targil2.m

```
% Load Data
train = [train1; train2];
train_labels = [ones(size(train1, 1), 1); zeros(size(train2, 1), 1)];
test = [test1; test2];
test_labels = [ones(size(test1, 1), 1); zeros(size(test2, 1), 1)];
train = double(train);
test = double(test);
figure;
for i = 1:5
    subplot(1, 5, i);
    img = reshape(train(i, :), [28, 28]);
    imshow(img, []);
    title(['Label: ', num2str(train_labels(i))]);
end
%Parameters
[m, n] = size(train);
w = zeros(n, 1);
learning_rate = 0.01;
num_iterations = 1000;
cost_history = zeros(num_iterations, 1);
eps = 1e-6;
for iter = 1:num_iterations
    z = train * w;
    predictions = 1 ./ (1 + exp(-z));
    predictions = max(min(predictions, 1 - eps), eps);
    gradient = (1/m) * (train' * (train_labels - predictions));
    w = w + learning_rate * gradient;
    cost = (1/m) * sum(train_labels .* log(predictions) + ...
                    (1 - train_labels) .* log(1 - predictions));
    cost_history(iter) = cost;
end
```


```
    cost = (1/m) * sum(train_labels .* log(predictions) + ...
                    (1 - train_labels) .* log(1 - predictions));
    cost_history(iter) = cost;
end
figure;
plot(1:num_iterations, cost_history, '-b', 'LineWidth', 2);
grid on;
xlabel('Iteration');
ylabel('Cost Function');
title('Cost Function vs. Iterations');
z_test = test * w;
test_predictions = 1 ./ (1 + exp(-z_test));
test_predictions_binary = test_predictions >= 0.5;
success_rate = mean(test_predictions_binary == test_labels) * 100;
disp(['Success Rate: ', num2str(success_rate), '%']);
```

