

قسمت اول

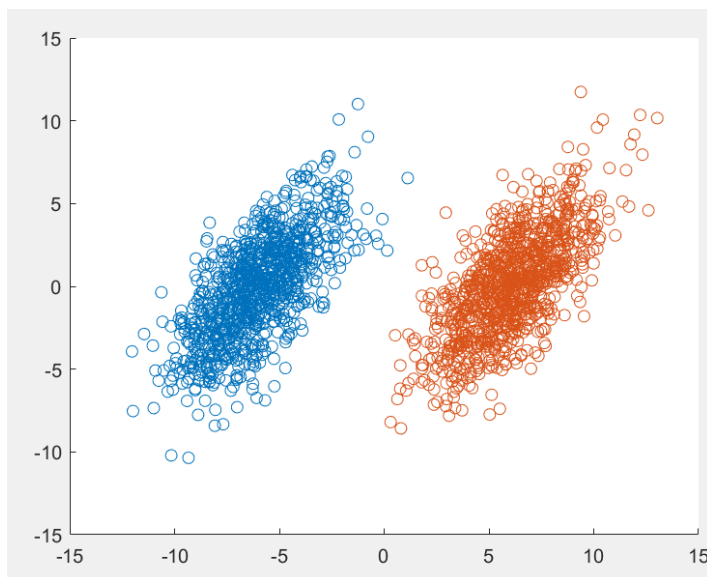
ک مساله دو کلاسه را در نظر گرفته و 1000 نمونه تصادفی با توزیع گاوسی با بردارهای میانگین و ماتریس کوواریانس داده شده در زیر برای هر کدام تولید کنید.

$$\mu_1 = [10, 10]^T, \mu_2 = [22, 10]^T, \Sigma_1 = \Sigma_2 = \begin{bmatrix} 4 & 4 \\ 4 & 9 \end{bmatrix}$$

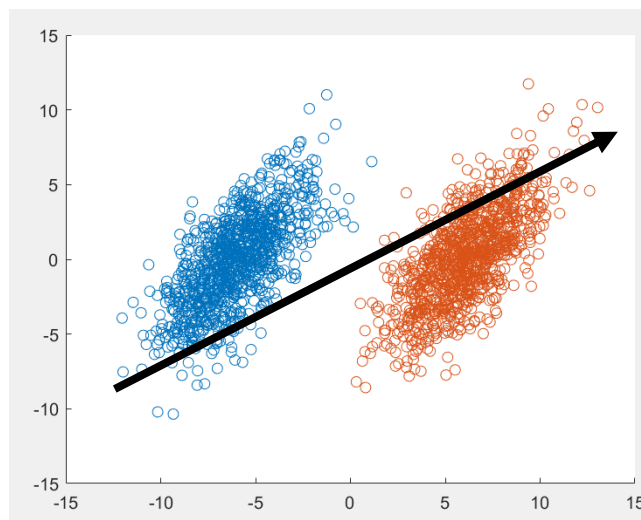
الف.

خطی را که PCA داده‌ها را روی آن تصویر می‌کند محاسبه و رسم کنید

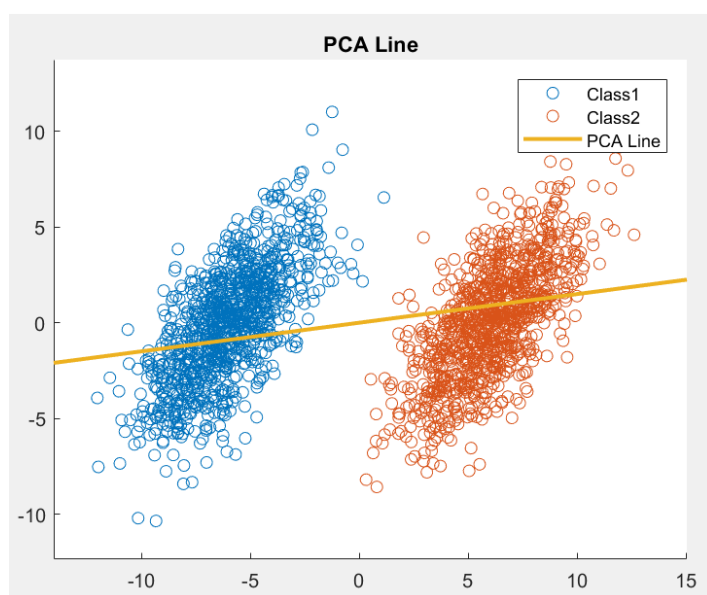
شکل زیر نقاط تولید شده را نشان می‌دهد.



با توجه به روش PCA که داده‌ها را در راستای بیشترین واریانس تصویر می‌کند، در نتیجه انتظار داریم که داده‌ها در راستای خطی که در شکل زیر مشخص شده‌اند، تصویر شوند زیرا این خط راستای بیشترین واریانس کل داده‌ها را نشان می‌دهد.



