```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Some Helper Function:

Softmax Function:

```
import numpy as np

def softmax(z):
    """
    Compute the softmax probabilities for a given input matrix.

    Parameters:
    z (numpy.ndarray): Logits (raw scores) of shape (m, n), where
                       - m is the number of samples.
                       - n is the number of classes.

    Returns:
    numpy.ndarray: Softmax probability matrix of shape (m, n), where
                   each row sums to 1 and represents the probability
                   distribution over classes.

    Notes:
    - The input to softmax is typically computed as: z = XW + b.
    - Uses numerical stabilization by subtracting the max value per row.
    """

    # Numerical stability trick: subtract max value in each row
    z_max = np.max(z, axis=1, keepdims=True)
    z_stable = z - z_max

    exp_z = np.exp(z_stable)  # Compute exponentials
    softmax_probs = exp_z / np.sum(exp_z, axis=1, keepdims=True)  # Normalize

    return softmax_probs
```

Softmax Test Case:

This test case checks that each row in the resulting softmax probabilities sums to 1, which is the fundamental property of softmax.

```
import numpy as np

def softmax(z):
    """
    Compute the softmax probabilities for a given input matrix.

    Parameters:
    z (numpy.ndarray): Logits (raw scores) of shape (m, n), where
                       - m is the number of samples.
                       - n is the number of classes.

    Returns:
    numpy.ndarray: Softmax probability matrix of shape (m, n), where
                   each row sums to 1 and represents the probability
                   distribution over classes.

    Notes:
    - The input to softmax is typically computed as: z = XW + b.
    - Uses numerical stabilization by subtracting the max value per row.
    """

    # Numerical stability trick: subtract max value in each row
    z_max = np.max(z, axis=1, keepdims=True)
    z_stable = z - z_max

    exp_z = np.exp(z_stable)  # Compute exponentials
```

```
    softmax_probs = exp_z / np.sum(exp_z, axis=1, keepdims=True)  # Normalize

    return softmax_probs

# Example test case
z_test = np.array([[2.0, 1.0, 0.1], [1.0, 1.0, 1.0]])
softmax_output = softmax(z_test)

# Verify if the sum of probabilities for each row is 1
row_sums = np.sum(softmax_output, axis=1)

# Assert that the sum of each row is 1
assert np.allclose(row_sums, 1), f"Test failed: Row sums are {row_sums}"

print("Softmax function passed the test case! ✅")
```

```
Softmax function passed the test case! ✅
```

Prediction Function:

```
import numpy as np

def predict_softmax(X, W, b):
    """
    Predict the class labels for a set of samples using the trained softmax model.

    Parameters:
    X (numpy.ndarray): Feature matrix of shape (n, d), where n is the number of samples and d is the number of features.
    W (numpy.ndarray): Weight matrix of shape (d, c), where c is the number of classes.
    b (numpy.ndarray): Bias vector of shape (c,).

    Returns:
    numpy.ndarray: Predicted class labels of shape (n,), where each value is the index of the predicted class.
    """
    logits = np.dot(X, W) + b  # Compute raw scores (logits)
    softmax_probs = softmax(logits)  # Apply softmax function
    predicted_classes = np.argmax(softmax_probs, axis=1)  # Get index of max probability for each sample

    return predicted_classes
```

Test Function for Prediction Function:

The test function ensures that the predicted class labels have the same number of elements as the input samples, verifying that the model produces a valid output shape.

```
import numpy as np

def softmax(z):
    """
    Compute the softmax probabilities for a given input matrix.

    Parameters:
    z (numpy.ndarray): Logits (raw scores) of shape (m, n), where
                       - m is the number of samples.
                       - n is the number of classes.

    Returns:
    numpy.ndarray: Softmax probability matrix of shape (m, n), where
                   each row sums to 1.
    """
    z_max = np.max(z, axis=1, keepdims=True)  # Numerical stability
    z_stable = z - z_max
    exp_z = np.exp(z_stable)
    return exp_z / np.sum(exp_z, axis=1, keepdims=True)

def predict_softmax(X, W, b):
    """
    Predict the class labels for a set of samples using the trained softmax model.

    Parameters:
```

```
    X (numpy.ndarray): Feature matrix of shape (n, d).
    W (numpy.ndarray): Weight matrix of shape (d, c).
    b (numpy.ndarray): Bias vector of shape (c,).

    Returns:
    numpy.ndarray: Predicted class labels of shape (n,).
    """
    logits = np.dot(X, W) + b  # Compute raw scores (logits)
    softmax_probs = softmax(logits)  # Apply softmax
    return np.argmax(softmax_probs, axis=1)  # Get the class with the highest probability

# **Test Case for Prediction Function**
X_test = np.array([[0.2, 0.8], [0.5, 0.5], [0.9, 0.1]])  # 3 samples, 2 features
W_test = np.array([[0.4, 0.2, 0.1], [0.3, 0.7, 0.5]])  # 2 features, 3 classes
b_test = np.array([0.1, 0.2, 0.3])  # Bias for 3 classes

# Expected output: Array of predicted class labels (each between 0 and 2)
y_pred_test = predict_softmax(X_test, W_test, b_test)

# Validate output shape
assert y_pred_test.shape == (3,), f"Test failed: Expected shape (3,), got {y_pred_test.shape}"

# Print the predicted labels
print("Predicted class labels:", y_pred_test)
```

```
Predicted class labels: [1 1 0]
```

Loss Function:

```
import numpy as np

def loss_softmax(y_pred, y):
    """
    Compute the cross-entropy loss for a single sample.

