Table of Contents

This code evaluates the test set.

```
% ** Important.  This script requires that:
% 1)'centroid_labels' be established in the workspace
% AND
% 2)'centroids' be established in the workspace
% AND
% 3)'test' be established in the workspace

% IMPORTANT!!:
% You should save 1) and 2) in a file named 'classifierdata.mat' as
 part of
% your submission.

predictions = zeros(200,1);
outliers = zeros(200,1);

% loop through the test set, figure out the predicted number
for i = 1:200

testing_vector=test(i,:);

% Extract the centroid that is closest to the test image
[prediction_index,
 vec_distance]=assign_vector_to_centroid(testing_vector,centroids);
predictions(i) = centroid_labels(prediction_index);

end
```

DESIGN AND IMPLEMENT A STRATEGY TO SET THE outliers VECTOR

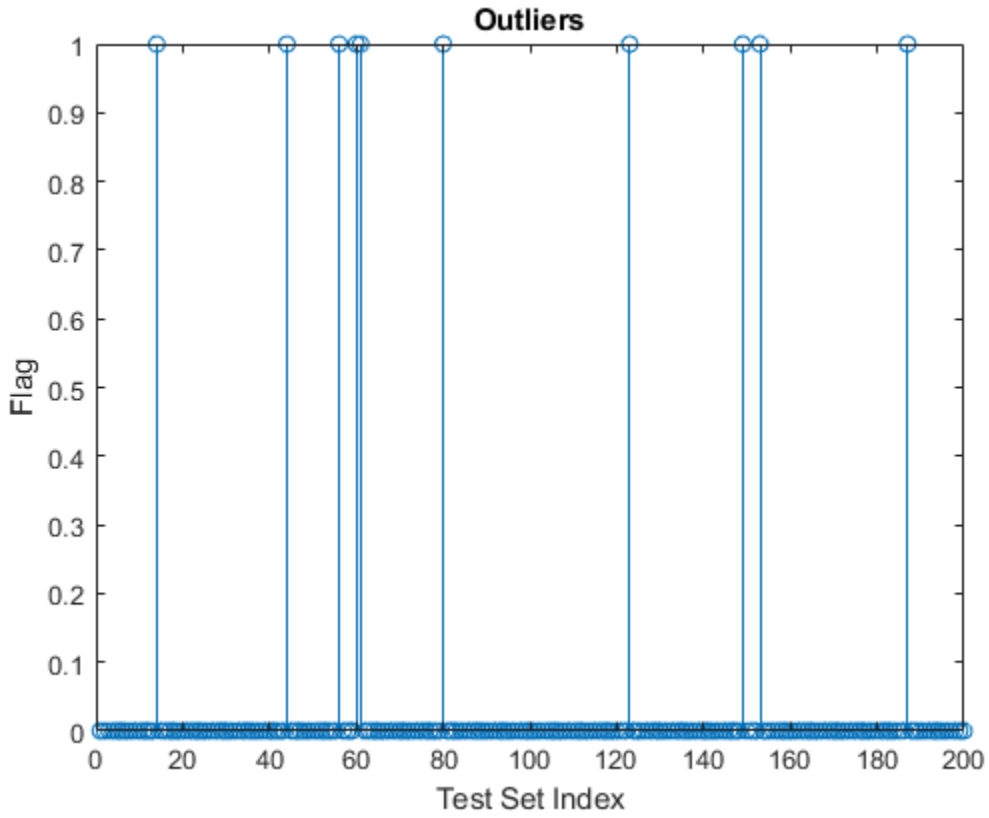
outliers(i) should be set to 1 if the i^th entry is an outlier otherwise, outliers(i) should be 0

```
%create a blank outliers vector
outliers = zeros(1, 200);
%iterate for each image in the test matrix
for i = 1:200
    %assign the i^th vector to a centroid
```

```
    [~, vec_distance] = assign_vector_to_centroid(test(i,
 1:784) ,centroids(:, 1:784));
    %find the distance from the vector to its centroid and put it in
 the
    %outliers vector in its index
    outliers(1, i) = vec_distance;
end
%find the 10 vectors with the largest distance from their centroid
[~, top] = maxk(outliers, 10);
%set the outliers vector to zeros.
outliers = zeros(1, 200);
%make the indexes of the outliers into ones.
outliers(1, top) = 1;
```

