EE511 Project 3 Summary

SilongHu email:silonghu@usc.edu

This summary is divided into 3 parts: Implementation, Conclusion and Source Code

I. Implementation

Problem 1. [Testing Faith]

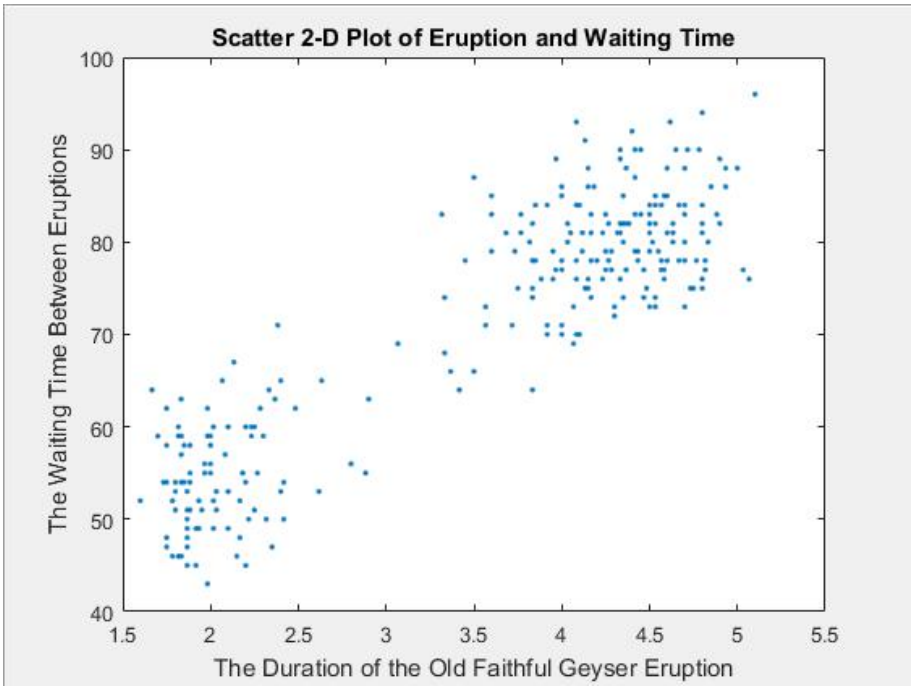

Figure 1. Scatter 2-D Plot of Eruption and Waiting Time

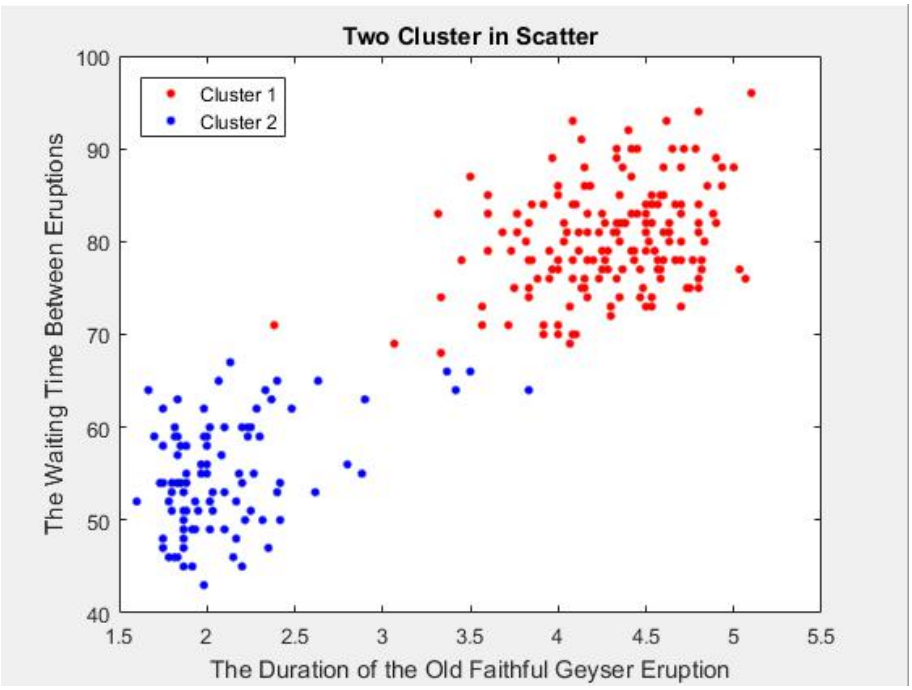

Figure 2. Two Cluster in Scatter

Using ‘kmeans’ function to get two clusters and using ‘scatter’ to plot them.

Problem 2. [Generating Mixed Samples]

