

```

clc;
clear;
close all;

colormap('jet');
hold on;

mu1 = [10, 3];
Sigma1 = [1, 0; 0, 1];
mu2 = [1, 1];
Sigma2 = [1.5, 0; 0, 1.5];
mu3 = [5, 4];
Sigma3 = [2, 0; 0, 2];

N = 200;

R1 = zeros(N, 3);
R2 = zeros(N, 3);
R3 = zeros(N, 3);

R1(:, 1:2) = mvnrnd(mu1, Sigma1, N);
R2(:, 1:2) = mvnrnd(mu2, Sigma2, N);
R3(:, 1:2) = mvnrnd(mu3, Sigma3, N);

R = [R1; R2; R3];

MinXY = min(min(min(R1(:, 1), R1(:, 2)), min(min(R2(:, 1), R2(:, 2)))));
MaxXY = max(max(max(R1(:, 1), R1(:, 2)), max(max(R2(:, 1), R2(:, 2)))));
MinXXYY = min(MinXY, min(min(R3(:, 1), R3(:, 2))));
MaxXXYY = max(MaxXY, max(max(R3(:, 1), R3(:, 2))));
X = MinXXYY:0.1:MaxXXYY;
Y = MinXXYY:0.1:MaxXXYY;

xlabel('x1');
ylabel('x2');
axis([MinXXYY, MaxXXYY, MinXXYY, MaxXXYY]);

plot(R1(:, 1), R1(:, 2), '+r');
plot(R2(:, 1), R2(:, 2), '+g');
plot(R3(:, 1), R3(:, 2), '+b');

index = randperm(numel(R(:, 1)));

c1 = [R(index(1), 1), R(index(1), 2)];
c2 = [R(index(2), 1), R(index(2), 2)];
c3 = [R(index(3), 1), R(index(3), 2)];

plot(c1(1, 1), c1(1, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'r', 'MarkerSize', 6);
plot(c2(1, 1), c2(1, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'g', 'MarkerSize', 6);
plot(c3(1, 1), c3(1, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'b', 'MarkerSize', 6);

```

```

f1 = zeros(size(X, 1), size(X, 1));
f2 = zeros(size(X, 1), size(X, 1));
f3 = zeros(size(X, 1), size(X, 1));

i = 1;
for y = MinXXYY:0.1:MaxXXYY
    j = 1;
    for x = MinXXYY:0.1:MaxXXYY
        f1(i, j) = mvar(x, y, mu1(1), mu1(2), Sigma1);
        f2(i, j) = mvar(x, y, mu2(1), mu2(2), Sigma2);
        f3(i, j) = mvar(x, y, mu3(1), mu3(2), Sigma3);
        j = j + 1;
    end;
    i = i + 1;
end;

contour(MinXXYY:0.1:MaxXXYY, MinXXYY:0.1:MaxXXYY, f1, 8, 'LineWidth', 1.5);
contour(MinXXYY:0.1:MaxXXYY, MinXXYY:0.1:MaxXXYY, f2, 8, 'LineWidth', 1.5);
contour(MinXXYY:0.1:MaxXXYY, MinXXYY:0.1:MaxXXYY, f3, 8, 'LineWidth', 1.5);

d1 = zeros(numel(R(:, 1)), 1);
d2 = zeros(numel(R(:, 1)), 1);
d3 = zeros(numel(R(:, 1)), 1);

for i = 1:20
    for j = 1:numel(R(:, 1))
        d1(j) = euclideanDistance(c1(1, 1), R(j, 1), c1(1, 2), R(j, 2));
        d2(j) = euclideanDistance(c2(1, 1), R(j, 1), c2(1, 2), R(j, 2));
        d3(j) = euclideanDistance(c3(1, 1), R(j, 1), c3(1, 2), R(j, 2));
    end

    d = [d1, d2, d3];

    for k = 1:numel(R(:, 1))
        [Value, Index] = min(d(k, :));
        R(k, 3) = Index;
    end

    ind1 = find(R(:, 3) == 1);
    ind2 = find(R(:, 3) == 2);
    ind3 = find(R(:, 3) == 3);

    Newc1_x = mean(R(ind1, 1));
    Newc1_y = mean(R(ind1, 2));
    c1 = [Newc1_x, Newc1_y];

    Newc2_x = mean(R(ind2, 1));
    Newc2_y = mean(R(ind2, 2));
    c2 = [Newc2_x, Newc2_y];

    Newc3_x = mean(R(ind3, 1));
    Newc3_y = mean(R(ind3, 2));
    c3 = [Newc3_x, Newc3_y];