MAKE A STEM PLOT OF THE OUTLIER FLAG

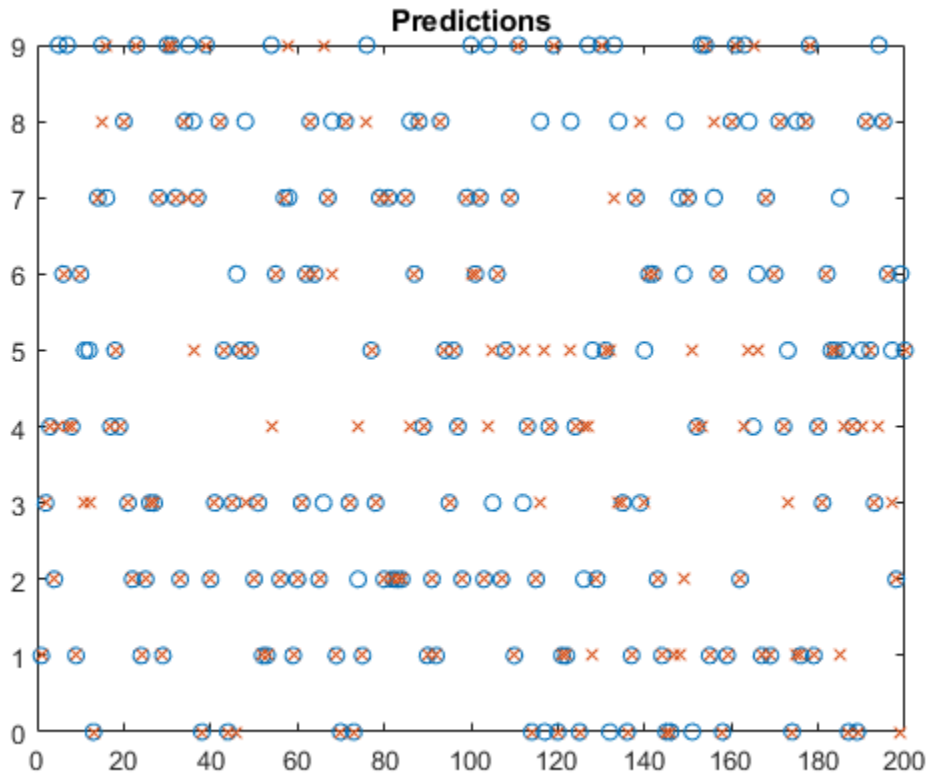
```
%make a stem plot of the outliers, add a title, and add axis labels.
figure;
stem(outliers);
title("Outliers");
xlabel("Test Set Index");
ylabel("Flag");
```



The following plots the correct and incorrect predictions

Make sure you understand how this plot is constructed

```
figure;
plot(correctlabels,'o');
hold on;
plot(predictions,'x');
title('Predictions');
```



The following line provides the number of instances where an entry in correctlabel is

equal to the corresponding entry in prediction However, remember that some of these are outliers

```
sum(correctlabels==predictions)
```

```
ans =

   150
```

Function to pick the Closest Centroid using norm/distance

This function takes two arguments, a vector and a set of centroids It returns the index of the assigned centroid and the distance between the vector and the assigned centroid.

```
function [index, vec_distance] =
 assign_vector_to_centroid(data,centroids)
    %find distance between the data vector and each centroid
    dist = pdist2(data, centroids, "euclidean", "Smallest", 1);
    %find the centroid closest to the data vector and find that
 distance.
    [vec_distance, index] = mink(dist, 1);
end
```

Published with MATLAB® R2021a