COGS 109: Assignment #4

Due on Sunday, November 8, 2015

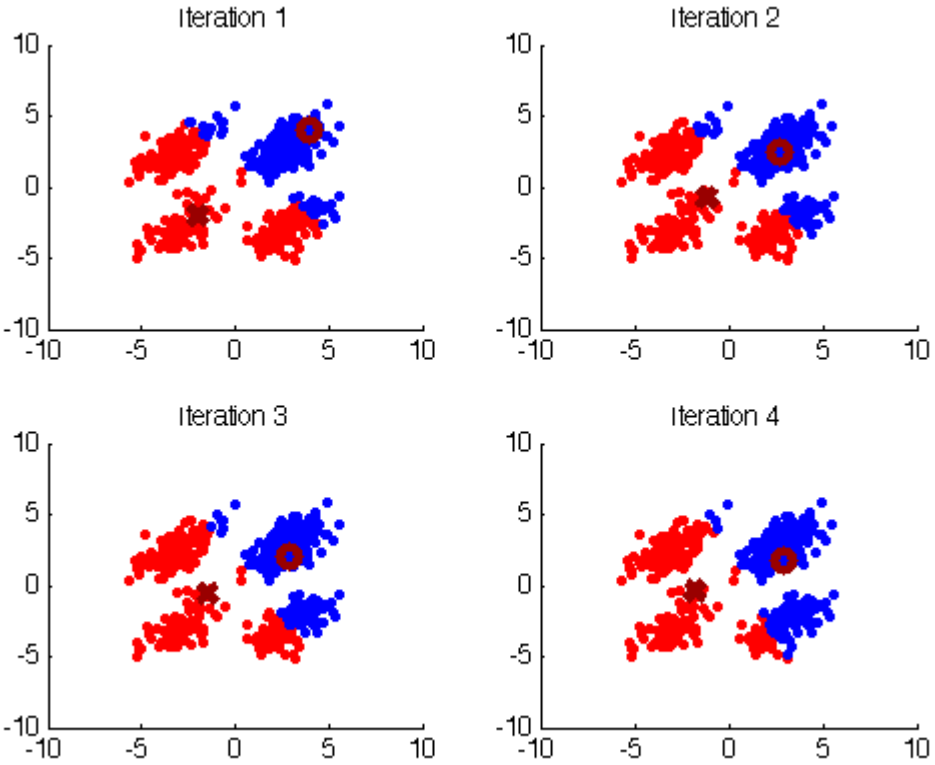
Tu, Zhuowen 2pm

Kyle Lee
A01614951

Problem 1

Problem: Step-by-step K-means. Let $k = 2$ and initialize the two means (cluster centers) at $[-2, -2]$ and $[4, 4]$ respectively.