```

```

        plot(c1(:, 1), c1(:, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'r', 'MarkerSize', 6);
        plot(c2(:, 1), c2(:, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'g', 'MarkerSize', 6);
        plot(c3(:, 1), c3(:, 2), 'ob', 'MarkerEdgeColor', 'k', 'MarkerFaceColor',
'b', 'MarkerSize', 6);
        drawnow
        title(['Iteration Number: ', num2str(i)]);
        pause(0.5);
end
print('Not_labeled', '-dpng', '-r150');

for i = 1:numel(R(:, 1))
    text(R(i, 1), R(i, 2), num2str(R(i, 3)));
end
print('Labeled', '-dpng', '-r150');

C1 = unique(R(1:N, 3));
C2 = unique(R(N+1:2*N, 3));
C3 = unique(R(2*N+1:3*N, 3));

Acc1 = zeros(3, 1);
for i = 1:numel(C1)
    Acc1(i, 1) = sum(R(1:N, 3) == C1(i, 1));
end

Acc2 = zeros(3, 1);
for i = 1:numel(C2)
    Acc2(i, 1) = sum(R(N+1:2*N, 3) == C2(i, 1));
end

Acc3 = zeros(3, 1);
for i = 1:numel(C3)
    Acc3(i, 1) = sum(R(2*N+1:3*N, 3) == C3(i, 1));
end

disp('Clustering accuracy - cluster 1');
disp('-----');
disp(['lable: ', num2str(mode(R(1:N, 3)))]);
disp('-----');
disp(max(Acc1)/numel(R1(:, 1)));

disp('Clustering accuracy - cluster 2');
disp('-----');
disp(['lable: ', num2str(mode(R(N+1:2*N, 3)))]);
disp('-----');
disp(max(Acc2)/numel(R2(:, 1)));

disp('Clustering accuracy - cluster 3');
disp('-----');
disp(['lable: ', num2str(mode(R(2*N+1:3*N, 3)))]);
disp('-----');
disp(max(Acc3)/numel(R3(:, 1)));

```

در ابتدا تعداد 100 عدد داده در سه دسته متفاوت، توسط توزیع گوسی با پارامترهای:

$$\mu_1 = \begin{bmatrix} 10 \\ 3 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mu_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}, \mu_3 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

