

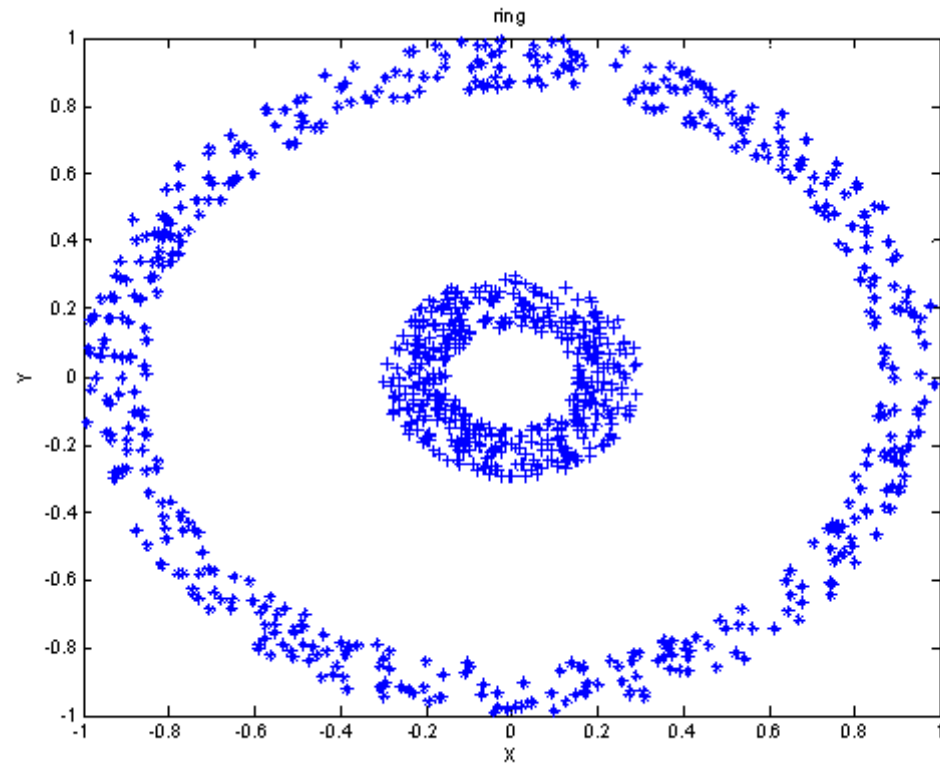
# 聚类——人工合成数据集

作者：凯鲁嘎吉 - 博客园 <http://www.cnblogs.com/kailugaji/>

## 1.Ring

```
function [data,real_label]=generate_ring(large_R,small_R,width,datanum_one_cluster)
%[data,real_label]=generate_ring(0.85,0.15,0.15,500)
%前两列是数据，最后一列是标签（1，2）
%datanum_one_cluster: 每一类点的个数
format long
seita_1=rand(1,datanum_one_cluster)*2*pi;%角度
r_1=rand(1,datanum_one_cluster)*width+large_R;%半径
x_1=r_1.*cos(seita_1);
y_1=r_1.*sin(seita_1);
A=[x_1; y_1; ones(1,datanum_one_cluster)];
plot(x_1,y_1,'b*');
hold on;
fid=fopen('ring.txt','w');
fprintf(fid,'% .2f % .2f %d\n',A);
seita_2=rand(1,datanum_one_cluster)*2*pi;%角度
r_2=rand(1,datanum_one_cluster)*width+small_R;%半径
x_2=r_2.*cos(seita_2);
y_2=r_2.*sin(seita_2);
B=[x_2; y_2; 2*ones(1,datanum_one_cluster)];
plot(x_2,y_2,'b+');
xlabel('X');
ylabel('Y');
title('ring');
fprintf(fid,'% .2f % .2f %d\n',B);
fclose(fid);
data_load=dlmread('ring.txt');
data=data_load(:,1:2);
real_label=data_load(:,3);
```