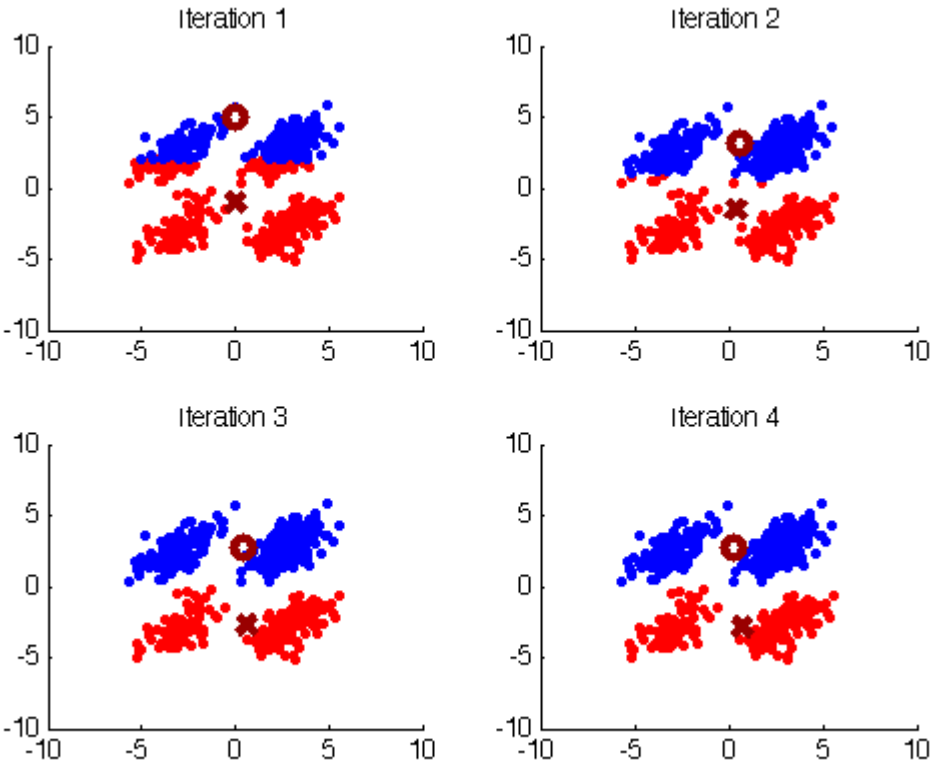
Solution:



Problem 2

Problem: Create another figure of 2 by 2 subplots under the same procedure as described in question 1. However, this time, use the initial centers at $[0, -1]$ and $[0, 5]$. Attach your figure here and describe your observations.

Solution:

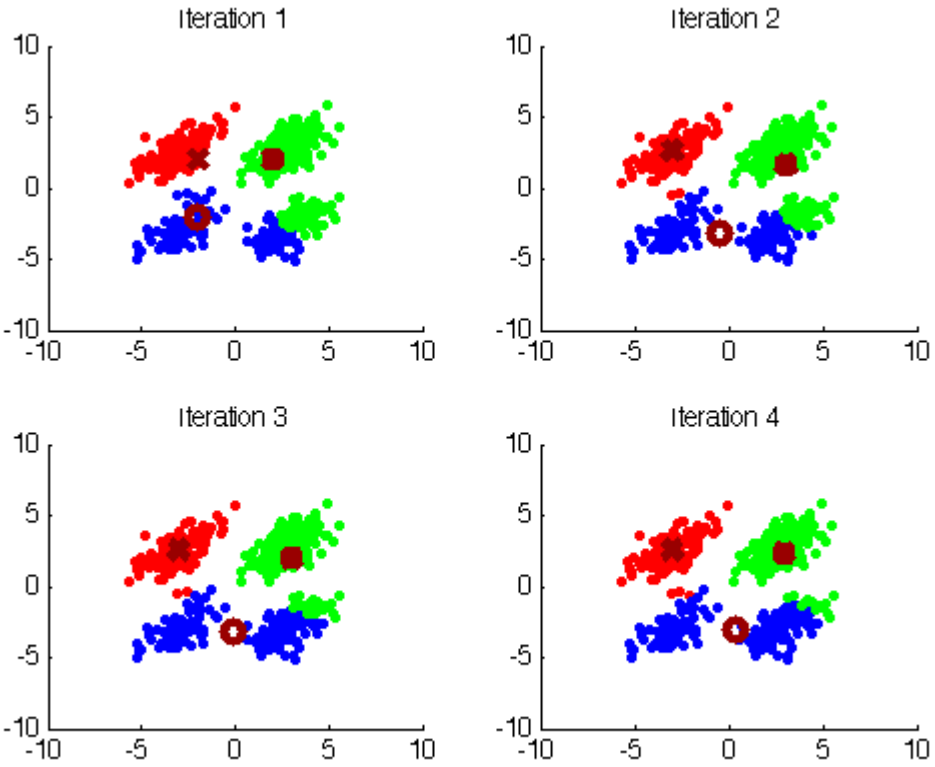


We observe that the initial centers greatly affect the clustering, even after each iteration. The clusters in one sense is perpendicular to the line containing the means which makes sense due to the fact that each point can only be classified as either Cluster 1 or Cluster 2.