Implement a random number generator for a random variable with the following mixture distribution. $f(x) = 0.4N(-1,1) + 0.6N(1,1)$. Using ‘rand’ to generate 1000*1 vector from 0 to 1. If element bigger than 0.6, f(x) takes N(-1,1), otherwise, f(x) takes N(1,1).

```
r = rand(1,N);
x1 = (r > 0.6).*R1;
x2 = (r <= 0.6).*R2;
```

Code Piece 1. Binomial Rejection Method

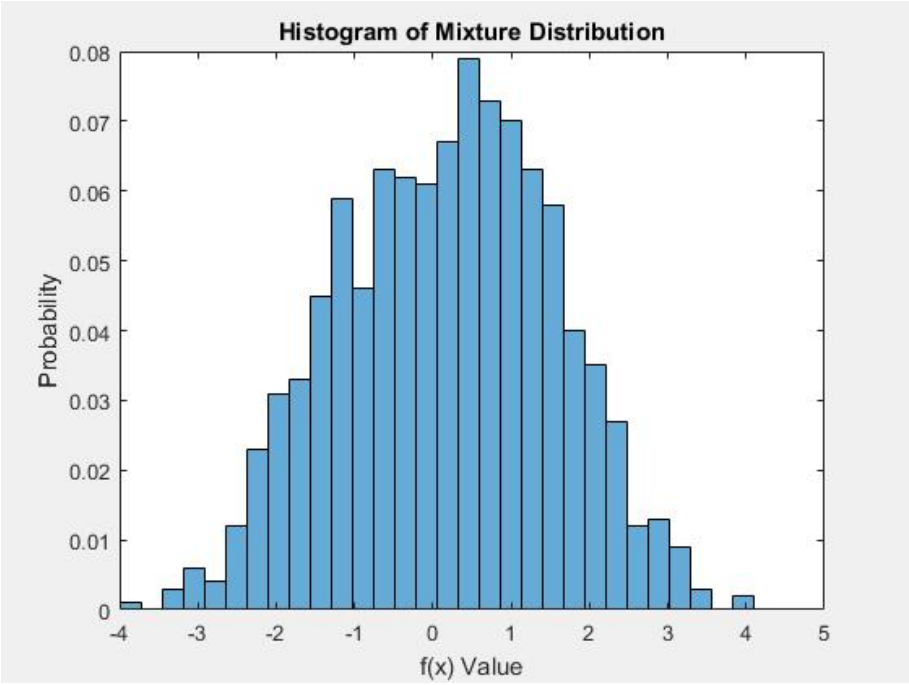

Figure 3. Mixture Distribution

After plotting the mixture distribution as shown in Figure 3, we use kmeans to separate 1000*1 vector into 2 cluster. Then histogram both of them. It shows in Figure 4.

