Università della Svizzera italiana

Faculty of Informatics

May 2, 2018

Stochastic Methods

Assignment 3

Sumeet Gyanchandani

1 L-1 Regularization

Listing 1. myl1reg.m

```
close all;
clear all;

load('data_3.mat');


d = size(X,2);
X = [ones(size(X, 1), 1) X];
theta = inv(X'*X) * X' * Y;

A = [eye(d), -eye(d);...
-eye(d), -eye(d);...
zeros(d), -eye(d);...
zeros(1,d), ones(1,d)];

A = [zeros(3*d+1,1) A];

x0 = zeros(2*d+1,1);
fun = @(x)(Y - (X*x(1:d+1)))'*(Y-(X*x(1:d+1)));

for i = 1:20
b = [zeros(1,(3*d)), C(i)]';
theta_reg(:,i) = fmincon(fun,x0,A,b);
error(i) = fun(theta_reg(:,i));
end

figure;
plot(C,error);

figure;
loglog(C,error);
```

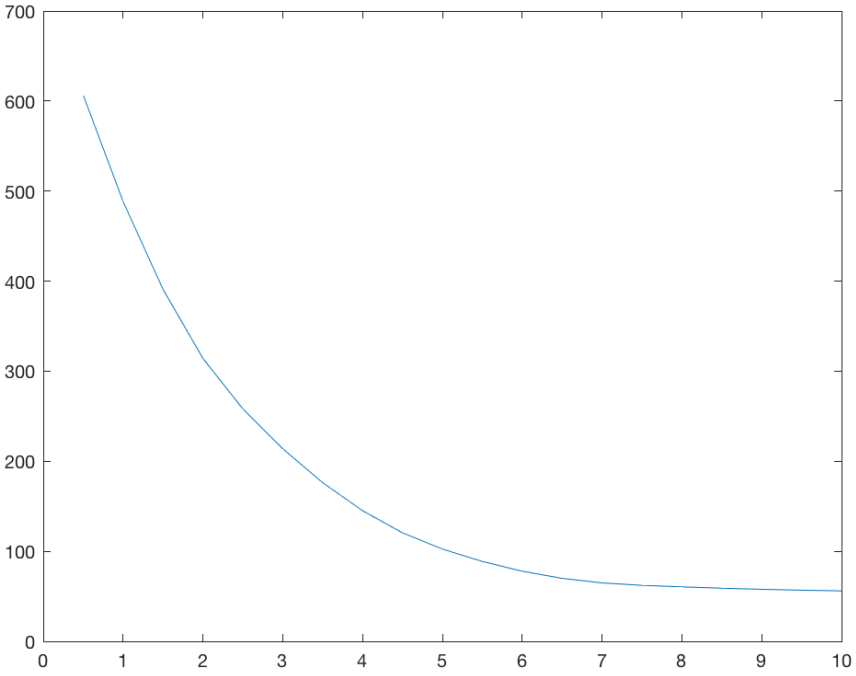

Figure 1. L-Curve

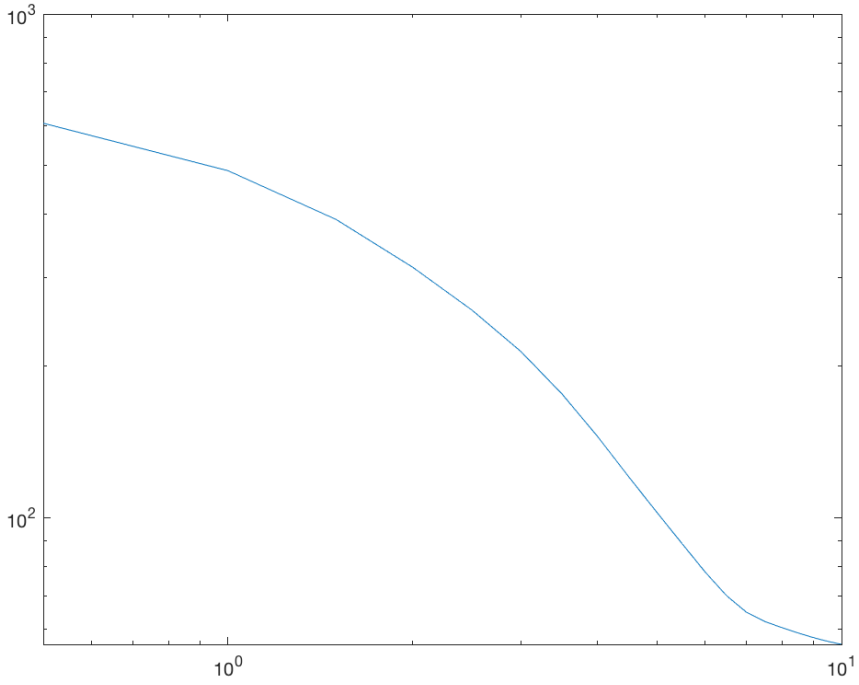

Figure 2. Log-Log

2 K-means Clustering

2(a) Visualize Dataset

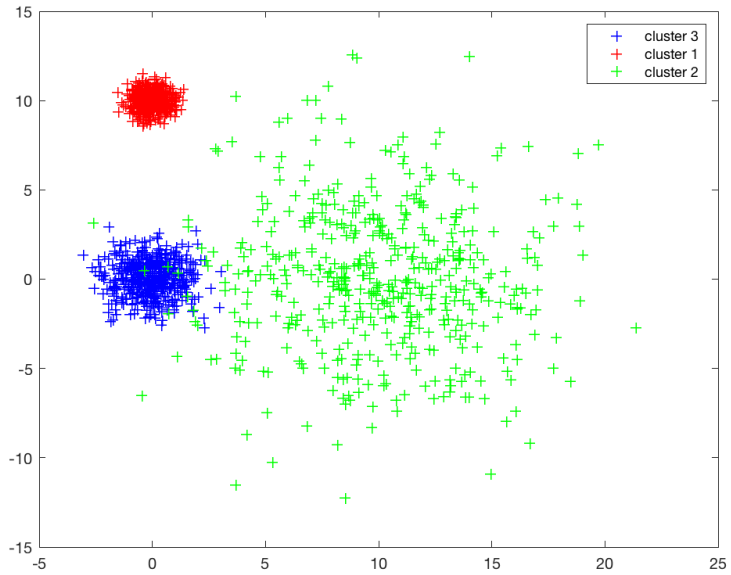

Figure 3. Visualizing the data given in **k_means_data.mat**

2(b) Clustering

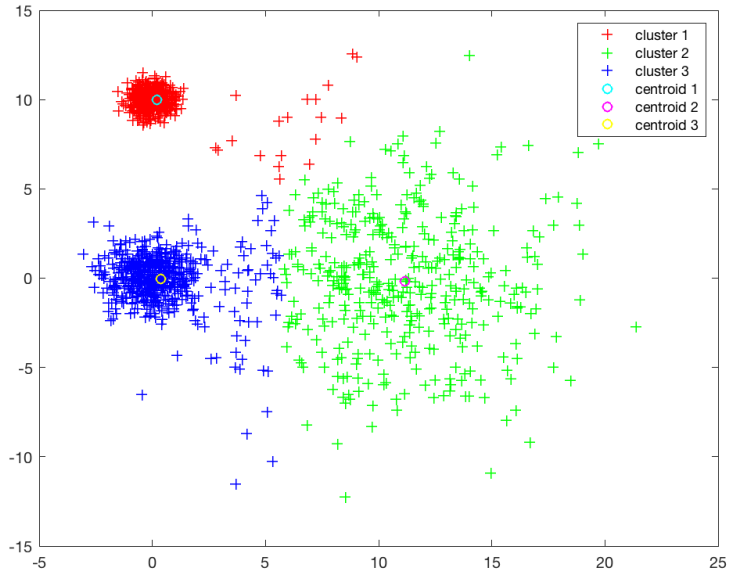

Figure 4. Data clustered by standard k-means clustering method