Problem 3

Problem: Create another figure of 2 by 2 subplots under the same procedure as described in question 1. However, this time, use $k = 3$ and initialize the cluster centers at $[2, -2]$,$[-2, -2]$,and$[2, 2]$. Attach your figure here and describe your observations.

Solution:

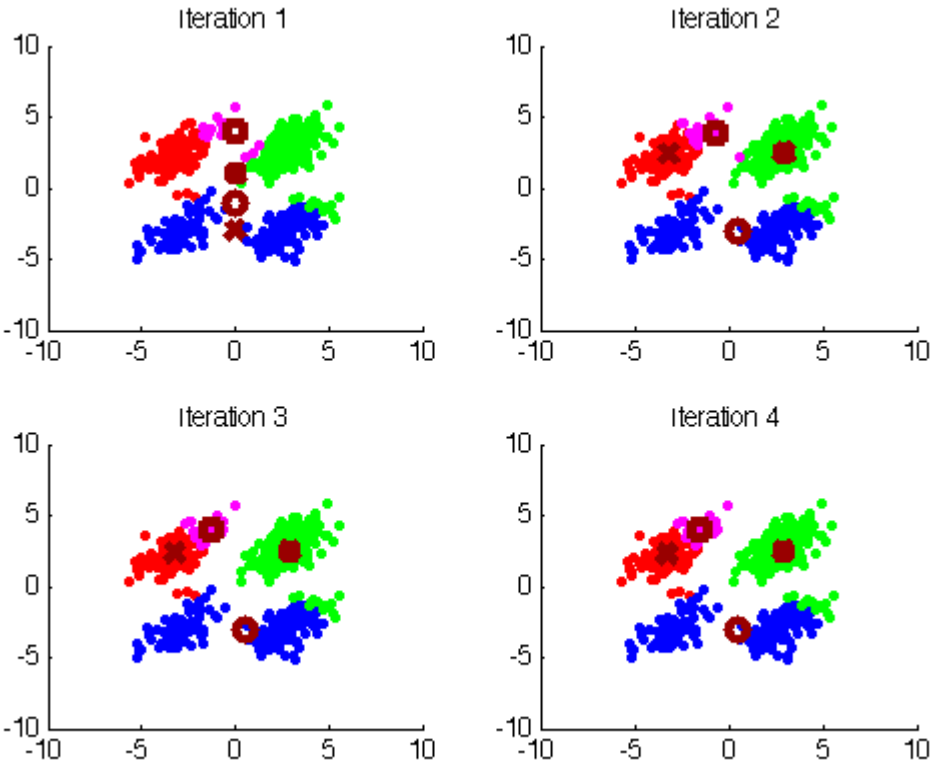


We observe that the initial centers greatly affect the clustering, even after each iteration. The bottom mean dominates the lower half of the data since the 2 top means seem to compete for the same territory which makes sense due to the proximity of the means with the clusters.

Problem 4

Problem: Create another figure of 2 by 2 subplots under the same procedure as described in question 1. However, this time, use $k = 4$ and initialize the cluster centers at $[0, -3]$, $[0, -1]$,$[0, 1]$, and $[0, 4]$. Attach your figure here and describe your observations.