    Parameters:
    y_pred (numpy.ndarray): Predicted probabilities of shape (c,) for a single sample,
                             where c is the number of classes.
    y (numpy.ndarray): True labels (one-hot encoded) of shape (c,), where c is the number of classes.

    Returns:
    float: Cross-entropy loss for the given sample.
    """
    # Add a small value (epsilon) to prevent log(0) errors
    epsilon = 1e-12
    y_pred = np.clip(y_pred, epsilon, 1.0 - epsilon)

    # Compute the cross-entropy loss
    loss = -np.sum(y * np.log(y_pred))

    return loss
```

Test case for Loss Function:

This test case Compares loss for correct vs. incorrect predictions.

- Expects low loss for correct predictions.
- Expects high loss for incorrect predictions.

```
import numpy as np

# Define correct predictions (low loss scenario)
y_true_correct = np.array([[1, 0, 0], [0, 1, 0], [0, 0, 1]])  # True one-hot labels
y_pred_correct = np.array([[0.9, 0.05, 0.05],
                           [0.1, 0.85, 0.05],
                           [0.05, 0.1, 0.85]])  # High confidence in the correct class

# Define incorrect predictions (high loss scenario)
y_pred_incorrect = np.array([[0.05, 0.05, 0.9],  # Highly confident in the wrong class
                              [0.1, 0.05, 0.85],
```

```
                              [0.85, 0.1, 0.05]])

# Compute loss for both cases
loss_correct = loss_softmax(y_pred_correct, y_true_correct)
loss_incorrect = loss_softmax(y_pred_incorrect, y_true_correct)

# Validate that incorrect predictions lead to a higher loss
assert loss_correct < loss_incorrect, f"Test failed: Expected loss_correct < loss_incorrect, but got {loss_correct:.4f} >= {loss_incorrect:.

# Print results
print(f"Cross-Entropy Loss (Correct Predictions): {loss_correct:.4f}")
print(f"Cross-Entropy Loss (Incorrect Predictions): {loss_incorrect:.4f}")
```

```
Cross-Entropy Loss (Correct Predictions): 0.4304
Cross-Entropy Loss (Incorrect Predictions): 8.9872
```

Cost Function:

```
import numpy as np

def cost_softmax(X, y, W, b):
    """
    Compute the average softmax regression cost (cross-entropy loss) over all samples.

    Parameters:
    X (numpy.ndarray): Feature matrix of shape (n, d), where n is the number of samples and d is the number of features.
    y (numpy.ndarray): True labels (one-hot encoded) of shape (n, c), where n is the number of samples and c is the number of classes.
    W (numpy.ndarray): Weight matrix of shape (d, c).
    b (numpy.ndarray): Bias vector of shape (c,).

    Returns:
    float: Average softmax cost (cross-entropy loss) over all samples.
    """
    # Compute logits: Z = XW + b
    logits = np.dot(X, W) + b

    # Compute softmax probabilities
    softmax_probs = softmax(logits)

    # Numerical stability: clip probabilities to avoid log(0) errors
    epsilon = 1e-12
    softmax_probs = np.clip(softmax_probs, epsilon, 1.0 - epsilon)

    # Compute cross-entropy loss for each sample
    total_loss = -np.sum(y * np.log(softmax_probs))

    # Compute the average loss
    n = X.shape[0]  # Number of samples
    return total_loss / n



# Example input data
X_test = np.array([[0.2, 0.8], [0.5, 0.5], [0.9, 0.1]])  # (3 samples, 2 features)
W_test = np.array([[0.4, 0.2, 0.1], [0.3, 0.7, 0.5]])  # (2 features, 3 classes)
b_test = np.array([0.1, 0.2, 0.3])  # (3 classes)

# True labels (one-hot encoded)
y_test = np.array([[0, 1, 0], [1, 0, 0], [0, 0, 1]])

# Compute cost
cost_value = cost_softmax(X_test, y_test, W_test, b_test)
print("Softmax cost:", cost_value)
```

```
Softmax cost: 1.09862371811174
```

Test Case for Cost Function:

The test case assures that the cost for the incorrect prediction should be higher than for the correct prediction, confirming that the cost function behaves as expected.