图像为：

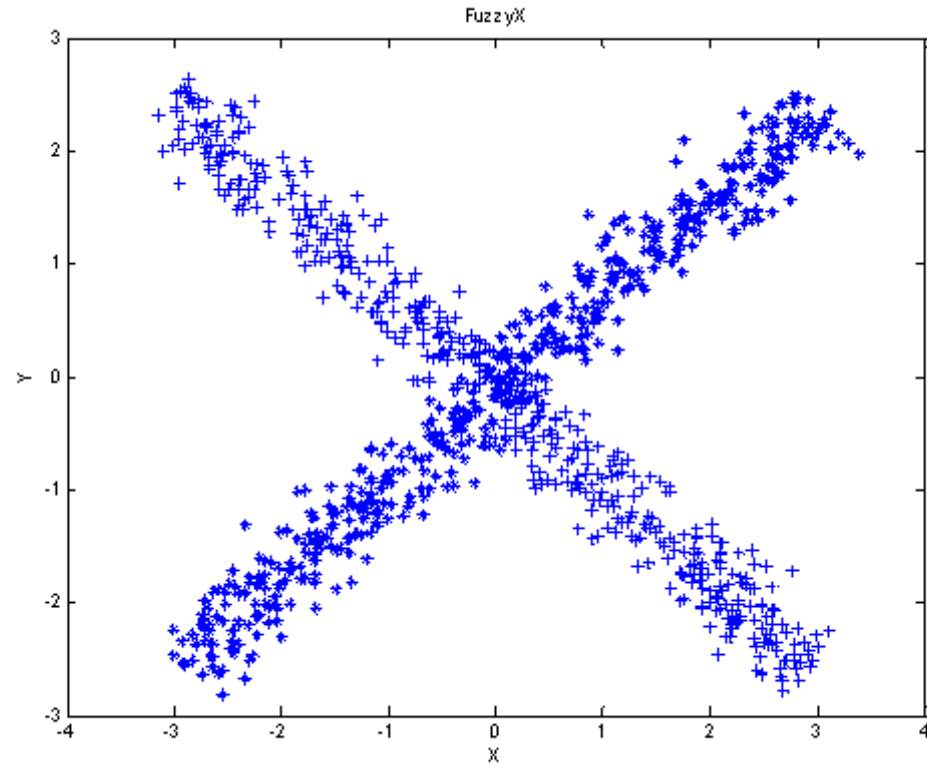


## 2. FuzzyX

```
function [data,real_label]=generate_FuzzyX(X_Range, Y_Range, width, datanum_one_cluster)
%datanum_one_cluster: 每类数据个数
% [data,real_label]=generate_FuzzyX([-3,3], [-2.5,2.5], 0.2, 500)
fid=fopen('FuzzyX.txt','w');
x_1=rand(1,datanum_one_cluster)*(X_Range(2)-X_Range(1))+X_Range(1);
x_2=rand(1,datanum_one_cluster)*(X_Range(2)-X_Range(1))+X_Range(1);
seital=atan((Y_Range(2)-Y_Range(1))/(X_Range(2)-X_Range(1)));%角度
y_1=x_1*tan(seital);
bias=randn(1,datanum_one_cluster)*width-width/2;
bias_x=bias*cos(pi/2-seital);
bias_y=bias*sin(pi/2-seital);
x_1=x_1-bias_x;
y_1=y_1+bias_y;
plot(x_1,y_1,'b*');
hold on;
```

```
A=[x_1; y_1; ones(1, datanum_one_cluster)];  
fprintf(fid, '%.2f %.2f %d\n', A);  
seita2=-seita1;%角度  
y_2=x_2*tan(seita2);  
bias=randn(1, datanum_one_cluster)*width-width/2;  
bias_x=bias*cos(pi/2-seita2);  
bias_y=bias*sin(pi/2-seita2);  
x_2=x_2-bias_x;  
y_2=y_2+bias_y;  
plot(x_2, y_2, 'b+');  
hold on;  
xlabel('X');  
ylabel('Y');  
title('FuzzyX');  
B=[x_2; y_2; 2*ones(1, datanum_one_cluster)];  
fprintf(fid, '%.2f %.2f %d\n', B);  
fclose(fid);  
data_load=dlmread('FuzzyX.txt');  
data=data_load(:, 1:2);  
real_label=data_load(:, 3);
```

