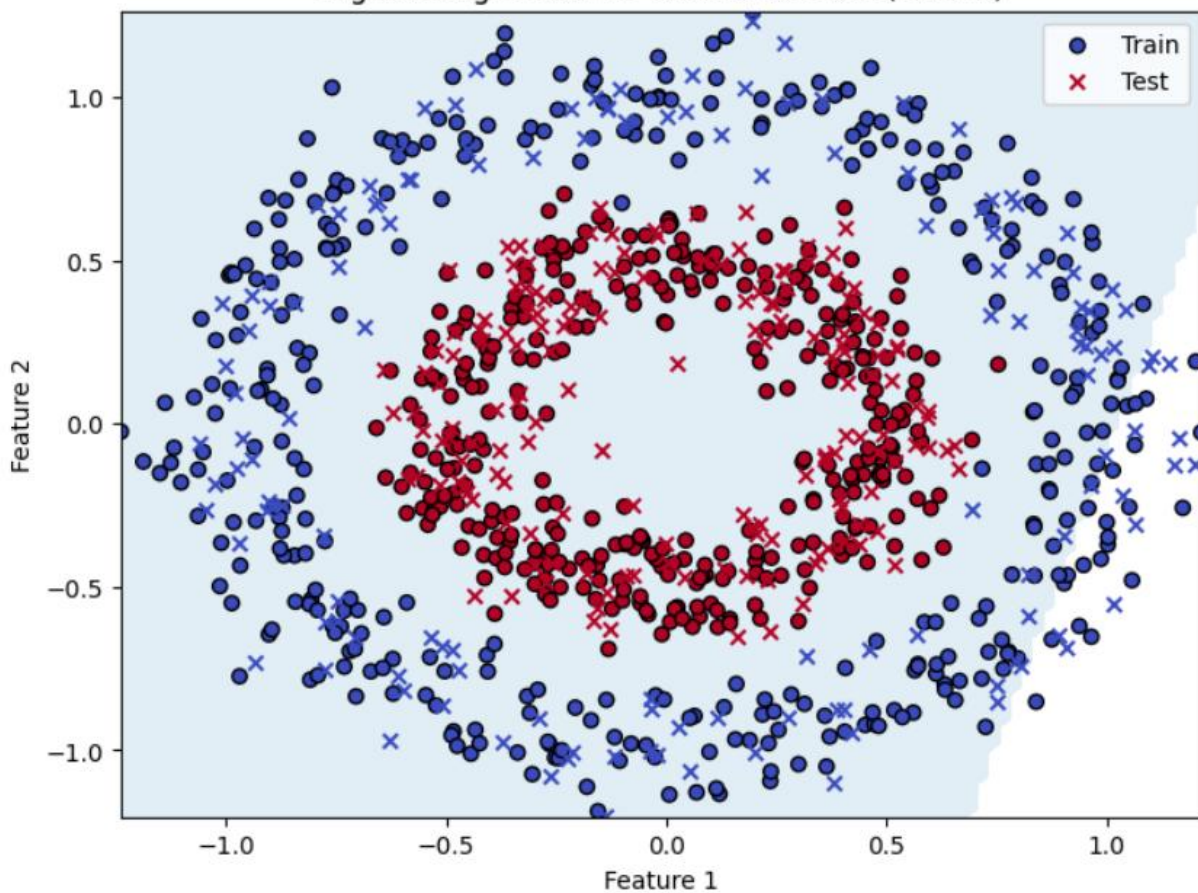
```
Iteration 0: Cost = 2.205990
Iteration 100: Cost = 0.607554
Iteration 200: Cost = 0.489663
Iteration 300: Cost = 0.440985
Iteration 400: Cost = 0.412865
Iteration 500: Cost = 0.393972
Iteration 600: Cost = 0.380134
Iteration 700: Cost = 0.369415
Iteration 800: Cost = 0.360783
Iteration 900: Cost = 0.353627
```