```
import numpy as np
```

```
# Example 1: Correct Prediction (Closer predictions)
X_correct = np.array([[1.0, 0.0], [0.0, 1.0]])  # Feature matrix for correct predictions
y_correct = np.array([[1, 0], [0, 1]])  # True labels (one-hot encoded, matching predictions)
W_correct = np.array([[5.0, -2.0], [-3.0, 5.0]])  # Weights for correct prediction
b_correct = np.array([0.1, 0.1])  # Bias for correct prediction

# Example 2: Incorrect Prediction (Far off predictions)
X_incorrect = np.array([[0.1, 0.9], [0.8, 0.2]])  # Feature matrix for incorrect predictions
y_incorrect = np.array([[1, 0], [0, 1]])  # True labels (one-hot encoded, incorrect predictions)
W_incorrect = np.array([[0.1, 2.0], [1.5, 0.3]])  # Weights for incorrect prediction
b_incorrect = np.array([0.5, 0.6])  # Bias for incorrect prediction

# Compute cost for correct predictions
cost_correct = cost_softmax(X_correct, y_correct, W_correct, b_correct)

# Compute cost for incorrect predictions
cost_incorrect = cost_softmax(X_incorrect, y_incorrect, W_incorrect, b_incorrect)

# Check if the cost for incorrect predictions is greater than for correct predictions
assert cost_incorrect > cost_correct, f"Test failed: Incorrect cost {cost_incorrect} is not greater than correct cost {cost_correct}"

# Print the costs for verification
print("Cost for correct prediction:", cost_correct)
print("Cost for incorrect prediction:", cost_incorrect)

print("Test passed!")
```

```
Cost for correct prediction: 0.0006234364133349324
Cost for incorrect prediction: 0.29930861359446115
Test passed!
```

Computing Gradients:

```
import numpy as np

def compute_gradient_softmax(X, y, W, b):
    """
    Compute the gradients of the cost function with respect to weights and biases.

    Parameters:
    X (numpy.ndarray): Feature matrix of shape (n, d).
    y (numpy.ndarray): True labels (one-hot encoded) of shape (n, c).
    W (numpy.ndarray): Weight matrix of shape (d, c).
    b (numpy.ndarray): Bias vector of shape (c,).

    Returns:
    tuple: Gradients with respect to weights (d, c) and biases (c,).
    """
    # Compute logits
    logits = np.dot(X, W) + b

    # Compute softmax probabilities
    softmax_probs = softmax(logits)

    # Compute the error (difference between predicted probabilities and true labels)
    error = softmax_probs - y  # Shape: (n, c)

    # Compute gradients
    n = X.shape[0]  # Number of samples
    grad_W = np.dot(X.T, error) / n  # Gradient w.r.t. weights, shape: (d, c)
    grad_b = np.sum(error, axis=0) / n  # Gradient w.r.t. biases, shape: (c,)

    return grad_W, grad_b



# Example input data
X_test = np.array([[0.2, 0.8], [0.5, 0.5], [0.9, 0.1]])  # (3 samples, 2 features)
W_test = np.array([[0.4, 0.2, 0.1], [0.3, 0.7, 0.5]])  # (2 features, 3 classes)
b_test = np.array([0.1, 0.2, 0.3])  # (3 classes)

# True labels (one-hot encoded)
y_test = np.array([[0, 1, 0], [1, 0, 0], [0, 0, 1]])

# Compute gradients
```

```
# Compute gradients
grad_W, grad_b = compute_gradient_softmax(X_test, y_test, W_test, b_test)

# Print gradients
print("Gradient w.r.t Weights:\n", grad_W)
print("Gradient w.r.t Biases:\n", grad_b)
```

```
Gradient w.r.t Weights:
 [[ 0.0031051   0.11805685 -0.12116196]
 [-0.03600547 -0.09320977  0.12921524]]
Gradient w.r.t Biases:
 [-0.03290036  0.02484708  0.00805328]
```

Test case for compute_gradient function:

The test checks if the gradients from the function are close enough to the manually computed gradients using np.allclose, which accounts for potential floating-point discrepancies.

```
import numpy as np

# Define a simple feature matrix and true labels
X_test = np.array([[0.2, 0.8], [0.5, 0.5], [0.9, 0.1]])  # Feature matrix (3 samples, 2 features)
y_test = np.array([[1, 0, 0], [0, 1, 0], [0, 0, 1]])  # True labels (one-hot encoded, 3 classes)

# Define weight matrix and bias vector
W_test = np.array([[0.4, 0.2, 0.1], [0.3, 0.7, 0.5]])  # Weights (2 features, 3 classes)
b_test = np.array([0.1, 0.2, 0.3])  # Bias (3 classes)

# Compute the gradients using the function
grad_W, grad_b = compute_gradient_softmax(X_test, y_test, W_test, b_test)

# Manually compute the predicted probabilities (using softmax function)
z_test = np.dot(X_test, W_test) + b_test
y_pred_test = softmax(z_test)

# Compute the manually computed gradients
grad_W_manual = np.dot(X_test.T, (y_pred_test - y_test)) / X_test.shape[0]
grad_b_manual = np.sum(y_pred_test - y_test, axis=0) / X_test.shape[0]