图像为：



### 3. Parabolic

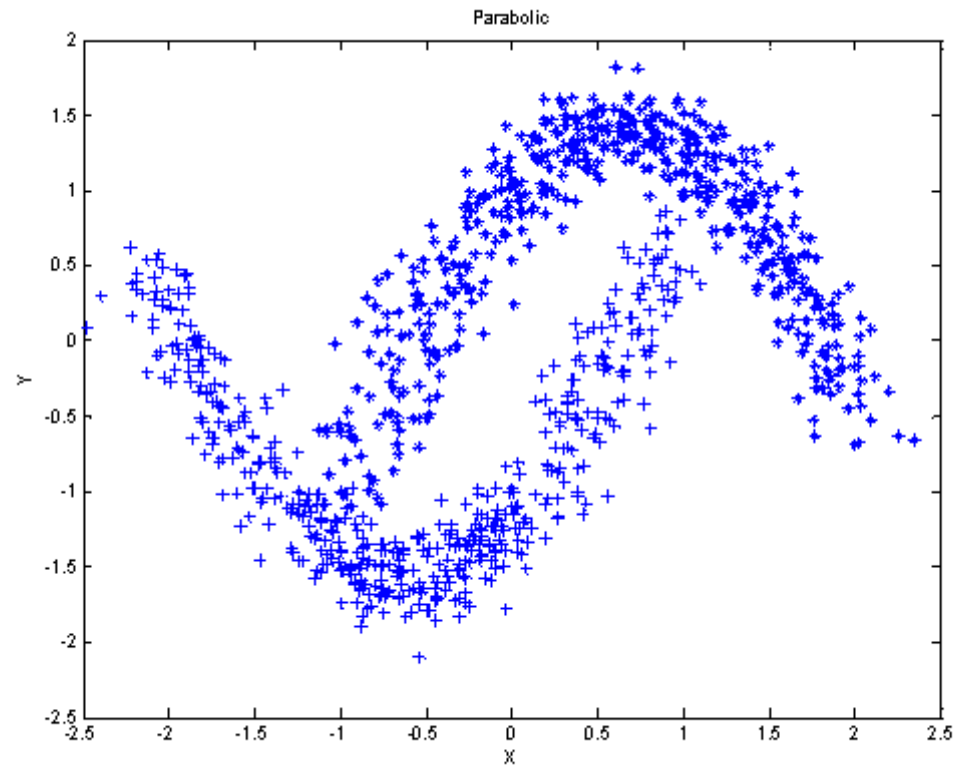
```
function [data,real_label]=generate_Parabolic(X1_Range, X2_Range, width, datanum_one_cluster)
%datanum_one_cluster: 每类数据个数
% [data,real_label]=generate_Parabolic([-2,1], [-1,2], 0.15, 500)
x_1=rand(1,datanum_one_cluster)*(X1_Range(2)-X1_Range(1))+X1_Range(1);
x_2=rand(1,datanum_one_cluster)*(X2_Range(2)-X2_Range(1))+X2_Range(1);
fid=fopen('Parabolic.txt','w');
y_1=(x_1+0.55).^2-1.5;
bias_x=randn(1,datanum_one_cluster)*width-width/2;
bias_y=randn(1,datanum_one_cluster)*width-width/2;
x_1=x_1+bias_x;
y_1=y_1+bias_y;
plot(x_1,y_1,'b+');
hold on;
A=[x_1; y_1; ones(1,datanum_one_cluster)];
fprintf(fid,'%0.2f %0.2f %d\n',A);
```

```

y_2=-(x_2-0.55).^2+1.5;
bias_x=randn(1,datanum_one_cluster)*width-width/2;
bias_y=randn(1,datanum_one_cluster)*width-width/2;
x_2=x_2-bias_x;
y_2=y_2+bias_y;
plot(x_2,y_2,'b*');
hold on;
xlabel('X');
ylabel('Y');
title('Parabolic');
B=[x_2; y_2; 2*ones(1,datanum_one_cluster)];
fprintf(fid,'% .2f %.2f %d\n',B);
fclose(fid);
data_load=dlmread('Parabolic.txt');
data=data_load(:,1:2);
real_label=data_load(:,3);

```

图像为:



## 4. Zigzag

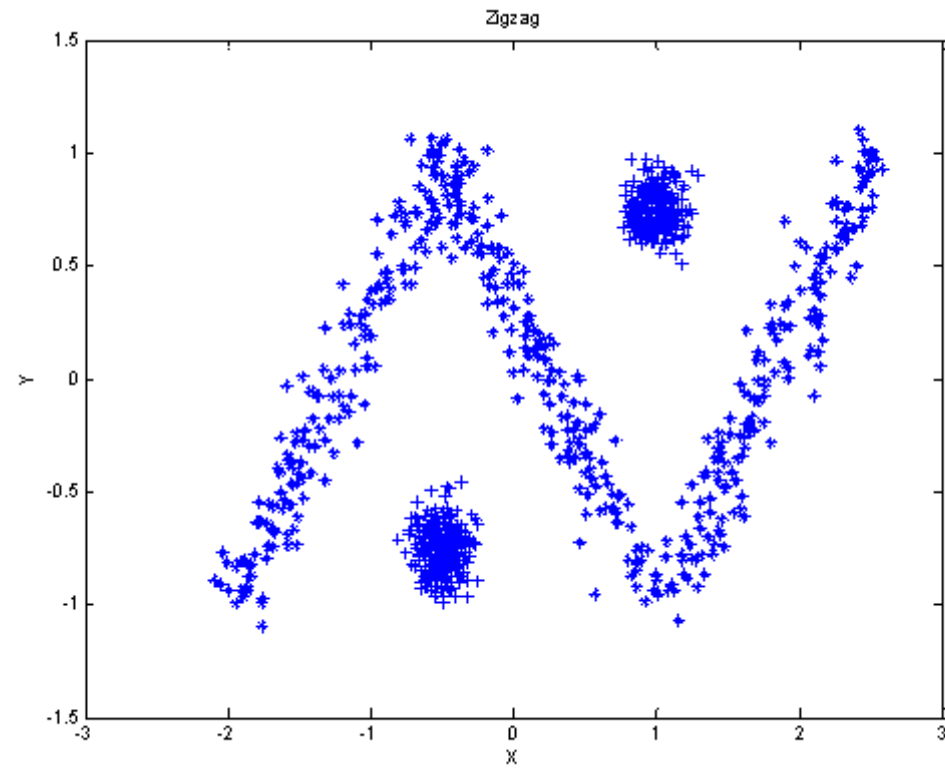
```
function [data,real_label]=generate_Zigzag(width, datanum_one_cluster)
%datanum_one_cluster: 每类数据个数
% [data,real_label]=generate_Zigzag(0.1, 501)
fid=fopen('Zigzag.txt','w');
seita1=atan(2/1.5);
seita2=atan(-2/1.5);
seita3=seita1;
x_1=rand(1,datanum_one_cluster/3)*1.5-2;
x_2=rand(1,datanum_one_cluster/3)*1.5-0.5;
x_3=rand(1,datanum_one_cluster/3)*1.5+1;
y_1=x_1*tan(seita1)+1.7;
y_2=x_2*tan(seita2)+0.4;
y_3=x_3*tan(seita3)-2.3;

bias=randn(1,datanum_one_cluster)*width-width/2;
bias_x_1=bias(1:datanum_one_cluster/3)*cos(pi/2-seita1);
bias_y_1=bias(1:datanum_one_cluster/3)*sin(pi/2-seita1);
bias_x_2=bias(datanum_one_cluster/3+1:2*datanum_one_cluster/3)*cos(pi/2-seita2);
bias_y_2=bias(datanum_one_cluster/3+1:2*datanum_one_cluster/3)*sin(pi/2-seita2);
bias_x_3=bias(2*datanum_one_cluster/3+1:3*datanum_one_cluster/3)*cos(pi/2-seita3);
bias_y_3=bias(2*datanum_one_cluster/3+1:3*datanum_one_cluster/3)*sin(pi/2-seita3);

x_1=x_1-bias_x_1;
y_1=y_1+bias_y_1;
x_2=x_2-bias_x_2;
y_2=y_2+bias_y_2;
x_3=x_3-bias_x_3;
y_3=y_3+bias_y_3;
x=[x_1';x_2';x_3'];
y=[y_1';y_2';y_3'];
plot(x,y,'b*');
hold on;
A=[x'; y'; ones(1,datanum_one_cluster)];
fprintf(fid,'% .2f % .2f %d\n',A);
mu1=[-0.5,-0.75];
sigma1=[0.01 0; 0 0.01];
z1=mvnrnd(mu1,sigma1,ceil(datanum_one_cluster/2));
mu2=[1,0.75];
sigma2=[0.01 0; 0 0.01];
z2=mvnrnd(mu2,sigma2,datanum_one_cluster-ceil(datanum_one_cluster/2));
z=[z1;z2];
plot(z(:,1),z(:,2),'b+');
hold on;
xlabel('X');
```

```
ylabel('Y');  
title('Zigzag');  
B=[(z(:,1))'; (z(:,2))'; 2*ones(1,datanum_one_cluster)];  
fprintf(fid,'% .2f %.2f %d\n',B);  
fclose(fid);  
data_load=dlmread('Zigzag.txt');  
data=data_load(:,1:2);  
real_label=data_load(:,3);
```

图像为：



## 5. demo与数据

demo.m

```
clear
clc
figure(1)
[data_1,real_label_1]=generate_ring(0.85,0.15,0.15,500);
saveas(gcf,sprintf('demo_Ring.jpg'),'bmp'); %保存图片
figure(2)
[data_2,real_label_2]=generate_FuzzyX([-3,3], [-2.5,2.5], 0.2, 500);
saveas(gcf,sprintf('demo_FuzzyX.jpg'),'bmp'); %保存图片
figure(3)
[data_3,real_label_3]=generate_Parabolic([-2,1], [-1,2], 0.15, 500);
saveas(gcf,sprintf('demo_Parabolic.jpg'),'bmp'); %保存图片
figure(4)
[data_4,real_label_4]=generate_Zigzag(0.1, 501);
saveas(gcf,sprintf('demo_Zigzag.jpg'),'bmp'); %保存图片
```

生成的数据已保存在当前文件夹下面的四个TXT文档中。可以用matlab打开。

