Silvia Ionescu
12/3/16

Assignment 8

Problem 8.1a: k-means vs. Spectral Clustering

For this part sample_circle.m and sample_spiral.m were used with k = 3 and 500 points for each cluster. The Matlab build in function kmeans was used to cluster D_1 and D_2, where 'Replicates' was set to 20 and 'Distance' to 'sqeuclidean'. K-means clustering was determined using k = 2, 3, 4.

i) Plots for D1 and D2 along with cluster assignment for k = 2, 3, 4 and cluster centers are shown below:

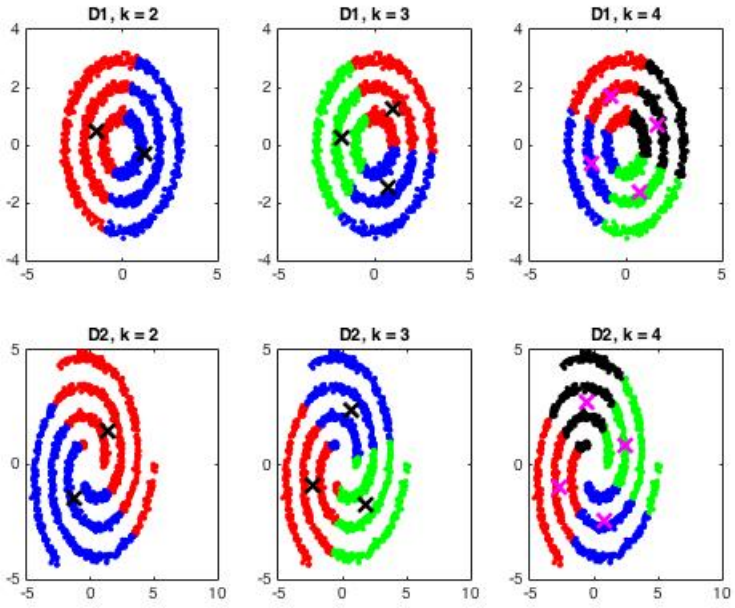


Fig 1: Plot of all data point in D1 and D2 along with cluster assignments and cluster centers.

ii) The within cluster sums of points to cluster center l_2 squared distances are shown below:

D_1 data points			
	k = 2	k =3	k = 4
Cluster 1	2165.7380	912.3421	473.3028
Cluster 2	2295.7806	939.2965	594.1163
Cluster 3		979.8277	464.6794
Cluster 4			564.7122
D_2 data points			

Silvia Ionescu
12/3/16

	k = 2	k =3	k = 4
Cluster 1	4730.2157	2089.0466	1076.1931
Cluster 2	5071.1917	1896.6135	1149.3859
Cluster 3		1765.5718	1262.9941
Cluster 4			925.2390

Problem 8.1b: Spectral Clustering

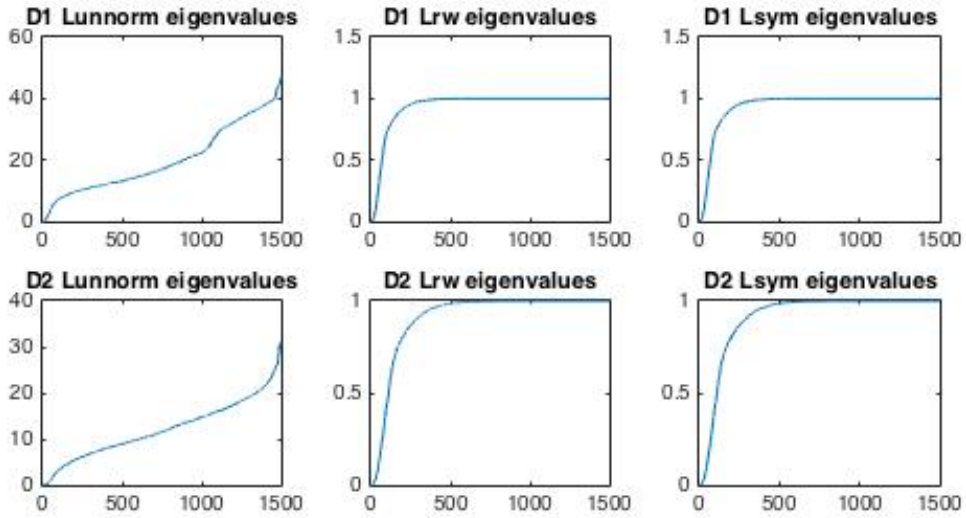
Implemented three spectral clustering algorithms. The Gaussian similarity, S(i , j) was calculated according to the homework description where σ = 0.2. Used a fully connected graph, therefore the adjacency matrix W = S.