# Assert that the gradients computed by the function match the manually computed gradients
assert np.allclose(grad_W, grad_W_manual), f"Test failed: Gradients w.r.t. W are not equal.\nExpected: {grad_W_manual}\nGot: {grad_W}"
assert np.allclose(grad_b, grad_b_manual), f"Test failed: Gradients w.r.t. b are not equal.\nExpected: {grad_b_manual}\nGot: {grad_b}"

# Print the gradients for verification
print("Gradient w.r.t. W:", grad_W)
print("Gradient w.r.t. b:", grad_b)

print("Test passed!")
```

```
Gradient w.r.t. W: [[ 0.1031051   0.01805685 -0.12116196]
 [-0.13600547  0.00679023  0.12921524]]
Gradient w.r.t. b: [-0.03290036  0.02484708  0.00805328]
Test passed!
```

Implementing Gradient Descent:

```
import numpy as np

def gradient_descent_softmax(X, y, W, b, alpha, n_iter, show_cost=False):
    """
    Perform gradient descent to optimize the weights and biases.

    Parameters:
    X (numpy.ndarray): Feature matrix of shape (n, d).
    y (numpy.ndarray): True labels (one-hot encoded) of shape (n, c).
    W (numpy.ndarray): Weight matrix of shape (d, c).
    b (numpy.ndarray): Bias vector of shape (c,).
    alpha (float): Learning rate.
    n_iter (int): Number of iterations.
    show_cost (bool): Whether to display the cost at intervals.

    Returns:
```

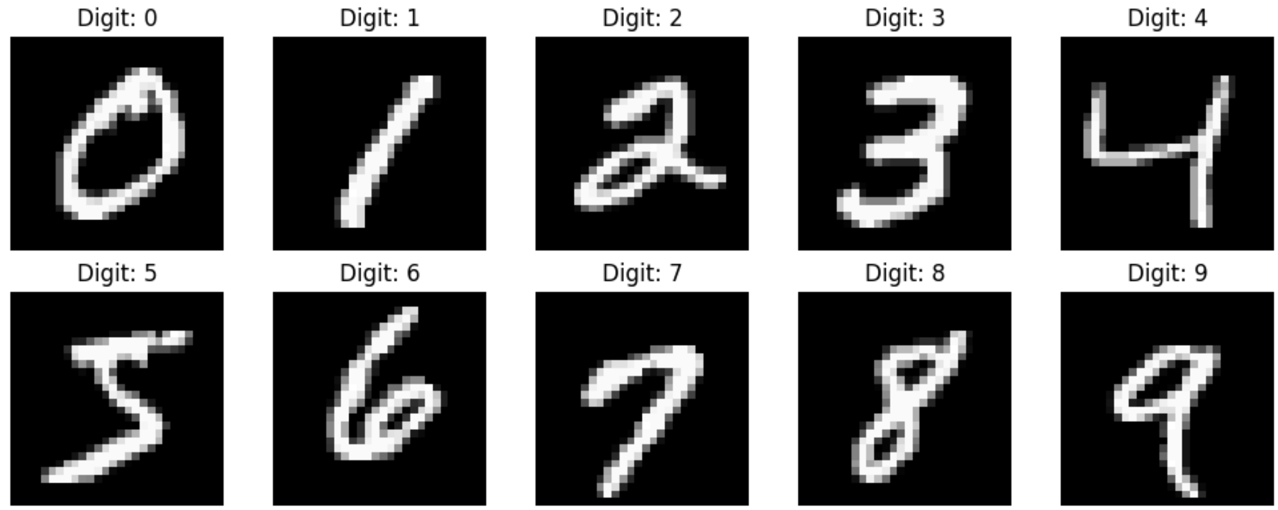
```
    tuple: Optimized weights, biases, and cost history.
    """
    cost_history = []

    for i in range(n_iter):
        # Compute gradients
        grad_W, grad_b = compute_gradient_softmax(X, y, W, b)

        # Update weights and biases using gradient descent
        W -= alpha * grad_W
        b -= alpha * grad_b

        # Compute cost
        cost = cost_softmax(X, y, W, b)
        cost_history.append(cost)

        # Display cost at intervals
        if show_cost and (i % 100 == 0 or i == n_iter - 1):
            print(f"Iteration {i}: Cost = {cost}")

    return W, b, cost_history



# Example input data
X_test = np.array([[0.2, 0.8], [0.5, 0.5], [0.9, 0.1]])  # (3 samples, 2 features)
W_test = np.random.rand(2, 3)  # Random initialization (2 features, 3 classes)
b_test = np.random.rand(3)  # Random initialization (3 classes)
y_test = np.array([[0, 1, 0], [1, 0, 0], [0, 0, 1]])  # One-hot encoded labels

# Hyperparameters
alpha = 0.1  # Learning rate
n_iter = 1000  # Number of iterations

# Run gradient descent
W_opt, b_opt, cost_hist = gradient_descent_softmax(X_test, y_test, W_test, b_test, alpha, n_iter, show_cost=True)

