ECE133A: Hw1 K-Means Clustering

Thomas Kost | UID: 504989794

Introduction

This PDF will be broken into two portions, a section containing the MATLAB code for the k-means clustering algorithm and a section detailing the results of the clustering. This completes the implemntation for problem 1.10 on k-means clustering.

MATLAB Code

The matlab code implementing the k-means clustering algorithm for the provided data set is shown below:

```matlab
%% Hw 1: K-means clustering
% Class: ECE133A
% Author:  Thomas Kost
% UID: 504989794
% Date: 1/15/2021
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear,clc,close all;

%% Read in data and set params
load mnist_train.mat;
digit_lim = 10000;
digits = digits(:, 1:digit_lim);
k =20;
min_error_term = 10e-5;

% Create initial group assignment
group = randi(k, 1, digit_lim);
reps = zeros(size(digits,1),k);

%% Execute K-means
for i =1:k
    Ind = find(group == i);
    reps(:,i) = mean(digits(:,Ind),2);
end
J_prev = inf;

while(abs(k_means_error(reps,digits,digit_lim,k) - J_prev) > ...
        min_error_term*k_means_error(reps,digits,digit_lim,k))

    J_prev = k_means_error(reps,digits,digit_lim,k);

    % recluster
    for i =1:digit_lim

        diff = digits(:,i)*ones(1,k) - reps(:,:);
        diff_norms = vecnorm(diff);
        diff_norms_sq = diff_norms.^2;
        [~,group(i)] = min(diff_norms_sq, [],2);
```

```matlab
        end

        % regroup
        for i =1:k
            Ind = find(group == i);
            reps(:,i) = mean(digits(:,Ind),2);
        end
    end

% display representatives
fig = figure(1);
for k=1:20
    subplot(4,5,k);
    imshow(reshape(reps(:,k), 28, 28));
end

saveas(fig, "clusters.jpg");
% error function
function error_term = k_means_error(reps, data, length,clusters);
    error_term = 0;
    for i =1:length
        diff = data(:,i)*ones(1,clusters) - reps(:,:);
        diff_norms = vecnorm(diff);
        diff_norms_sq = diff_norms.^2;
        error_term = error_term + min(diff_norms_sq, [],2);
    end
    error_term = error_term/length;
end
```

Results

The resulting clusters are shown in the image in Figure 1. We can see that each image is fairly distinct and indicates that the data can be clustered well (that is that there is a general comminality between the sets of images that they can be divided into this way). This also suggests that our clustering has been done properly as the images seem to be representative of the handwritten data in general (we can clearly see the digits written).