The graph three Laplacian(L, L_{rw}, and L_{sym}) were calculated according to the formulas below:

$$L = D - W$$
$$L_{rw} = D^{-1}L$$
$$L_{sym} = D^{-1/2}LD^{-1/2}$$

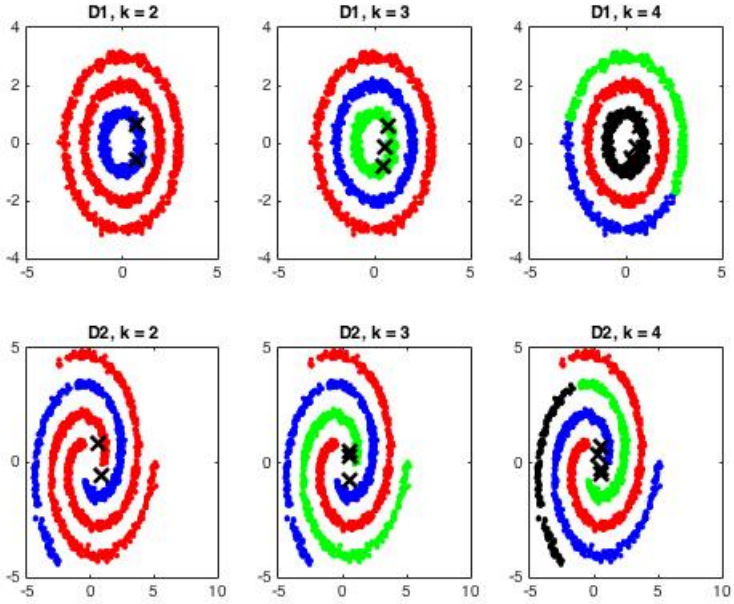
Where D is the degree matrix.

i) A plot of the eigenvalues of L, L_{rw}, and L_{sym} for D1 and D2 in ascending order is shown below:

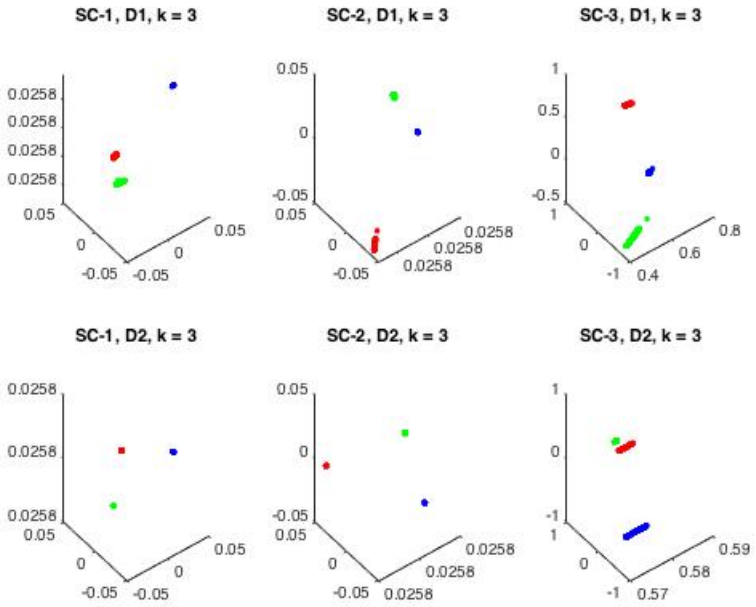


ii) A plot of all the data points in D1 and D2 with K= 2, 3, 4 for SC-3 indicating cluster assignments is shown below:

Silvia Ionescu
12/3/16



iii) A plot of the V matrices in SC-1, SC-2, and SC-3 for k = 3 and both D1 and D2 is shown below:

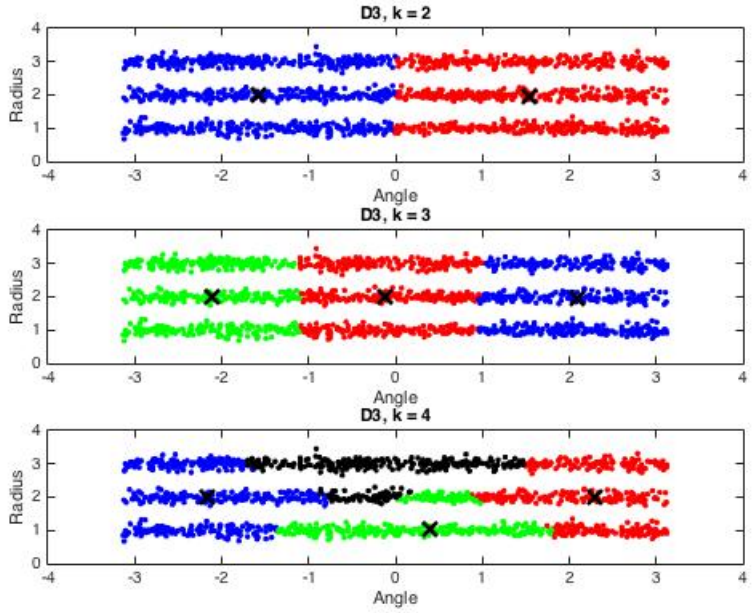


Problem 8.1c

Silvia Ionescu
12/3/16

Transformed the Cartesian coordinates representation of the data points in D1 into polar coordinates. Applied the k-means algorithm to D3 using 'Replicates' of 20 and 'Distance' to 'cityblock' for k = 2, 3, 4.

i) A plot of all the data points in D3 for K = 2, 3, 4 is shown below:



ii) The overall within-cluster sums of points to cluster centroid (Euclidian) squared distances are shown below for D3:

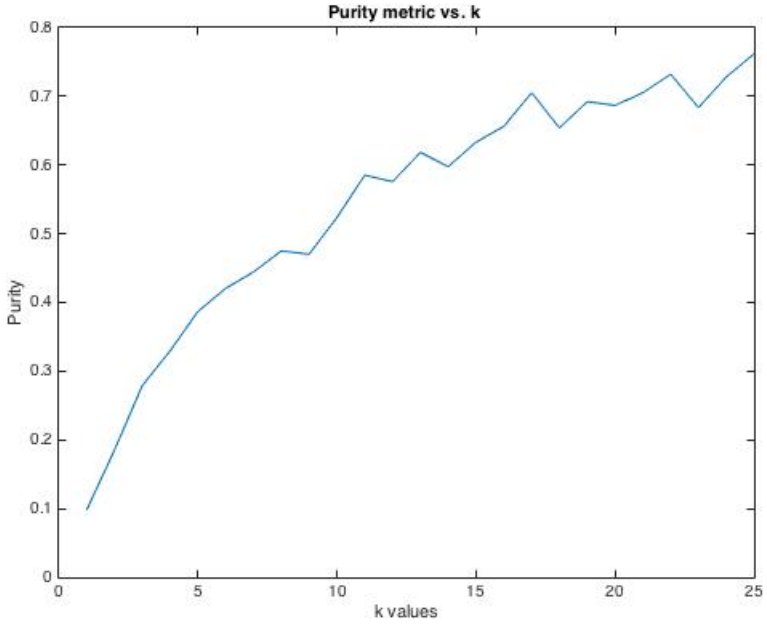
D_3 data points			
	k = 2	k =3	k = 4
Cluster 1	122.4183	43.9835	27.4472
Cluster 2	43.75309	32.1188	7.74382
Cluster 3		26.8592	24.7113
Cluster 4			9.21778

Problem 8.2: Spectral Clustering on Airbnb data

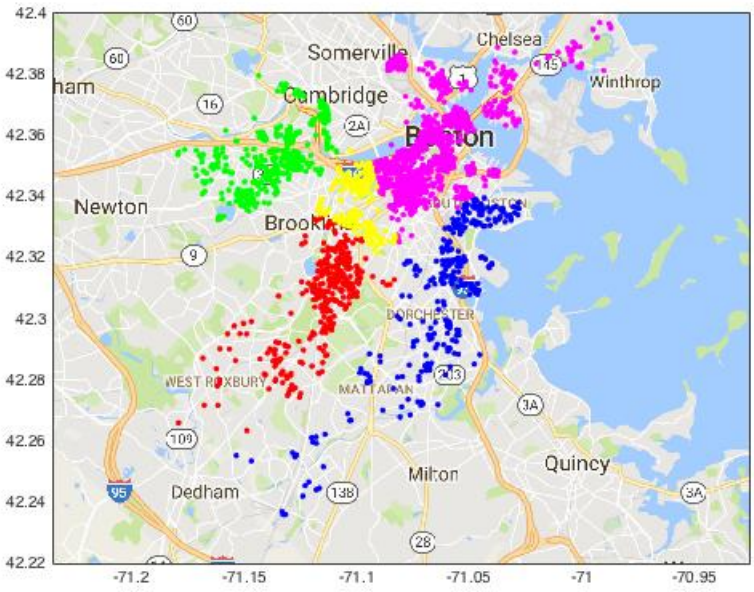
The longitude, latitude, and neighborhood were imported from the 'BostonListing.mat' file.

8.2a) The Gaussian similarity distance was used and a full graph was constructed. The σ = 0.01 and the symmetrical normalized Laplacian, L_{sym} , was used for spectral clustering. For k = 1, 2, … ,25 the purity metric was calculated and is shown below:

Silvia Ionescu
12/3/16



8.2b) Using the plot_google.map all data points were plotted and k = 5 different colors are show below:




```
% Problem 8_1a

clear; clc; close all;

input = load('BostonListing.mat');
latitude = input.latitude;
```

```
longitude = input.longitude;
neighbourhood = input.neighbourhood;

k = 3;
points_per_cluster = [500,500,500];
[D1, D1_label] = sample_circle(k, points_per_cluster );
[D2, D2_label] = sample_spiral(k, points_per_cluster );

% D = pdist2(D1(1,:),D2(1,:), 'euclidean');

rng(2);
% D1, k = 2
figure(1)
subplot(2,3,1)
[idx, C, sumd_D1_k2] = kmeans(D1, 2, 'Replicates', 20, 'Distance',
'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'br','xo')

plot(D1(idx==1,1),D1(idx==1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx==2,1),D1(idx==2,2),'b.','MarkerSize',10)
plot(C(:,1),C(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D1, k = 2')

subplot(2,3,2)
[idx_D1_k3, C_D1_k3, sumd_D1_k3] = kmeans(D1, 3, 'Replicates', 20,
'Distance', 'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'brg','xo+')
plot(D1(idx_D1_k3==1,1),D1(idx_D1_k3==1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx_D1_k3==2,1),D1(idx_D1_k3==2,2),'b.','MarkerSize',10)
hold on
plot(D1(idx_D1_k3==3,1),D1(idx_D1_k3==3,2),'g.','MarkerSize',10)
plot(C_D1_k3(:,1),C_D1_k3(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D1, k = 3')

subplot(2,3,3)
[idx_D1_k4, C_D1_k4, sumd_D1_k4] = kmeans(D1, 4, 'Replicates', 20,
'Distance', 'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'brgk','xo+*')
plot(D1(idx_D1_k4 == 1,1), D1(idx_D1_k4 == 1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx_D1_k4 == 2,1), D1(idx_D1_k4 == 2,2),'b.','MarkerSize',10)
hold on
plot(D1(idx_D1_k4 == 3,1),D1(idx_D1_k4 == 3,2),'g.','MarkerSize',10)
hold on
plot(D1(idx_D1_k4 == 4,1),D1(idx_D1_k4 == 4,2),'k.','MarkerSize',10)
plot(C_D1_k4(:,1),C_D1_k4(:,2),'mx',...
     'MarkerSize',12,'LineWidth',3)
title('D1, k = 4')


subplot(2,3,4)
[idx_D2_k2, C_D2_k2, sumd_D2_k2] = kmeans(D2, 2, 'Replicates', 20,
```

```
'Distance', 'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'br','xo')
plot(D2(idx_D2_k2==1,1),D2(idx_D2_k2==1,2),'r.','MarkerSize',10)
hold on
plot(D2(idx_D2_k2==2,1),D2(idx_D2_k2==2,2),'b.','MarkerSize',10)
plot(C_D2_k2(:,1),C_D2_k2(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D2, k = 2')

subplot(2,3,5)
[idx_D2_k3, C_D2_k3, sumd_D2_k3] = kmeans(D2, 3, 'Replicates', 20,
'Distance', 'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'brg','xo+')
plot(D2(idx_D2_k3==1,1),D2(idx_D2_k3==1,2),'r.','MarkerSize',10)
hold on
plot(D2(idx_D2_k3==2,1),D2(idx_D2_k3==2,2),'b.','MarkerSize',10)
hold on
plot(D2(idx_D2_k3==3,1),D2(idx_D2_k3==3,2),'g.','MarkerSize',10)
plot(C_D2_k3(:,1),C_D2_k3(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D2, k = 3')

subplot(2,3,6)
[idx_D2_k4, C_D2_k4, sumd_D2_k4] = kmeans(D2, 4, 'Replicates', 20,
'Distance', 'sqeuclidean');
% gscatter(D1(:,1),D1(:,2), idx, 'brgk','xo+*')

plot(D2(idx_D2_k4==1,1),D2(idx_D2_k4==1,2),'r.','MarkerSize',10)
hold on
plot(D2(idx_D2_k4==2,1),D2(idx_D2_k4==2,2),'b.','MarkerSize',10)
hold on
plot(D2(idx_D2_k4==3,1),D2(idx_D2_k4==3,2),'g.','MarkerSize',10)
hold on
plot(D2(idx_D2_k4==4,1),D2(idx_D2_k4==4,2),'k.','MarkerSize',10)
plot(C_D2_k4(:,1),C_D2_k4(:,2),'mx',...
     'MarkerSize',12,'LineWidth',3)
title('D2, k = 4')



%
% % part 8.1a part ii
%
% test = D1(idx==1,:);
% C_test = C(1, :);
% distance = pdist2(test,C_test, 'euclidean');

% part 8.1b
sigma = 0.2;
D1_distance = pdist2(D1, D1, 'euclidean');
D1_similarity_matrix = exp(-((D1_distance.^2)/(2*sigma^2)));
D1_W = D1_similarity_matrix;
D1_degree_matrix = sum(D1_W, 2);