# Print final optimized weights and biases
print("Optimized Weights:\n", W_opt)
print("Optimized Biases:\n", b_opt)
```

```
Iteration 0: Cost = 1.1798757910598126
Iteration 100: Cost = 0.7870631729938221
Iteration 200: Cost = 0.6268705497085009
Iteration 300: Cost = 0.5367725728401082
Iteration 400: Cost = 0.4755970867696704
Iteration 500: Cost = 0.42951381754933693
Iteration 600: Cost = 0.3926222985407403
Iteration 700: Cost = 0.36194683735713107
Iteration 800: Cost = 0.3357930781515659
Iteration 900: Cost = 0.31310300343208225
Iteration 999: Cost = 0.29335414772356894
Optimized Weights:
 [[ 0.76754981 -3.46676061  4.09294856]
 [ 0.93835218  4.42541883 -3.6698721 ]]
Optimized Biases:
 [ 0.68905651  0.03164296 -0.09163724]
```

Preparing Dataset:

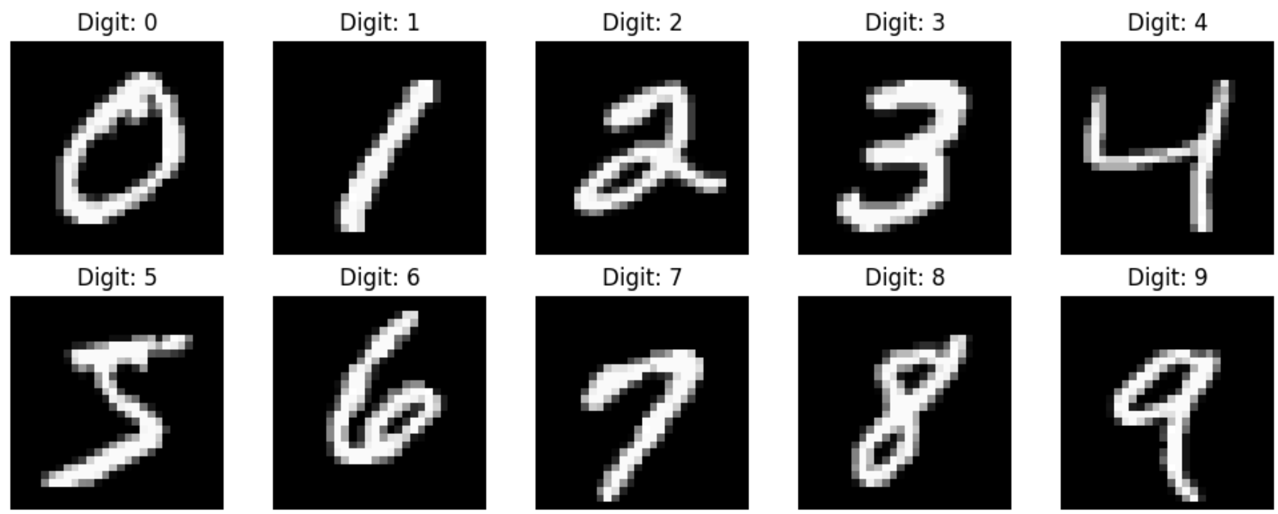
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

# Define the dataset path
csv_file = "/content/drive/MyDrive/AI/week2/worksheet-2/mnist_dataset.csv"

def load_and_prepare_mnist(csv_file, test_size=0.2, random_state=42):
    """
    Reads the MNIST CSV file, splits data into train/test sets, and plots one image per class.

    Arguments:
    csv_file (str)       : Path to the CSV file containing MNIST data.
    test_size (float)    : Proportion of the data to use as the test set (default: 0.2).
```

```
        random_state (int)   : Random seed for reproducibility (default: 42).

        Returns:
        X_train, X_test, y_train, y_test : Split dataset.
        """

        # Load dataset
        df = pd.read_csv(csv_file)

        # Separate labels and features
        y = df.iloc[:, 0].values  # First column is the label
        X = df.iloc[:, 1:].values  # Remaining columns are pixel values

        # Normalize pixel values (optional but recommended)
        X = X / 255.0  # Scale values between 0 and 1

        # Split data into train and test sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=random_state)

        # Plot one sample image per class
        plot_sample_images(X, y)

        return X_train, X_test, y_train, y_test

    def plot_sample_images(X, y):
        """
        Plots one sample image for each digit class (0-9).

        Arguments:
        X (np.ndarray): Feature matrix containing pixel values.
        y (np.ndarray): Labels corresponding to images.
        """

        plt.figure(figsize=(10, 4))
        unique_classes = np.unique(y)  # Get unique class labels

        for i, digit in enumerate(unique_classes):
            index = np.where(y == digit)[0][0]  # Find first occurrence of the class
            image = X[index].reshape(28, 28)  # Reshape 1D array to 28x28

            plt.subplot(2, 5, i + 1)
            plt.imshow(image, cmap='gray')
            plt.title(f"Digit: {digit}")
            plt.axis('off')

        plt.tight_layout()
        plt.show()

    # Load and process the MNIST dataset
    X_train, X_test, y_train, y_test = load_and_prepare_mnist(csv_file)

    # Print dataset shapes
    print(f"Training set shape: {X_train.shape}, Labels: {y_train.shape}")
    print(f"Test set shape: {X_test.shape}, Labels: {y_test.shape}")
```




```
Training set shape: (48000, 784), Labels: (48000,)
Test set shape: (12000, 784), Labels: (12000,)
```

```
csv_file_path = "/content/drive/MyDrive/AI/week2/worksheet-2/mnist_dataset.csv"  # Path to saved dataset
X_train, X_test, y_train, y_test = load_and_prepare_mnist(csv_file_path)
```