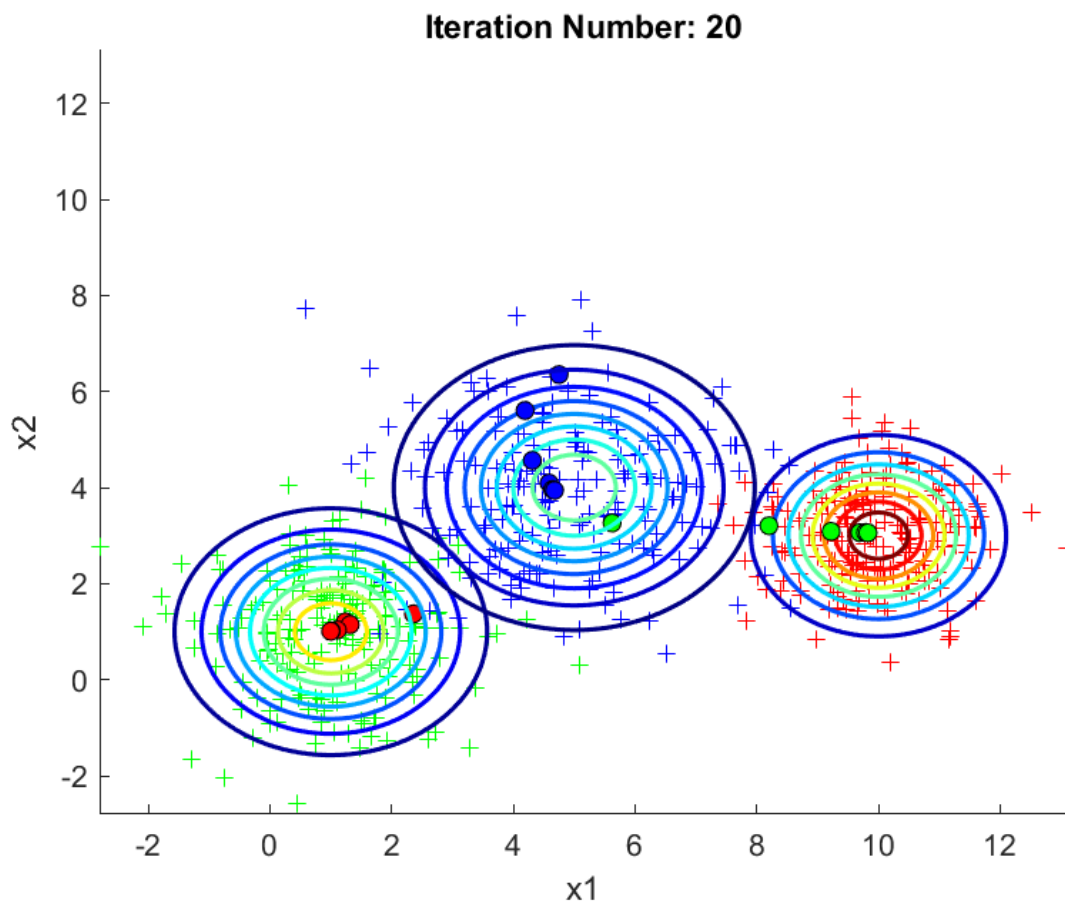
تولید می‌گردند؛ بنابراین، هر دسته شامل 200 عدد متغیر تصادفی (ایجاد شده به واسطه پارامترهای در نظر گرفته شده) در فضای دو بُعدی خواهد بود. پس‌از آن، با استفاده از تابع randperm مراکز اولیه سه خوشه، از میان 600 عدد تصادفی تولید شده به صورت تصادفی (بدون جایگذاری) انتخاب می‌شوند. گفتنی است که نمودارهای هم‌تراز تنها به دلیل بصری به نمودارها اضافه گردیده‌اند و تأثیری در روند اجرای برنامه نخواهند داشت.

در ادامه فاصله اقلیدسی هر نقطه از مراکز در نظر گرفته شده اندازه‌گیری شده و به عنوان مثال، اگر داده‌ای با مختصات (x1, x2) به مرکز C1 نزدیک بود برچسب مرکز C1 (1) را کسب می‌نماید. پس‌از آن، در یک روند تکراری داده‌های با برچسب یکسان در تولید میانگین خود و در نتیجه مرکز خوشه مربوط به خود مشارکت نموده و مجدداً برچسب‌گذاری می‌شوند.