D1_degree_matrix = diag(D1_degree_matrix);

% D1 -calculate Laplacians
D1_L_unnorm = D1_degree_matrix - D1_W;
```

```
D1_L_rw = (D1_degree_matrix^-1) * D1_L_unnorm;
D1_L_sym = (D1_degree_matrix^-(1/2)) * D1_L_unnorm * (D1_degree_matrix^-
(1/2));

% D1 - calculate eigenvalues

D1_L_unnorm_eigval = sort(eig(D1_L_unnorm), 'ascend');
D1_L_rw_eigval = sort(eig(D1_L_rw), 'ascend');
D1_L_sym_eigval = sort(eig(D1_L_sym), 'ascend');

% D2
D2_distance = pdist2(D2, D2, 'euclidean');
D2_similarity_matrix = exp(-((D2_distance.^2)/(2*sigma^2)));
D2_W = D2_similarity_matrix;
D2_degree_matrix = sum(D2_W, 2);
D2_degree_matrix = diag(D2_degree_matrix);

% D1 -calculate Laplacians
D2_L_unnorm = D2_degree_matrix - D2_W;
D2_L_rw = (D2_degree_matrix^-1) * D2_L_unnorm;
D2_L_sym = (D2_degree_matrix^-(1/2)) * D2_L_unnorm * (D2_degree_matrix^-
(1/2));

% D1 - calculate eigenvalues

D2_L_unnorm_eigval = sort(eig(D2_L_unnorm), 'ascend');
D2_L_rw_eigval = sort(eig(D2_L_rw), 'ascend');
D2_L_sym_eigval = sort(eig(D2_L_sym), 'ascend');


% 8.1bi plot
figure(2)
subplot(3,3,1);
plot(D1_L_unnorm_eigval);
title('D1 Lunnorm eigenvalues');
subplot(3,3,2);
plot(D1_L_rw_eigval);
title('D1 Lrw eigenvalues');
subplot(3,3,3);
plot(D1_L_sym_eigval);
title('D1 Lsym eigenvalues');
subplot(3,3,4);
plot(D2_L_unnorm_eigval);
title('D2 Lunnorm eigenvalues');
subplot(3,3,5);
plot(D2_L_rw_eigval);
title('D2 Lrw eigenvalues');
subplot(3,3,6);
plot(D2_L_sym_eigval);
title('D2 Lsym eigenvalues');


% find the smallest eigenvectors for D1_L_sym
[D1_L_sym_eigvect_k2 val1] =  eigs(D1_L_sym, 2,'sm');
[D1_L_sym_eigvect_k3 val2] =  eigs(D1_L_sym, 3,'sm');
[D1_L_sym_eigvect_k4 val3] =  eigs(D1_L_sym, 4,'sm');
```

```
% V matrix of eigenvectors

% find the smallest eigenvectors for D2_L_sym
[D2_L_sym_eigvect_k2 val4] =  eigs(D2_L_sym, 2,'sm');
[D2_L_sym_eigvect_k3 val5] =  eigs(D2_L_sym, 3,'sm');
[D2_L_sym_eigvect_k4 val6] =  eigs(D2_L_sym, 4,'sm');

D1_norm_k2 = sqrt(sum(D1_L_sym_eigvect_k2.^2,2));
D1_norm_k3 = sqrt(sum(D1_L_sym_eigvect_k3.^2,2));
D1_norm_k4 = sqrt(sum(D1_L_sym_eigvect_k4.^2,2));

D2_norm_k2 = sqrt(sum(D2_L_sym_eigvect_k2.^2,2));
D2_norm_k3 = sqrt(sum(D2_L_sym_eigvect_k3.^2,2));
D2_norm_k4 = sqrt(sum(D2_L_sym_eigvect_k4.^2,2));

% normalize V
for i = 1:length(D2_L_sym_eigvect_k2)

    D1_Lsym_k2_norm(i,:) = D1_L_sym_eigvect_k2(i,:)/ D1_norm_k2(i);
    D1_Lsym_k3_norm(i,:) = D1_L_sym_eigvect_k3(i,:)/ D1_norm_k3(i);
    D1_Lsym_k4_norm(i,:) = D1_L_sym_eigvect_k4(i,:)/ D1_norm_k4(i);

    D2_L_sym_k2_norm(i,:) = D2_L_sym_eigvect_k2(i,:)/ D2_norm_k2(i);
    D2_L_sym_k3_norm(i,:) = D2_L_sym_eigvect_k3(i,:)/ D2_norm_k3(i);
    D2_L_sym_k4_norm(i,:) = D2_L_sym_eigvect_k4(i,:)/ D2_norm_k4(i);

end

% [idx,C] = kmeans(D1_Lsym_k2_norm,2);
%
% figure;
% plot(D1(idx==1,1),D1(idx==1,2),'r.','MarkerSize',12)
% hold on
% plot(D1(idx==2,1),D1(idx==2,2),'b.','MarkerSize',12)
% plot(C(:,1),C(:,2),'kx',...
%      'MarkerSize',15,'LineWidth',3)
```

```
% Problem 8_1b

clear; clc; close all;

input = load('BostonListing.mat');
latitude = input.latitude;
```

```
longitude = input.longitude;
neighbourhood = input.neighbourhood;

k = 3;
points_per_cluster = [500,500,500];
[D1, D1_label] = sample_circle(k, points_per_cluster );
[D2, D2_label] = sample_spiral(k, points_per_cluster );


% part 8.1b
sigma = 0.2;
D1_distance = pdist2(D1, D1, 'euclidean');
D1_similarity_matrix = exp(-((D1_distance.^2)/(2*sigma^2)));
D1_W = D1_similarity_matrix;
D1_degree_matrix = sum(D1_W, 2);
D1_degree_matrix = diag(D1_degree_matrix);

% D1 -calculate Laplacians
D1_L_unnorm = D1_degree_matrix - D1_W;
D1_L_rw = (D1_degree_matrix^-1) * D1_L_unnorm;
D1_L_sym = (D1_degree_matrix^-(1/2)) * D1_L_unnorm * (D1_degree_matrix^-