A Quick debugging Step:

```
# Assert that X and y have matching lengths
assert len(X_train) == len(y_train), f"Error: X and y have different lengths! X={len(X_train)}, y={len(y_train)}"
print("Move forward: Dimension of Feture Matrix X and label vector y matched.")
```

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

Train the Model:

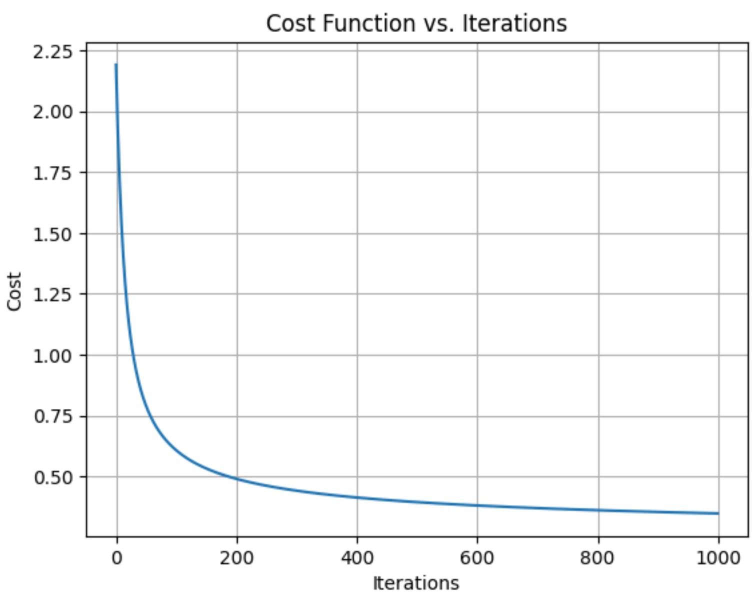
```
print(f"Training data shape: {X_train.shape}")
print(f"Test data shape: {X_test.shape}")
```

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

```
from sklearn.preprocessing import OneHotEncoder

# Check if y_train is one-hot encoded
if len(y_train.shape) == 1:
    encoder = OneHotEncoder(sparse_output=False)  # Use sparse_output=False for newer versions of sklearn
    y_train = encoder.fit_transform(y_train.reshape(-1, 1))  # One-hot encode labels
    y_test = encoder.transform(y_test.reshape(-1, 1))  # One-hot encode test labels
```

```
# Now y_train is one-hot encoded, and we can proceed to use it
d = X_train.shape[1]  # Number of features (columns in X_train)
c = y_train.shape[1]  # Number of classes (columns in y_train after one-hot encoding)

# Initialize weights with small random values and biases with zeros
W = np.random.randn(d, c) * 0.01  # Small random weights initialized
b = np.zeros(c)  # Bias initialized to 0

# Set hyperparameters for gradient descent
alpha = 0.1  # Learning rate
n_iter = 1000  # Number of iterations to run gradient descent

# Train the model using gradient descent
W_opt, b_opt, cost_history = gradient_descent_softmax(X_train, y_train, W, b, alpha, n_iter, show_cost=True)

# Plot the cost history to visualize the convergence
plt.plot(cost_history)
plt.title('Cost Function vs. Iterations')
plt.xlabel('Iterations')
plt.ylabel('Cost')
plt.grid(True)
plt.show()
```

```
Iteration 0: Cost = 2.1907523908880377
Iteration 100: Cost = 0.6071497628881436
Iteration 200: Cost = 0.4896775892376777
Iteration 300: Cost = 0.44108426078870444
Iteration 400: Cost = 0.4129891218247169
Iteration 500: Cost = 0.39410246103548874
Iteration 600: Cost = 0.3802628935679442
Iteration 700: Cost = 0.36954004085975195
Iteration 800: Cost = 0.3609019197961151
Iteration 900: Cost = 0.35374051873910445
Iteration 999: Cost = 0.34772746084789385
```