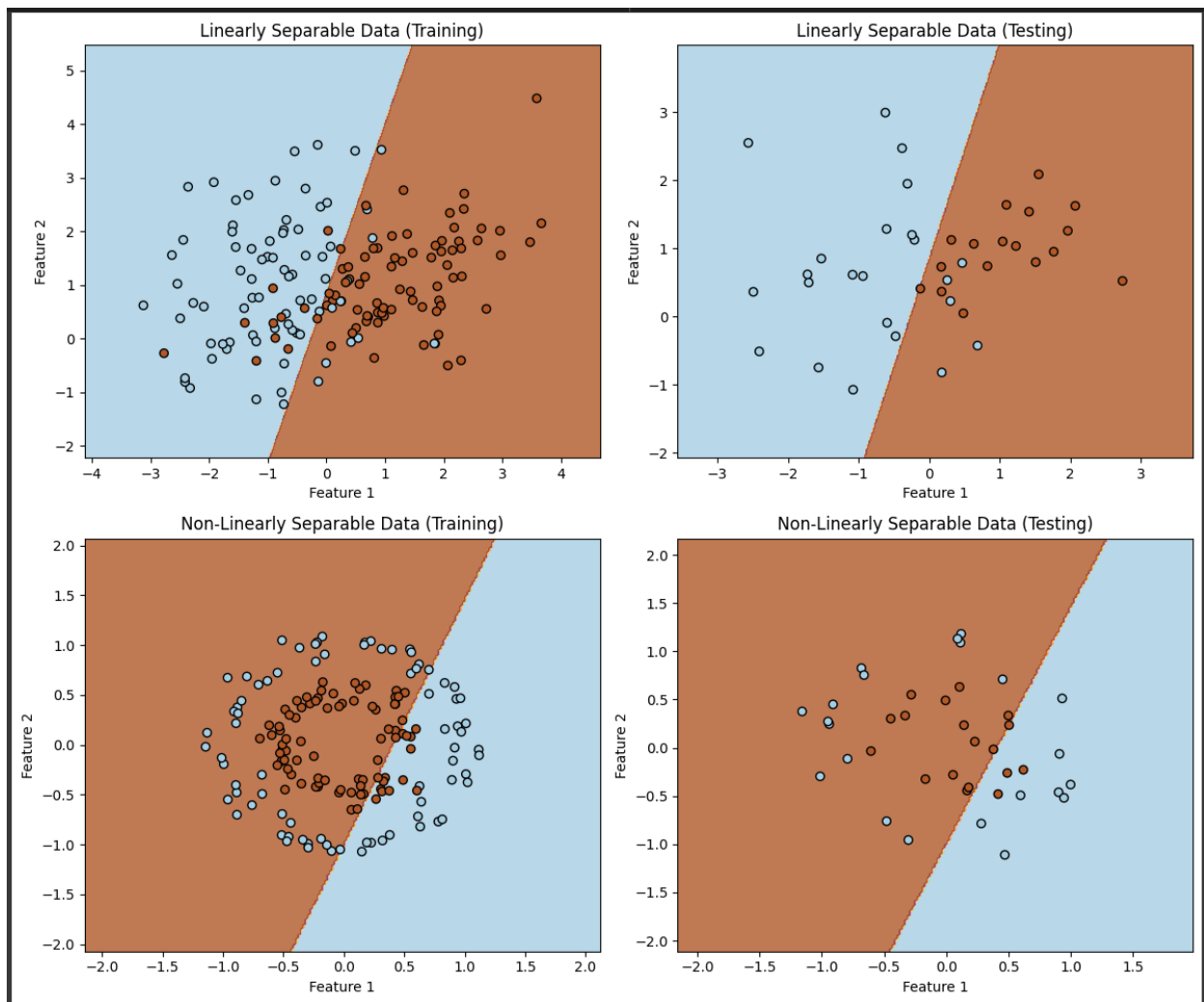




```
plt.scatter(X_test[:, 0], X_test[:, 1], c=y_test, cmap='coolwarm', edgecolors='k', marker='x')
```

Logistic Regression on Non-linear Data (Circles)





- For linearly separable data, the decision boundary effectively separates the two classes, leading to accurate predictions in both training and testing. This indicates that logistic regression is well-suited for linear classification problems.
- However, for non-linearly separable data, the linear nature of the decision boundary fails to separate the circular data points correctly. As a result, many points are misclassified, showing that logistic regression is not suitable for non-linear classification tasks.