(1/2));

% D1 - calculate eigenvalues

D1_L_unnorm_eigval = sort(eig(D1_L_unnorm), 'ascend');
D1_L_rw_eigval = sort(eig(D1_L_rw), 'ascend');
D1_L_sym_eigval = sort(eig(D1_L_sym), 'ascend');

% D2
D2_distance = pdist2(D2, D2, 'euclidean');
D2_similarity_matrix = exp(-((D2_distance.^2)/(2*sigma^2)));
D2_W = D2_similarity_matrix;
D2_degree_matrix = sum(D2_W, 2);
D2_degree_matrix = diag(D2_degree_matrix);

% D1 -calculate Laplacians
D2_L_unnorm = D2_degree_matrix - D2_W;
D2_L_rw = (D2_degree_matrix^-1) * D2_L_unnorm;
D2_L_sym = (D2_degree_matrix^-(1/2)) * D2_L_unnorm * (D2_degree_matrix^-
(1/2));

% D1 - calculate eigenvalues

D2_L_unnorm_eigval = sort(eig(D2_L_unnorm), 'ascend');
D2_L_rw_eigval = sort(eig(D2_L_rw), 'ascend');
D2_L_sym_eigval = sort(eig(D2_L_sym), 'ascend');


% 8.1bi plot
figure(1)
subplot(3,3,1);
plot(D1_L_unnorm_eigval);
title('D1 Lunnorm eigenvalues');
subplot(3,3,2);
```

```
plot(D1_L_rw_eigval);
title('D1 Lrw eigenvalues');
subplot(3,3,3);
plot(D1_L_sym_eigval);
title('D1 Lsym eigenvalues');
subplot(3,3,4);
plot(D2_L_unnorm_eigval);
title('D2 Lunnorm eigenvalues');
subplot(3,3,5);
plot(D2_L_rw_eigval);
title('D2 Lrw eigenvalues');
subplot(3,3,6);
plot(D2_L_sym_eigval);
title('D2 Lsym eigenvalues');

% find the V matrix for L, k = 3
[D1_V_L_k3 L3_val1] =  eigs(D1_L_unnorm, 3,'sm');
[D2_V_L_k3 L3_val2] =  eigs(D2_L_unnorm, 3,'sm');


% find the V matrix for Lrw, k = 3
[D1_V_Lrw_k3 Lrw3_val1] =  eigs(D1_L_rw, 3,'sm');
[D2_V_Lrw_k3 Lrw3_val2] =  eigs(D2_L_rw, 3,'sm');

% find the smallest eigenvectors for D1_L_sym
[D1_L_sym_eigvect_k2 val1] =  eigs(D1_L_sym, 2,'sm');
[D1_L_sym_eigvect_k3 val2] =  eigs(D1_L_sym, 3,'sm');
[D1_L_sym_eigvect_k4 val3] =  eigs(D1_L_sym, 4,'sm');

% V matrix of eigenvectors

% find the smallest eigenvectors for D2_L_sym
[D2_L_sym_eigvect_k2 val4] =  eigs(D2_L_sym, 2,'sm');
[D2_L_sym_eigvect_k3 val5] =  eigs(D2_L_sym, 3,'sm');
[D2_L_sym_eigvect_k4 val6] =  eigs(D2_L_sym, 4,'sm');

D1_norm_k2 = sqrt(sum(D1_L_sym_eigvect_k2.^2,2));
D1_norm_k3 = sqrt(sum(D1_L_sym_eigvect_k3.^2,2));
D1_norm_k4 = sqrt(sum(D1_L_sym_eigvect_k4.^2,2));

D2_norm_k2 = sqrt(sum(D2_L_sym_eigvect_k2.^2,2));
D2_norm_k3 = sqrt(sum(D2_L_sym_eigvect_k3.^2,2));
D2_norm_k4 = sqrt(sum(D2_L_sym_eigvect_k4.^2,2));

% normalize V
for i = 1:length(D2_L_sym_eigvect_k2)

    D1_Lsym_k2_norm(i,:) = D1_L_sym_eigvect_k2(i,:)/ D1_norm_k2(i);
    D1_Lsym_k3_norm(i,:) = D1_L_sym_eigvect_k3(i,:)/ D1_norm_k3(i);
    D1_Lsym_k4_norm(i,:) = D1_L_sym_eigvect_k4(i,:)/ D1_norm_k4(i);

    D2_Lsym_k2_norm(i,:) = D2_L_sym_eigvect_k2(i,:)/ D2_norm_k2(i);
    D2_Lsym_k3_norm(i,:) = D2_L_sym_eigvect_k3(i,:)/ D2_norm_k3(i);
    D2_Lsym_k4_norm(i,:) = D2_L_sym_eigvect_k4(i,:)/ D2_norm_k4(i);
```

```
end

[idx2,C2] = kmeans(D1_Lsym_k2_norm,2);

figure(2)
subplot(2,3,1)
plot(D1(idx2==1,1),D1(idx2==1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx2==2,1),D1(idx2==2,2),'b.','MarkerSize',10)
plot(C2(:,1),C2(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
title('D1, k = 2')

[idx3,C3] = kmeans(D1_Lsym_k3_norm,3);
subplot(2,3,2)
plot(D1(idx3==1,1),D1(idx3==1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx3==2,1),D1(idx3==2,2),'b.','MarkerSize',10)
hold on
plot(D1(idx3==3,1),D1(idx3==3,2),'g.','MarkerSize',10)
plot(C3(:,1),C3(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
title('D1, k = 3')

[idx4,C4] = kmeans( D1_Lsym_k4_norm, 4);
subplot(2,3,3)
plot(D1(idx4==1,1),D1(idx4==1,2),'r.','MarkerSize',10)
hold on
plot(D1(idx4==2,1),D1(idx4==2,2),'b.','MarkerSize',10)
hold on
plot(D1(idx4 == 3,1),D1(idx4 == 3,2),'g.','MarkerSize',10)
hold on
plot(D1(idx4 == 4,1),D1(idx4 == 4,2),'k.','MarkerSize',10)