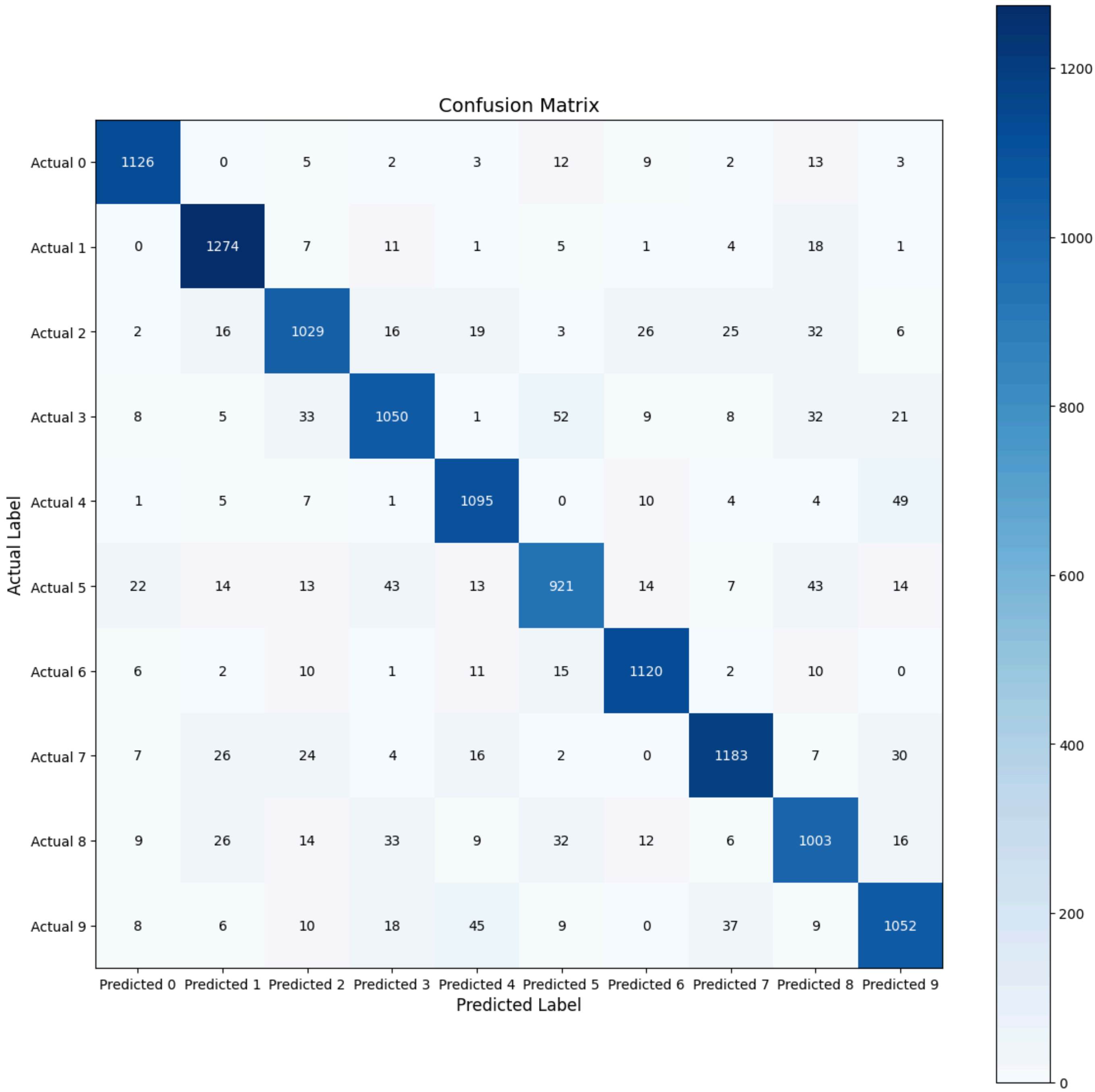


Evaluating the Model:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, precision_score, recall_score, f1_score

# Evaluation Function
def evaluate_classification(y_true, y_pred):
    """
    Evaluate classification performance using confusion matrix, precision, recall, and F1-score.

    Parameters:
    y_true (numpy.ndarray): True labels
    y_pred (numpy.ndarray): Predicted labels
```

```
        Returns:
        tuple: Confusion matrix, precision, recall, F1 score
        """
        # Compute confusion matrix
        cm = confusion_matrix(y_true, y_pred)

        # Compute precision, recall, and F1-score
        precision = precision_score(y_true, y_pred, average='weighted')
        recall = recall_score(y_true, y_pred, average='weighted')
        f1 = f1_score(y_true, y_pred, average='weighted')

        return cm, precision, recall, f1



    # Predict on the test set
    y_pred_test = predict_softmax(X_test, W_opt, b_opt)

    # Evaluate accuracy
    y_test_labels = np.argmax(y_test, axis=1)  # True labels in numeric form

    # Evaluate the model
    cm, precision, recall, f1 = evaluate_classification(y_test_labels, y_pred_test)

    # Print the evaluation metrics
    print("\nConfusion Matrix:")
    print(cm)
    print(f"Precision: {precision:.2f}")
    print(f"Recall: {recall:.2f}")
    print(f"F1-Score: {f1:.2f}")

    # Visualizing the Confusion Matrix
    fig, ax = plt.subplots(figsize=(12, 12))
    cax = ax.imshow(cm, cmap='Blues')  # Use a color map for better visualization

    # Dynamic number of classes
    num_classes = cm.shape[0]
    ax.set_xticks(range(num_classes))
    ax.set_yticks(range(num_classes))
    ax.set_xticklabels([f'Predicted {i}' for i in range(num_classes)])
    ax.set_yticklabels([f'Actual {i}' for i in range(num_classes)])