As depicted by the above two images, the standard k-means clustering method does a good job of clustering the dataset, with only exception being the boundary points between two clusters. But it is still very **unreliable**. If the initial selection of centroids is not good, we will end up with clusters as shown below.

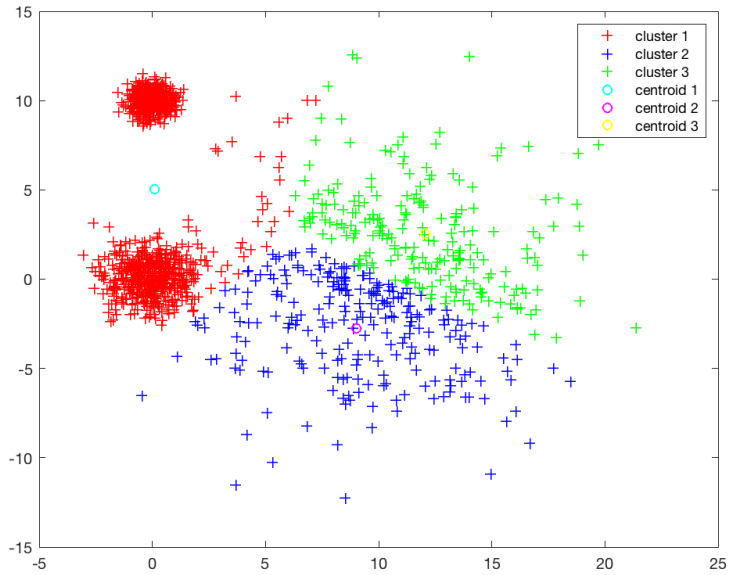

Figure 5. Example of bad clustering

Listing 2. mykmeans.m

```
close all;
clear all;

%Visualize Dataset
load('k_means_data.mat');
figure;
gscatter(x(1,:), x(2,:), g,'brg','+++');
legend('cluster 3','cluster 1','cluster 2','Location','northeast')

%Clustering
rng(3);
[IDX_i, C] = kmeans(x', 3);
figure;
gscatter(x(1,:), x(2,:), IDX_i,'rgb','+++');
hold on
plot(C(1,1),C(1,2),'co');
plot(C(2,1),C(2,2),'mo');
plot(C(3,1),C(3,2),'yo');
legend('cluster 1','cluster 2','cluster 3','centroid 1','centroid 2','centroid 3','Location','northeast↩
    ')
hold off

%Centroid gone wrong
rng(27);
[IDX_i, C] = kmeans(x', 3);
figure;
gscatter(x(1,:), x(2,:), IDX_i,'rbg','+++');
hold on
plot(C(1,1),C(1,2),'co');
plot(C(2,1),C(2,2),'mo');
plot(C(3,1),C(3,2),'yo');
legend('cluster 1','cluster 2','cluster 3','centroid 1','centroid 2','centroid 3','Location','northeast↩
    ')
hold off
```