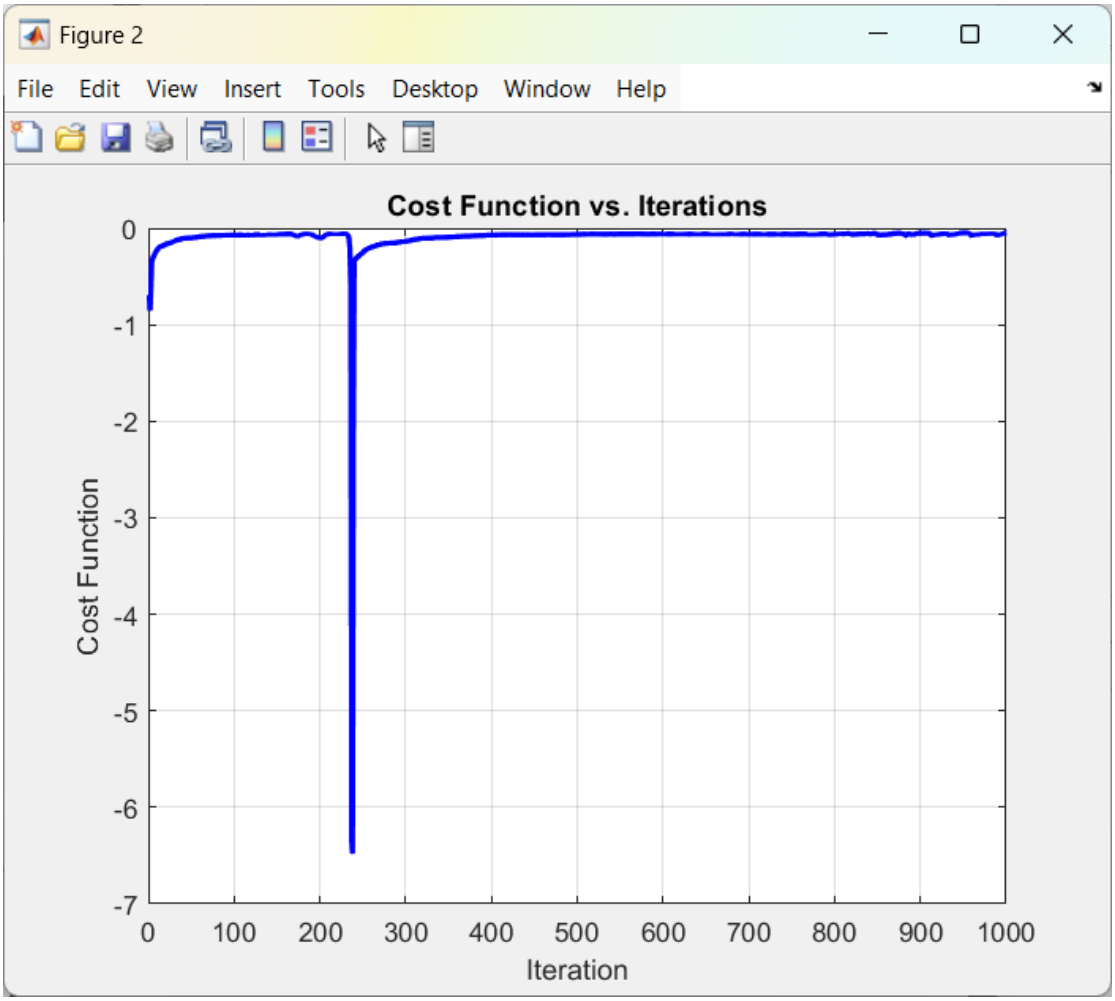
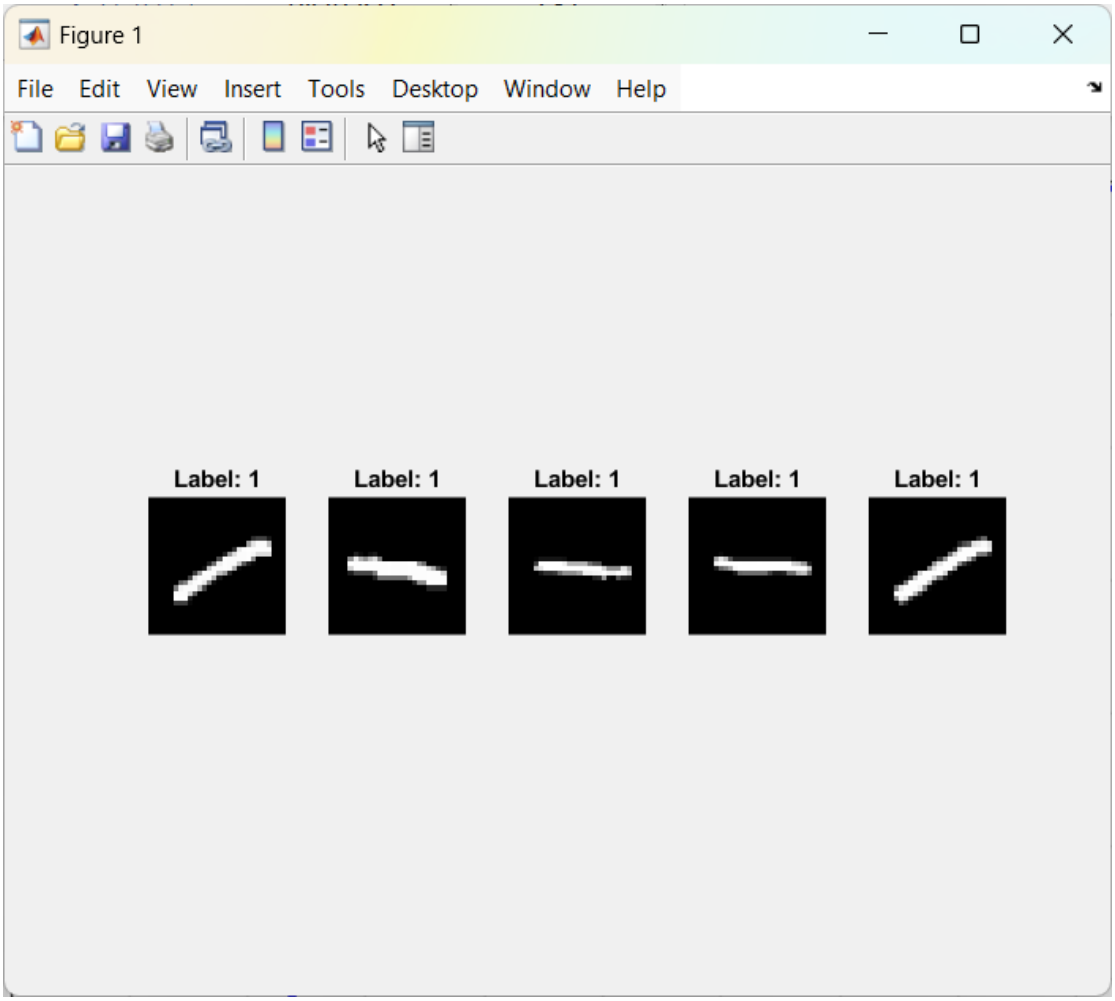


Figure 2
File Edit View Insert Tools Desktop Window Help
Cost Function vs. Iterations
Cost Function
0
-1
-2
-3
-4
-5
-6
-7
0 100 200 300 400 500 600 700 800 900 1000
Iteration



Success Rate: 98.9386%

7) $I(w) = \frac{1}{N} \sum_{t=1}^{N} \log (P(s_t | x_t))$

i)

$$\frac{\partial I(w)}{\partial w} = \frac{1}{N} \sum_{t=1}^{N} (s_t - \sigma(w x_t)) x_t^i$$

$$\frac{\partial I(w)}{\partial w} = \frac{\partial}{\partial w} \frac{1}{N} \sum_{t=1}^{N} \log \left(\sigma(w x_t)^{s_t} \left(1 - \sigma(w x_t)\right)^{1-s_t} \right)$$

$s_t \cdot \ldots \qquad (1-s_t) \cdot \ldots$

$$= \frac{\partial}{\partial w} \frac{1}{N} \sum_{t=1}^{N} s_t \log (\sigma(w x_t)) (1-s) \log (w x_t) = \frac{1}{N} \sum_{t=1}^{N} s_t x_t (1 - \sigma(w x_t)) - (1 - s_t) x_t \sigma(w x_t))$$

$P(s|x) = \sigma(wx)^s (1 - \sigma(wx))^{1-s}$

$$= \frac{1}{N} \sum_{t=1}^{N} (s_t - \sigma(w x_t)) x_t^i$$