حال اگر مولفه های اصلی را بدست آوریم و داده ها با توجه به این مولفه ها تصویر کنیم، خط PCA به صورت زیر بدست می آید.



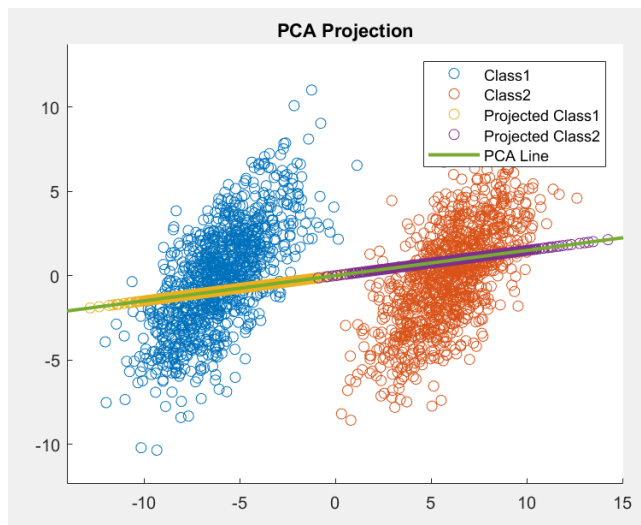
ب.

همه نقاط داده را روی خط PCA حاصله تصویر کرده و نتیجه را نمایش دهید.

برای این کار باید، داده ها را تحت تبدیل خطی بردار ویژه حاصل از PCA قرار دهیم که از رابطه زیر این کار انجام میشود

$$X_{projected} = X * V^T$$

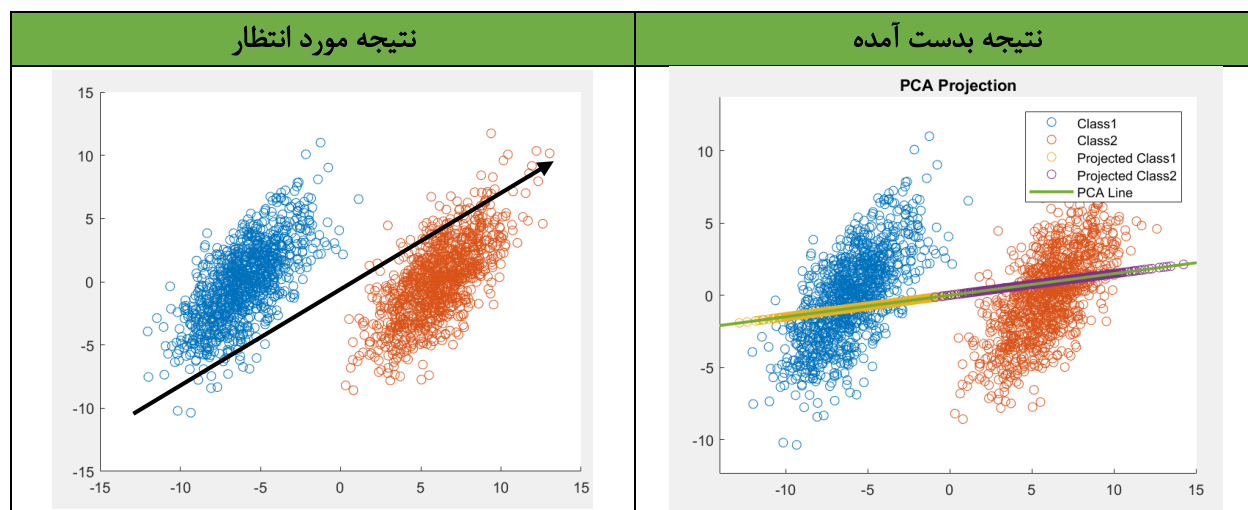
حال داده ها را تحت تبدیل PCA بر روی خطی که در قسمت قبل نشان داده شد، تصویر میکنیم.



ج.

آیا چیزی که مشاهده می کنید مورد انتظارتان بود؟ توضیح دهید.

با توجه به جدول زیر، میبینیم که خط PCA مطابق با پیش بینی ما بوده است. البته با کمی تغییر در جهت. ولی راستای بیشترین واریانس صحیح بوده است.

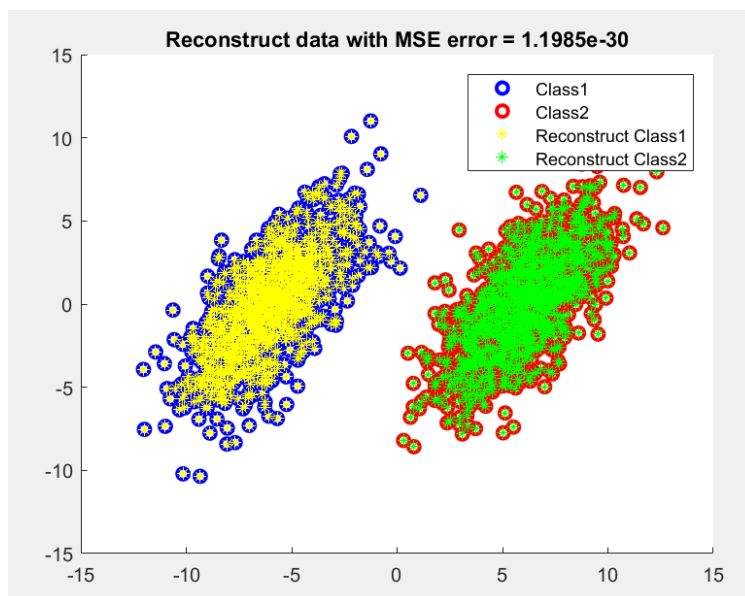


د.