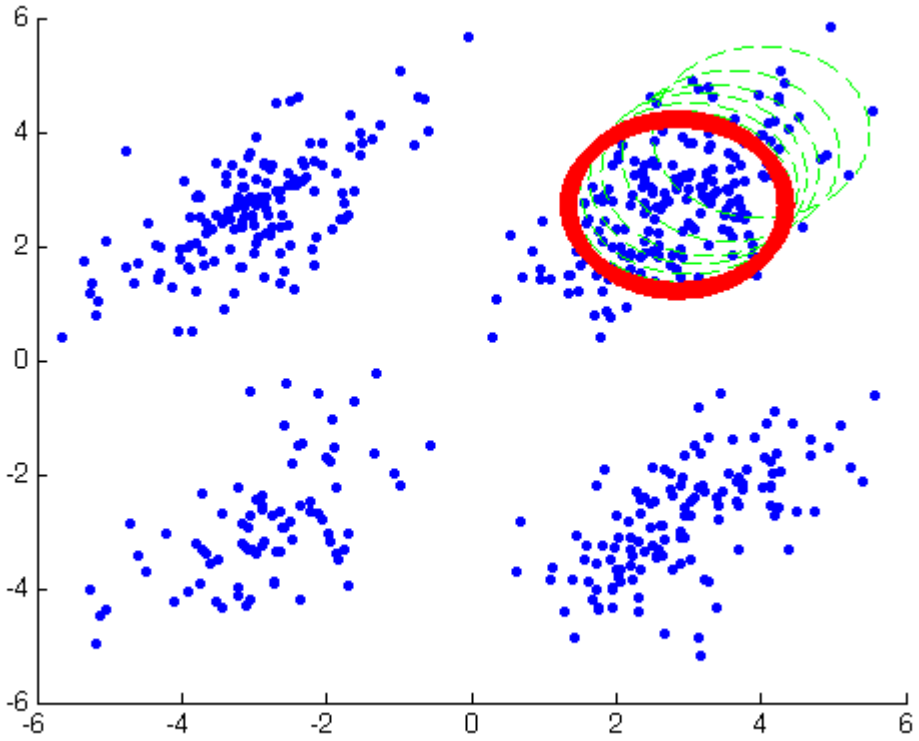
Solution:



We observe that the initial centers greatly affect the clustering, even after each iteration. Note that the initial cluster means are perpendicular to each other, but as we recalibrate the mean clusters, they sporadically move due to the mean's proximity with the surrounding means. Note in the 4th iteration, we see that the magenta region shares in the same proximity as the red region, which is due to the previous iterations where 3 of the means are moving upwards while there is an obvious divide between the blue points and the upper half points. If we ran a few more iterations, we can expect that the green points in the bottom half will be clustered to the blue mean.