plot(C4(:,1),C4(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
title('D1, k = 4')

[D2_idx2,D2_C2] = kmeans(D2_Lsym_k2_norm,2);

subplot(2,3,4)
plot(D2(D2_idx2==1,1),D2(D2_idx2==1,2),'r.','MarkerSize',10)
hold on
plot(D2(D2_idx2==2,1),D2(D2_idx2==2,2),'b.','MarkerSize',10)
plot(D2_C2(:,1), D2_C2(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
title('D2, k = 2')

[D2_idx3,D2_C3] = kmeans(D2_Lsym_k3_norm,3);
subplot(2,3,5)
plot(D2(D2_idx3==1,1),D2(D2_idx3==1,2),'r.','MarkerSize',10)
hold on
plot(D2(D2_idx3==2,1),D2(D2_idx3==2,2),'b.','MarkerSize',10)
hold on
plot(D2(D2_idx3==3,1),D2(D2_idx3==3,2),'g.','MarkerSize',10)
plot(D2_C3(:,1),D2_C3(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
```

```
title('D2, k = 3')


[D2_idx4, D2_C4] = kmeans( D2_Lsym_k4_norm, 4);
subplot(2,3,6)
plot(D2(D2_idx4==1,1),D2(D2_idx4==1,2),'r.','MarkerSize',10)
hold on
plot(D2(D2_idx4==2,1),D2(D2_idx4==2,2),'b.','MarkerSize',10)
hold on
plot(D2(D2_idx4 == 3,1),D2(D2_idx4 == 3,2),'g.','MarkerSize',10)
hold on
plot(D2(D2_idx4 == 4,1),D2(D2_idx4 == 4,2),'k.','MarkerSize',10)
plot(D2_C4(:,1), D2_C4(:,2),'kx',...
     'MarkerSize', 12,'LineWidth',3)
title('D2, k = 4')

% kmeans for L and k = 3
[LD1_idx3,LD1_C3] = kmeans(D1_V_L_k3,3);
[LD2_idx3,LD2_C3] = kmeans(D2_V_L_k3,3);

% kmeans for Lrw and k = 3
[LrwD1_idx3,LrwD1_C3] = kmeans(D1_V_Lrw_k3, 3);
[LrwD2_idx3,LrwD2_C3] = kmeans(D2_V_Lrw_k3, 3);

% part 8b_iii

figure(3)
subplot(2,3,1)
plot3(D1_V_L_k3(LD1_idx3==1,1),D1_V_L_k3(LD1_idx3==1,2),D1_V_L_k3(LD1_idx3==1
,3),'r.','MarkerSize',10)
hold on
plot3(D1_V_L_k3(LD1_idx3==2,1),D1_V_L_k3(LD1_idx3==2,2),D1_V_L_k3(LD1_idx3==2
,3) ,'b.','MarkerSize',10)
hold on
plot3(D1_V_L_k3(LD1_idx3==3,1),D1_V_L_k3(LD1_idx3==3,2),D1_V_L_k3(LD1_idx3==3
,3),'g.','MarkerSize',10)

title('SC-1, D1, k = 3')

subplot(2,3,2)
plot3(D1_V_Lrw_k3(LrwD1_idx3==1,1),D1_V_Lrw_k3(LrwD1_idx3==1,2),D1_V_Lrw_k3(L
rwD1_idx3==1,2),'r.','MarkerSize',10)
hold on
plot3(D1_V_Lrw_k3(LrwD1_idx3==2,1),D1_V_Lrw_k3(LrwD1_idx3==2,2),D1_V_Lrw_k3(L
rwD1_idx3==2,3) ,'b.','MarkerSize',10)
hold on
plot3(D1_V_Lrw_k3(LrwD1_idx3==3,1),D1_V_Lrw_k3(LrwD1_idx3==3,2),D1_V_Lrw_k3(L
rwD1_idx3==3,3),'g.','MarkerSize',10)

title('SC-2, D1, k = 3')


subplot(2,3,3)
plot3(D1_Lsym_k3_norm(idx3==1,1),D1_Lsym_k3_norm(idx3==1,2),D1_Lsym_k3_norm(i
dx3==1,3),'r.','MarkerSize',10)
```

```
hold on
plot3(D1_Lsym_k3_norm(idx3==2,1),D1_Lsym_k3_norm(idx3==2,2),D1_Lsym_k3_norm(i
dx3==2,3) ,'b.','MarkerSize',10)
hold on
plot3(D1_Lsym_k3_norm(idx3==3,1),D1_Lsym_k3_norm(idx3==3,2),
D1_Lsym_k3_norm(idx3==3,3),'g.','MarkerSize',10)

title('SC-3, D1, k = 3')


% plot D2

subplot(2,3,4)
plot3(D2_V_L_k3(LD2_idx3==1,1),D2_V_L_k3(LD2_idx3==1,2),D2_V_L_k3(LD2_idx3==1
,3),'r.','MarkerSize',10)
hold on
plot3(D2_V_L_k3(LD2_idx3==2,1),D2_V_L_k3(LD2_idx3==2,2),D2_V_L_k3(LD2_idx3==2
,3) ,'b.','MarkerSize',10)
hold on
plot3(D2_V_L_k3(LD2_idx3==3,1),D2_V_L_k3(LD2_idx3==3,2),D2_V_L_k3(LD2_idx3==3
,3),'g.','MarkerSize',10)

title('SC-1, D2, k = 3')


subplot(2,3,5)
plot3(D2_V_Lrw_k3(LrwD2_idx3==1,1),D2_V_Lrw_k3(LrwD2_idx3==1,2),D2_V_Lrw_k3(L
rwD2_idx3==1,3),'r.','MarkerSize',10)
hold on
plot3(D2_V_Lrw_k3(LrwD2_idx3==2,1),D2_V_Lrw_k3(LrwD2_idx3==2,2),D2_V_Lrw_k3(L
rwD2_idx3==2,3) ,'b.','MarkerSize',10)
hold on
plot3(D2_V_Lrw_k3(LrwD2_idx3==3,1),D2_V_Lrw_k3(LrwD2_idx3==3,2),D2_V_Lrw_k3(L
rwD2_idx3==3,3),'g.','MarkerSize',10)

title('SC-2, D2, k = 3')


subplot(2,3,6)
plot3(D2_Lsym_k3_norm(D2_idx3==1,1),D2_Lsym_k3_norm(D2_idx3==1,2),D2_Lsym_k3_
norm(D2_idx3==1,3),'r.','MarkerSize',10)
hold on
plot3(D2_Lsym_k3_norm(D2_idx3==2,1),D2_Lsym_k3_norm(D2_idx3==2,2),D2_Lsym_k3_
norm(D2_idx3==2,3) ,'b.','MarkerSize',10)