نقاط داده را در فضای دوبعدی بازسازی کرده و خطای بازسازی را محاسبه کنید.

بعد از اینکه داده ها را به فضای جدید و تحت تبدیل PCA بردیم، با همان بردار ویژه میتوانیم مقادیر را به فضای قبلی برگردانیم و از رابطه زیر استفاده میکنیم

$$X_{reconstruct} = X_{projected} * V$$



در شکل بالا مبینیم که نقاط بازسازی شده که به صورت ستاره هستند (با رنگ زرد و سبز) دقیقا در داخل نقاط اصلی که به صورت دایره قرمز و آبی مشخص شده اند، قرار گرفته است. خطای MSE حاصل از بازسازی نیز در بالای شکل درج شده است که تقریبا برابر با صفر است.

کدهای قسمت اول

Run_me_1
<p>پاک کردن فضای متلب</p> <pre>%% Clear workspace clc; clear; close all; randn('seed',10);</pre> <p>ساخت داده‌گان</p> <pre>%% Make data mu1=[10 10]'; mu2=[22 10]';</pre>

```

sig1 = [4,4;4,9];
sig2 = sig1 ;
N = 1000;

% ساخت داده با دستور mvnrnd
x1 = mvnrnd(mu1,sig1,N);
x2=mvnrnd(mu2,sig2,N);

% ترکیب داده های دو کلاس و قرار دادن در یک متغیر
x = [x1;x2];

%% Normalize
% نرمال سازی داده ها
x = x - mean(x);

%% Do PCA
% اعمال pca روی داده ها با دستور pca
[coeff,score,latent,tsquared,explained,mu] = pca(x,'Centered',false);

% خروجی اول = ضرایب تبدیل
% خروجی دوم = داده های تصویر شده

%% Find PCA line
% شیب خط pca با توجه به مقادیر بردار ویژه
m = coeff(2)/coeff(1);

% تابع با شیب بدست آمده
fun=@(x)m*x;

% تصویر کردن داده ها روی خط pca
%% Project data on
pc_vector = coeff(:,1);
x_projection = score(:,1);
x_projection1=(x_projection(1:N)./(pc_vector'*pc_vector))*pc_vector';
x_projection2=(x_projection(N+1:2*N)./(pc_vector'*pc_vector))*pc_vector';

% نمایش خط pca
%% Plot PCA line (part a)
figure;
scatter(x(1:N,1), x(1:N,2));
hold on;
scatter(x(N+1:end,1), x(N+1:end,2));
fplot(fun,[min(x(:,1))-2,max(x(:,1))+2],'LineWidth',2);
title('PCA Line');
legend('Class1','Class2','PCA Line');
ylim([min(x(:,2))-2,max(x(:,2))+2]);
xlim([min(x(:,1))-2,max(x(:,1))+2]);

% نمایش نقاط تصویر شده
%% Plot projected data (part b)
figure;
scatter(x(1:N,1), x(1:N,2));
hold on;
scatter(x(N+1:end,1), x(N+1:end,2));

```

```

scatter(x_projection1(:,1), x_projection1(:,2));
scatter(x_projection2(:,1), x_projection2(:,2));
fplot(fun, [min(x(:,1))-2,max(x(:,1))+2], 'LineWidth',2);
title('PCA Projection');
legend('Class1','Class2','Projected Class1','Projected Class2','PCA Line');
ylim([min(x(:,2))-2,max(x(:,2))+2]);
xlim([min(x(:,1))-2,max(x(:,1))+2]);

```

قسمت بازسازی داده ها

```
%% Reconstruct data using PCA part(d)
```

ضرب نقاط تصویر شده در بردار ویژه

```
rec = (score)*coeff';
```

محاسبه خطای بازسازی

```
Mse_Error = mse(rec,x);
```

نمایش نقاط اصلی و نقاط بازسازی شده

```

figure;
scatter(x(1:N,1), x(1:N,2), 'bo', 'linewidth',2);
hold on;
scatter(x(N+1:end,1), x(N+1:end,2), 'ro', 'linewidth',2);
scatter(rec(1:N,1), rec(1:N,2), 'y*');
scatter(rec(N+1:end,1), rec(N+1:end,2), 'g*');
legend('Class1','Class2','Reconstruct Class1','Reconstruct Class2');
title(['Reconstruct data with MSE error = ' num2str(Mse_Error)]);

```