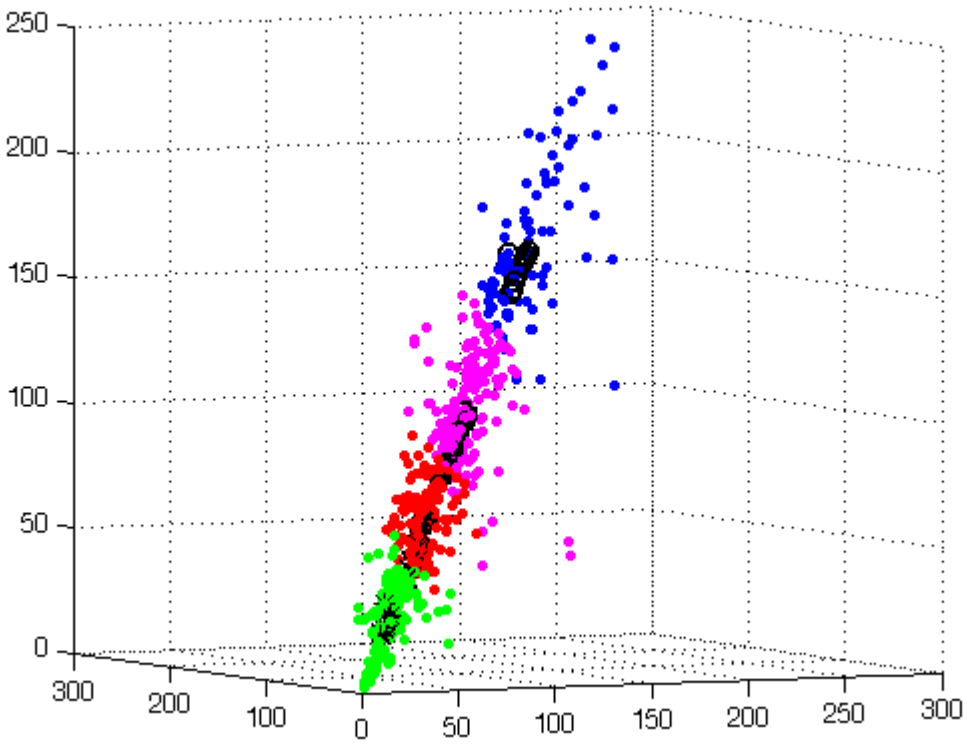
Problem 5

Problem: Implement mean-shift algorithm and let $k = 1$ for simplicity. Use a uniformkernel of radius 1.5.
Solution:



Problem 6

Extra Credit
Problem: Combine the PCA and K-Means
Solution:




```
%%%% Homework #4 %%%%

%% QUESTION 1 %%
% PART A %
% Initialize means
centers = [ -2, -2; 4, 4];
% PART B %
% Calculate distance between data points
figure
for i = 1:4
    subplot(2,2,i);

    % Find distance between kmeandata and the centers
    for j=1:2
        dist(:,j) = sqrt((kmeandata(:,1) - centers(j,1)).^2 ...
            +(kmeandata(:,2) - centers(j,2)).^2);
    end

    % PART C %
    % Assign each data point with the id of its nearest cluster
```

```
    [v g_ind] = min(dist,[],2);

    % PART D %
    % Cluster data and show centers
    % Cluster 1
    scatter(kmeandata(g_ind == 1,1), kmeandata(g_ind == 1,2),20,'r', ...
        'filled');
    hold on
    scatter(centers(1,1),centers(1,2),80,'Marker','x','MarkerEdgeColor',...
        [0.6 0 0], 'LineWidth',4);
    % Cluster 2
    scatter(kmeandata(g_ind == 2,1), kmeandata(g_ind == 2,2),20,'b',...
        'filled');
    hold on
    scatter(centers(2,1),centers(2,2),80,'Marker','o','MarkerEdgeColor',...
        [0.6 0 0], 'LineWidth',4);

    % PART E %
    % Recalibrate centers
    centers(1,:) = mean(kmeandata(g_ind ==1,:));
    centers(2,:) = mean(kmeandata(g_ind ==2,:));

    % Label iteration
    title(sprintf('Iteration %d', i))
end
% PART F - repeat 4 times in loop %


%% QUESTION 2 %%
% Initialize new centers
centers2 = [ 0, -1; 0, 5];
% Start new figure
figure
for i = 1:4
    subplot(2,2,i);
    % Find distance between kmeandata and centers
    for j=1:2
        dist2(:,j) = sqrt((kmeandata(:,1) - centers2(j,1)).^2 ...
            +(kmeandata(:,2) - centers2(j,2)).^2);
    end

    % Assign each data point with the id of its nearest cluster
    [v2 g_ind2] = min(dist2,[],2);

    % Cluster data and show centers
    % Cluster 1
    scatter(kmeandata(g_ind2 == 1,1), kmeandata(g_ind2 == 1,2),20,'r',...
        'filled');
    hold on
    scatter(centers2(1,1),centers2(1,2),80,'Marker','x',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 2
    scatter(kmeandata(g_ind2 == 2,1), kmeandata(g_ind2 == 2,2),20,'b',...
```

```
        'filled');
    hold on
    scatter(centers2(2,1),centers2(2,2),80,'Marker','o',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);

    % Recalibrate centers
    centers2(1,:) = mean(kmeandata(g_ind2 ==1,:));
    centers2(2,:) = mean(kmeandata(g_ind2 ==2,:));

    % Label iteration
    title(sprintf('Iteration %d', i))
end


%% QUESTION 3 %%
%Initialize 3 means
centers3 = [ -2, 2; -2, -2; 2,2];

figure
for i = 1:4
    subplot(2,2,i);

    % Calculate distance between data points and centers
    for j=1:3
        dist3(:,j) = sqrt((kmeandata(:,1) - centers3(j,1)).^2 ...
            +(kmeandata(:,2) - centers3(j,2)).^2);
    end

    % Assign each data point with the id of its nearest cluster
    [v3 g_ind3] = min(dist3,[],2);

    % Cluster data and show centers
    % Cluster 1
    scatter(kmeandata(g_ind3 == 1,1), kmeandata(g_ind3 == 1,2),20,...
        'r', 'filled');
    hold on
    scatter(centers3(1,1),centers3(1,2),80,'Marker','x',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 2
    scatter(kmeandata(g_ind3 == 2,1), kmeandata(g_ind3 == 2,2),20,...
        'b', 'filled');
    hold on
    scatter(centers3(2,1),centers3(2,2),80,'Marker','o',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 3
    scatter(kmeandata(g_ind3 == 3,1), kmeandata(g_ind3 == 3,2),20,'g',...
        'filled');
    hold on
    scatter(centers3(3,1),centers3(3,2),80,'Marker','*',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);

    % Recalibrate centers
    centers3(1,:) = mean(kmeandata(g_ind3 ==1,:));
```

```
    centers3(2,:) = mean(kmeandata(g_ind3 ==2,:));
    centers3(3,:) = mean(kmeandata(g_ind3 ==3,:));

    % Label iteration
    title(sprintf('Iteration %d', i));
end

%% QUESTION 4 %%
%Initialize 4 means
centers4 = [ 0, -3; 0, -1; 0, 1; 0, 4];

figure
for i = 1:4
    subplot(2,2,i);
    % Calculate distance between data points
    for j=1:4
        dist3(:,4) = sqrt((kmeandata(:,1) - centers4(j,1)).^2 ...
            +(kmeandata(:,2) - centers4(j,2)).^2);
    end

    % Assign each data point with the id of its nearest cluster
    [v4 g_ind4] = min(dist3,[],2);

    % Cluster data and show centers
    % Cluster 1