    # Add labels to each cell in the confusion matrix
    for i in range(cm.shape[0]):
        for j in range(cm.shape[1]):
            ax.text(j, i, cm[i, j], ha='center', va='center', color='white' if cm[i, j] > np.max(cm) / 2 else 'black')

    # Add grid lines and axis labels
    ax.grid(False)
    plt.title('Confusion Matrix', fontsize=14)
    plt.xlabel('Predicted Label', fontsize=12)
    plt.ylabel('Actual Label', fontsize=12)

    # Adjust layout
    plt.tight_layout()
    plt.colorbar(cax)
    plt.show()
```

```
Confusion Matrix:
[[1126    0    5    2    3   12    9    2   13    3]
 [   0 1274    7   11    1    5    1    4   18    1]
 [   2   16 1029   16   19    3   26   25   32    6]
 [   8    5   33 1050    1   52    9    8   32   21]
 [   1    5    7    1 1095    0   10    4    4   49]
 [  22   14   13   43   13  921   14    7   43   14]
 [   6    2   10    1   11   15 1120    2   10    0]
 [   7   26   24    4   16    2    0 1183    7   30]
 [   9   26   14   33    9   32   12    6 1003   16]
 [   8    6   10   18   45    9    0   37    9 1052]]
Precision: 0.90
Recall: 0.90
F1-Score: 0.90
```



Linear Seperability and Logistic Regression:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_classification, make_circles
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression

# Set random seed for reproducibility
np.random.seed(42)

# Generate a synthetic dataset that is linearly separable
X_linear_separable, y_linear_separable = make_classification(
    n_samples=200, n_features=2, n_informative=2, n_redundant=0, n_clusters_per_class=1, random_state=42
)

# Split the dataset into training and testing sets (80% train, 20% test)
X_train_linear, X_test_linear, y_train_linear, y_test_linear = train_test_split(
    X_linear_separable, y_linear_separable, test_size=0.2, random_state=42
)

# Train a Logistic Regression model on the linearly separable dataset
logistic_model_linear_separable = LogisticRegression()
logistic_model_linear_separable.fit(X_train_linear, y_train_linear)

# Generate a synthetic dataset that is non-linearly separable (circles pattern)
X_non_linear_separable, y_non_linear_separable = make_circles(
    n_samples=200, noise=0.1, factor=0.5, random_state=42
)

# Split the dataset into training and testing sets (80% train, 20% test)
X_train_non_linear, X_test_non_linear, y_train_non_linear, y_test_non_linear = train_test_split(
    X_non_linear_separable, y_non_linear_separable, test_size=0.2, random_state=42
)

# Train a Logistic Regression model on the non-linearly separable dataset
logistic_model_non_linear_separable = LogisticRegression()
logistic_model_non_linear_separable.fit(X_train_non_linear, y_train_non_linear)

# Function to plot the decision boundary of a trained model
def plot_decision_boundary(ax, model, X, y, title):
    h = 0.02  # Step size for mesh grid
    x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
    y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1

    # Create a mesh grid over the feature space
    xx, yy = np.meshgrid(np.arange(x_min, x_max, h), np.arange(y_min, y_max, h))

    # Predict class labels for each point in the grid
    Z = model.predict(np.c_[xx.ravel(), yy.ravel()])
    Z = Z.reshape(xx.shape)

    # Plot the decision boundary using contour plot
    ax.contourf(xx, yy, Z, alpha=0.8, cmap=plt.cm.Paired)

    # Scatter plot of the actual data points
    ax.scatter(X[:, 0], X[:, 1], c=y, edgecolors='k', cmap=plt.cm.Paired)

    # Formatting the plot
    ax.set_title(title)
    ax.set_xlabel('Feature 1')
    ax.set_ylabel('Feature 2')

# Create subplots to visualize decision boundaries
fig, axes = plt.subplots(2, 2, figsize=(12, 10))

# Plot decision boundary for linearly separable data (Training set)
plot_decision_boundary(axes[0, 0], logistic_model_linear_separable, X_train_linear, y_train_linear, 'Linearly Separable Data (Training)')

# Plot decision boundary for linearly separable data (Testing set)
plot_decision_boundary(axes[0, 1], logistic_model_linear_separable, X_test_linear, y_test_linear, 'Linearly Separable Data (Testing)')

# Plot decision boundary for non-linearly separable data (Training set)
plot_decision_boundary(axes[1, 0], logistic_model_non_linear_separable, X_train_non_linear, y_train_non_linear, 'Non-Linearly Separable Data
```

```
# Plot decision boundary for non-linearly separable data (Testing set)
plot_decision_boundary(axes[1, 1], logistic_model_non_linear_separable, X_test_non_linear, y_test_non_linear, 'Non-Linearly Separable Data (Te

# Adjust layout for better spacing
plt.tight_layout()

# Save the plot as a PNG file
plt.savefig('decision_boundaries.png')

# Display the plots
plt.show()
```