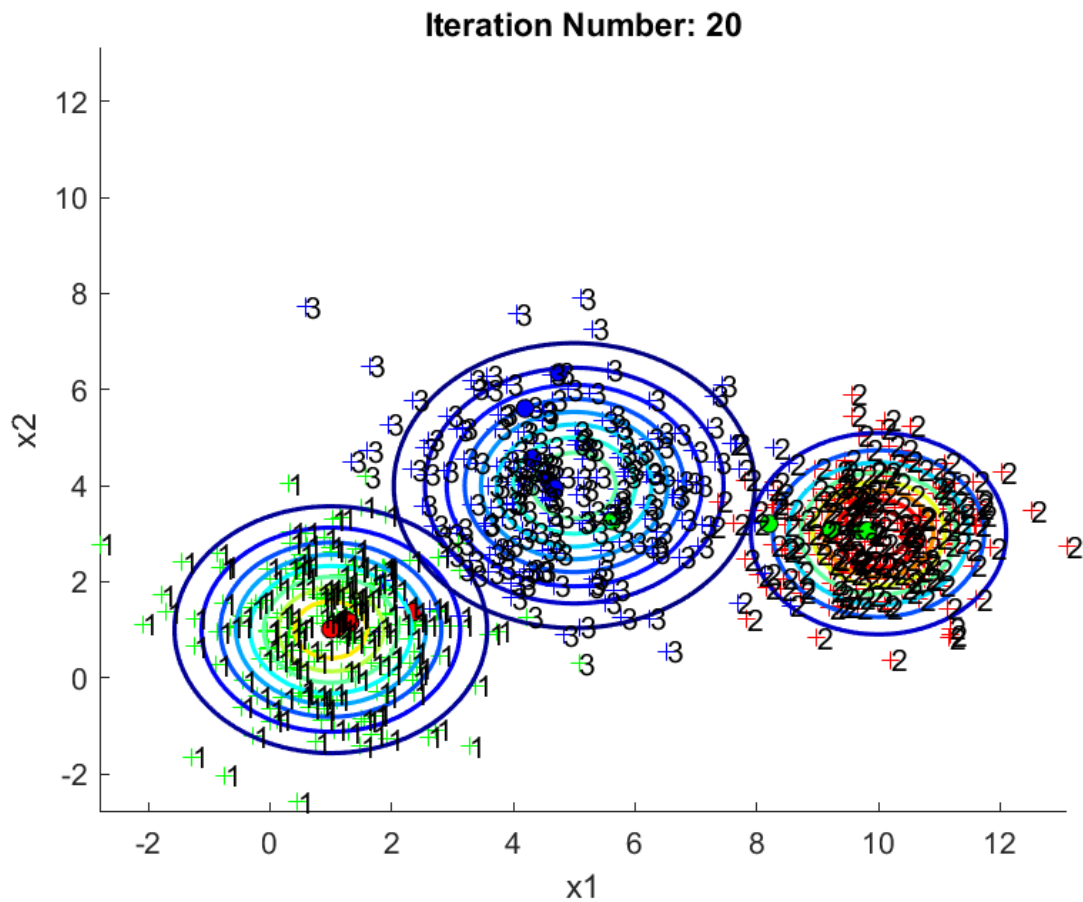
در طی این روند مراکز نهایی خوشه‌ها شناسایی شده و هر مرکز خوشه، در دل دسته‌ای از متغیرهای تصادفی ایجاد شده جای خواهد گرفت.

در انتهای فرایند، تمامی داده‌ها خوشه‌بندی شده و از آنجایی که می‌دانیم داده‌های 1 الی 200 مربوط به دسته 1 (خوشه 1) بوده‌اند بنابراین بیشترین تعداد برچسب تکرار شده در این دسته از اعداد را شمارش کرده و بر 200 تقسیم می‌کنیم. چنین روشی را برای خوشه 2 و با داده‌های 201 الی 400 و خوشه 3 با داده‌های 401 الی 600 به کار خواهیم گرفت.

در پایان گفتن این نکته ضروری است که به واسطه ذات الگوریتم، برچسب‌های اختصاص داده شده به هر یک از خوشه‌ها حتماً با اندیس خوشه نمایش داده شده در بررسی دقت یکی نخواهند بود.



نتیجۀ نهایی خوشه‌بندی k-mean - بدون درج برچسب‌ها



نتیجه نهایی خوشه‌بندی k-mean - همراه با درج برچسب‌ها

دقت حاصل از خوشه‌بندی:

Clustering accuracy - cluster 1

lable: 2

1

Clustering accuracy - cluster 2

lable: 1

0.9700

Clustering accuracy - cluster 3

lable: 3

0.9200