    scatter(kmeandata(g_ind4 == 1,1), kmeandata(g_ind4 == 1,2),20,'r',...
        'filled');
    hold on
    scatter(centers4(1,1),centers4(1,2),80,'Marker','x',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 2
    scatter(kmeandata(g_ind4 == 2,1), kmeandata(g_ind4 == 2,2),20,'b',...
        'filled');
    hold on
    scatter(centers4(2,1),centers4(2,2),80,'Marker','o',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 3
    scatter(kmeandata(g_ind4 == 3,1), kmeandata(g_ind4 == 3,2),20,'g',...
        'filled');
    hold on
    scatter(centers4(3,1),centers4(3,2),80,'Marker','*',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);
    % Cluster 4
    scatter(kmeandata(g_ind4 == 4,1), kmeandata(g_ind4 == 4,2),20,'m',...
        'filled');
    hold on
    scatter(centers4(4,1),centers4(4,2),80,'Marker','square',...
        'MarkerEdgeColor',[0.6 0 0], 'LineWidth',4);

    % Recalibrate centers
    centers4(1,:) = mean(kmeandata(g_ind4 ==1,:));
    centers4(2,:) = mean(kmeandata(g_ind4 ==2,:));
    centers4(3,:) = mean(kmeandata(g_ind4 ==3,:));
```

```
    centers4(4,:) = mean(kmeandata(g_ind4 ==4,:));

    %Label iteration
    title(sprintf('Iteration %d', i))
end

%% QUESTION 5 %%
figure
% PART A
% initialize mean and radius
center5 = [4,4];
radius = 1.5;

% PART B
% plot all points and hold on
scatter(kmeandata(:,1),kmeandata(:,2),20,'b','filled');
hold on;

% PART C
% plot disk of radius 1.5 centered at current mean
plot(center5(:,1) + radius*cos([0:0.01:2*pi]),center5(:,2)...
    + radius*sin([0:0.01:2*pi]),'--','Color',[0,1,0],'LineWidth',1);

% PART D
% update mean using Mean Shift Alg.
dist5(:,1) = sqrt((kmeandata(:,1) - center5(1,1)).^2 ...
    + (kmeandata(:,2) - center5(1,2)).^2);
ind = dist5 <= radius;
center5 = mean(kmeandata(ind,1:2));

% PART E
for i=1:9
    % plot disk of radius 1.5 centered at current mean
    plot(center5(:,1) + radius*cos([0:0.01:2*pi]),center5(:,2)...
        + radius*sin([0:0.01:2*pi]),'--','Color',[0,1,0],'LineWidth',1);

    % update mean using Mean Shift Alg.
    dist5(:,1) = sqrt((kmeandata(:,1) - center5(1,1)).^2 ...
        + (kmeandata(:,2) - center5(1,2)).^2);
    ind = dist5 <= radius;
    center5 = mean(kmeandata(ind,1:2));
end
% PART F
 plot(center5(:,1) + radius*cos([0:0.01:2*pi]),center5(:,2)...
     + radius*sin([0:0.01:2*pi]),'*','Color',[1,0,0],'LineWidth',1);


 %% Extra Credit %%
 % PART A %
 face_redu = facemat(1:3,1:500)';
 % PART B %
 % calculate 4 random faces
 index = randperm(size(face_redu,1),4);
```

```
% take 4 random faces and store into center6
center6 = face_redu(index,:)
figure
for i = 1:10
    % Calculate distance
   for j=1:4
       dist6(:,j) = sqrt((face_redu(:,1) - center6(j,1)).^2 ...
           +(face_redu(:,2) - center6(j,2)).^2 + (face_redu(:,3) ...
           - center6(j,3)).^2);
   end

   % assign each data point with the id of its nearest cluster
   [v6 g_ind6] = min(dist6,[],2);

   % Cluster 1
   scatter3(face_redu(g_ind6 == 1,1), face_redu(g_ind6 == 1,2),...
       face_redu(g_ind6 == 1,3),20,'r', 'filled');
   hold on
   scatter3(center6(1,1),center6(1,2),center6(1,3),80,'Marker','x',...
       'MarkerEdgeColor',[0 0 0], 'LineWidth',2);
   hold on
   % Cluster 2
   scatter3(face_redu(g_ind6 == 2,1), face_redu(g_ind6 == 2,2),...
       face_redu(g_ind6 == 2,3),20,'b', 'filled');
   hold on
   scatter3(center6(2,1),center6(2,2),center6(2,3),80,'Marker','o',...
       'MarkerEdgeColor',[0 0 0], 'LineWidth',2);
   hold on
   % Cluster 3
   scatter3(face_redu(g_ind6 == 3,1), face_redu(g_ind6 == 3,2),...
       face_redu(g_ind6 == 3,3),20,'g', 'filled');
   hold on
   scatter3(center6(3,1),center6(3,2),center6(3,3),80,'Marker','*',...
       'MarkerEdgeColor',[0 0 0], 'LineWidth',1);
   hold on
   % Cluster 4
   scatter3(face_redu(g_ind6 == 4,1), face_redu(g_ind6 == 4,2),...
       face_redu(g_ind6 == 4,3),20,'m', 'filled');
   hold on
   scatter3(center6(4,1),center6(4,2),center6(4,3),80,'Marker','+',...
       'MarkerEdgeColor',[0 0 0], 'LineWidth',4);

   % Calcuate new centers
   center6(1,:) = mean(face_redu(g_ind6 ==1,:));
   center6(2,:) = mean(face_redu(g_ind6 ==2,:));
   center6(3,:) = mean(face_redu(g_ind6 ==3,:));
   center6(4,:) = mean(face_redu(g_ind6 ==4,:));
   center6
end

% PART C %
subplot(1,1,1)
```