hold on
plot3(D2_Lsym_k3_norm(D2_idx3==3,1),D2_Lsym_k3_norm(D2_idx3==3,2),D2_Lsym_k3_
norm(D2_idx3==3,3),'g.','MarkerSize',10)

title('SC-3, D2, k = 3')
```

```
% Problem 8_1c
clear; clc; close all;

input = load('BostonListing.mat');
latitude = input.latitude;
longitude = input.longitude;
neighbourhood = input.neighbourhood;

k = 3;
points_per_cluster = [500,500,500];
[D1, D1_label] = sample_circle(k, points_per_cluster );
[D2, D2_label] = sample_spiral(k, points_per_cluster );

[D3_theta D3_ro] = cart2pol(D1(:,1), D1(:,2));
```

```
% linear transform to [0,1]
D3_theta = mat2gray(D3_theta);
D3_ro = mat2gray(D3_ro);
D3 = [D3_theta, D3_ro];

rng(2);

% D3, k = 2
figure(1)
subplot(3,1,1)
[idx_k2, C_k2] = kmeans(D3, 2, 'Replicates', 20, 'Distance', 'cityblock');

plot(D3_theta(idx_k2==1),D3_ro(idx_k2==1),'r.','MarkerSize',10)
hold on
plot(D3_theta(idx_k2==2),D3_ro(idx_k2==2),'b.','MarkerSize',10)
plot(C_k2(:,1),C_k2(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D3, k = 2')
xlabel('Angle');
ylabel('Radius');

subplot(3,1,2)
[idx_k3, C_k3] = kmeans(D3, 3, 'Replicates', 20, 'Distance', 'cityblock');

plot(D3_theta(idx_k3==1),D3_ro(idx_k3==1),'r.','MarkerSize',10)
hold on
plot(D3_theta(idx_k3==2),D3_ro(idx_k3==2),'b.','MarkerSize',10)
hold on
plot(D3_theta(idx_k3==3),D3_ro(idx_k3==3),'g.','MarkerSize',10)
plot(C_k3(:,1),C_k3(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D3, k = 3')
xlabel('Angle');
ylabel('Radius');

subplot(3,1,3)
[idx_k4, C_k4] = kmeans(D3, 4, 'Replicates', 20, 'Distance', 'cityblock');
plot(D3_theta(idx_k4==1),D3_ro(idx_k4==1),'r.','MarkerSize',10)
hold on
plot(D3_theta(idx_k4==2),D3_ro(idx_k4==2),'b.','MarkerSize',10)
hold on
plot(D3_theta(idx_k4==3),D3_ro(idx_k4==3),'g.','MarkerSize',10)
hold on
plot(D3_theta(idx_k4==4),D3_ro(idx_k4==4),'k.','MarkerSize',10)
plot(C_k4(:,1),C_k4(:,2),'kx',...
     'MarkerSize',12,'LineWidth',3)
title('D3, k = 4')
xlabel('Angle');
ylabel('Radius');

% 81c_ii l2 distance

[~, ~, sumdk2] = kmeans(D3, 2, 'Replicates', 20, 'Distance', 'sqeuclidean');
[~, ~, sumdk3] = kmeans(D3, 3, 'Replicates', 20, 'Distance', 'sqeuclidean');
[~, ~, sumdk4] = kmeans(D3, 4, 'Replicates', 20, 'Distance', 'sqeuclidean');
```

```
% Problem 2

clear; clc; close all;
input = load('BostonListing.mat');

latitude = input.latitude;
longitude = input.longitude;
nbh = input.neighbourhood;

position = [latitude, longitude];

sigma = 0.01;
distance = pdist2(position, position, 'euclidean');
similarity_matrix = exp(-((distance.^2)/(2*sigma^2)));
W = similarity_matrix;
degree_matrix = sum(W, 2);
degree_matrix = diag(degree_matrix);
```

```
% D1 -calculate Laplacians
L_unnorm = degree_matrix - W;
L_sym = (degree_matrix^-(1/2)) * L_unnorm * (degree_matrix^-(1/2));

for k = 1:25
    [Lsym_eigvect val] =  eigs(L_sym, k,'sm');

    Lsym_eigvect_norm = normr(Lsym_eigvect);
    [idx, C] = kmeans(Lsym_eigvect_norm, k);
    if k == 5
        idx5 = idx;
    end
    sum = 0;
    for j = 1:k
        neighbourhood = categorical(nbh(find(idx == j)));
        num = countcats(neighbourhood);
        max_ni = max(num);
        sum = sum + max_ni/2558;
    end
    purity(k) = sum;
end

figure(1)
plot(1:25, purity);
title('Purity metric vs. k ')
ylabel('Purity');
xlabel('k values');

% Hw8_2b

figure(2)
plot(longitude(idx5==1),latitude(idx5==1),'r.','MarkerSize',10)
hold on
plot(longitude(idx5==2),latitude(idx5==2),'b.','MarkerSize',10)
hold on
plot(longitude(idx5==3),latitude(idx5==3),'g.','MarkerSize',10)
hold on
plot(longitude(idx5==4),latitude(idx5==4),'m.','MarkerSize',10)
hold on
plot(longitude(idx5==5),latitude(idx5==5),'y.','MarkerSize',10)
plot_google_map
```