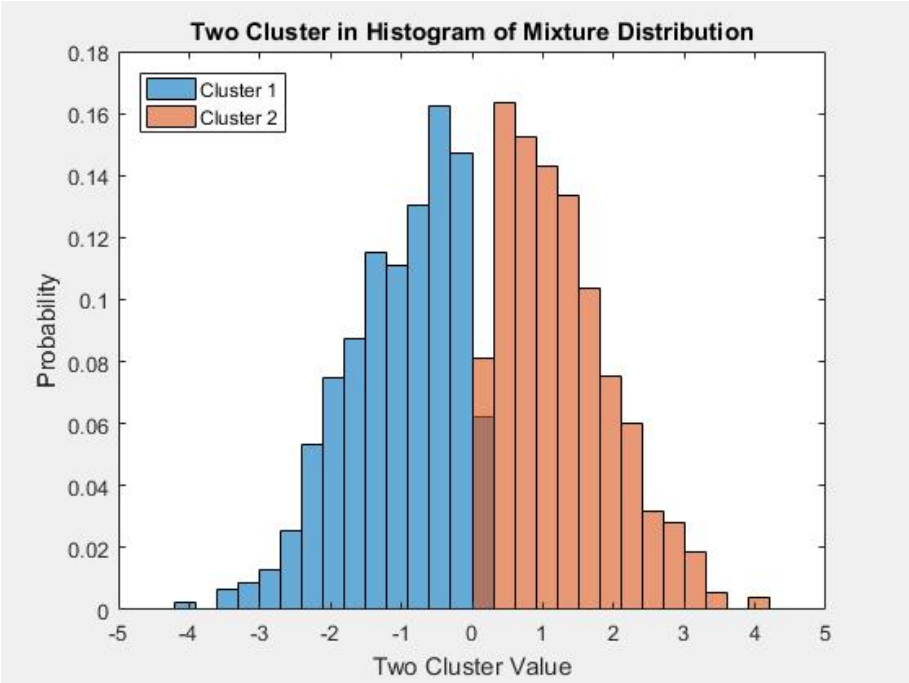

Figure 4. Two Cluster in Histogram of Mixture Distribution

II. Conclusion

In this project, we learned an important function named “kmeans” and by using that function we could get the clusters we want. By scattering the clusters in one figure, we could recognize the different cluster obviously.

For the mixed distribution, using binomial rejection method to generate. Also we could just multiply the 0.4 or 0.6 directly. From figure 3 and figure 4, we could clearly separate two clusters in histogram.

III. Source Code

1. faith.m

```
clear,clc;
A = load('Trim_faithful.txt');
A = A(:,2:3);
plot(A(:,1),A(:,2),'.');
title('Scatter 2-D Plot of Eruption and Waiting Time');
xlabel('The Duration of the Old Faithful Geyser Eruption');
ylabel('The Waiting Time Between Eruptions');
idx = kmeans(A,2) ;

figure;
plot(A(idx==1,1),A(idx==1,2),'r.','MarkerSize',12)
hold on
```

```
plot(A(idx==2,1),A(idx==2,2),'b.','MarkerSize',12)
title('Two Cluster in Scatter');
xlabel('The Duration of the Old Faithful Geyser Eruption');
ylabel('The Waiting Time Between Eruptions');
legend('Cluster 1','Cluster 2',...
      'Location','NW')
```

2. mixed.m

```
clear,clc;
R1 = normrnd(-1,1,[1,1000]);
R2 = normrnd(1,1,[1,1000]);
mu  = [-1 1];
std = [1 1];
N = 1000;
r = rand(1,N);
x1 = (r > 0.6).*R1;
x2 = (r <= 0.6).*R2;
%x1 = (r >0.6).*(mu(1) + std(1)*randn(1,N));
%x2 = (r<=0.6).*(mu(2) + std(2)*randn(1,N));
x = x1+x2;
%figure;
%hold on
histogram(x,30,'Normalization','probability');
title('Histogram of Mixture Distribution');
xlabel('f(x) Value');
ylabel('Probability');


A = x';
%A = [x1;x2]';
idx = kmeans(A,2) ;
A1 = A(idx==1);
A2 = A(idx==2);
figure;
histogram(A1,'Normalization','probability');
hold on
histogram(A2,'Normalization','probability');

title('Two Cluster in Histogram of Mixture Distribution');
xlabel('Two Cluster Value');
ylabel('Probability');
legend('Cluster 1','Cluster 2',...
      'Location','NW')
```