

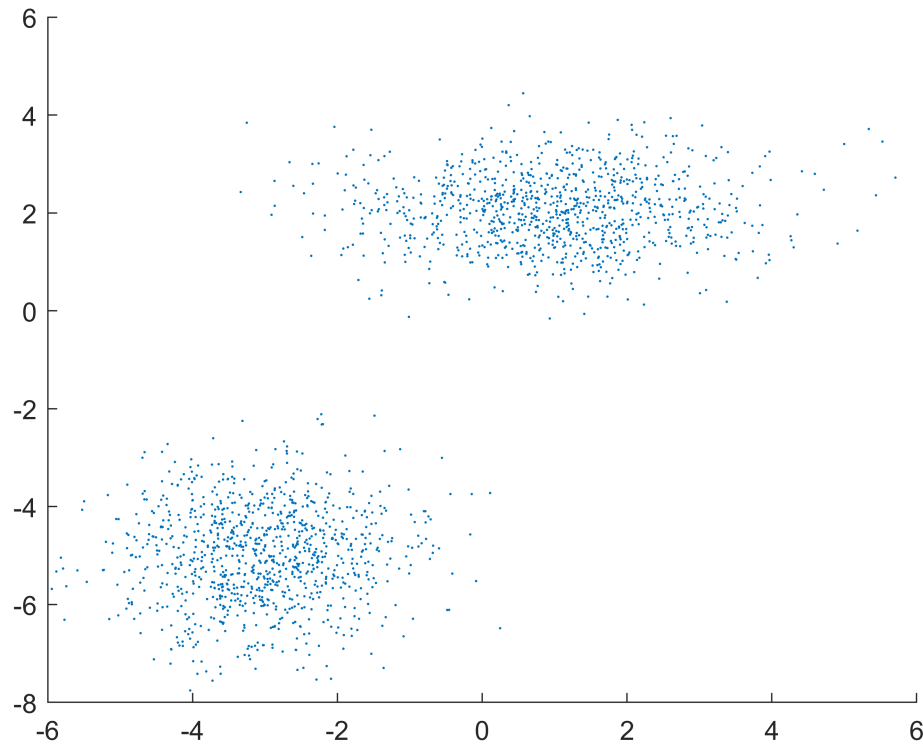
Fit a Gaussian Mixture Model to Data

This example shows how to simulate data from a multivariate normal distribution, and then fit a Gaussian mixture model (GMM) to the data using [fitgmdist](#). To create a known, or fully specified, GMM object, see [docid:stats_ug.bus2n9t](#).

`fitgmdist` requires a matrix of data (X) and the number of components in the GMM (k). To create a useful GMM, you must choose k carefully. Too few components fails to model the data accurately (i.e., underfitting to the data). Too many components leads to an over-fit model with singular covariance matrices.

Simulate data from a mixture of two bivariate Gaussian distributions using [mvnrnd](#).

```
MU1 = [1 2];  
SIGMA1 = [2 0; 0 .5];  
MU2 = [-3 -5];  
SIGMA2 = [1 0; 0 1];  
  
rng(1); % For reproducibility  
X = [mvnrnd(MU1,SIGMA1,1000);  
     mvnrnd(MU2,SIGMA2,1000)];  
  
figure;  
scatter(X(:,1),X(:,2),10,'.')
```

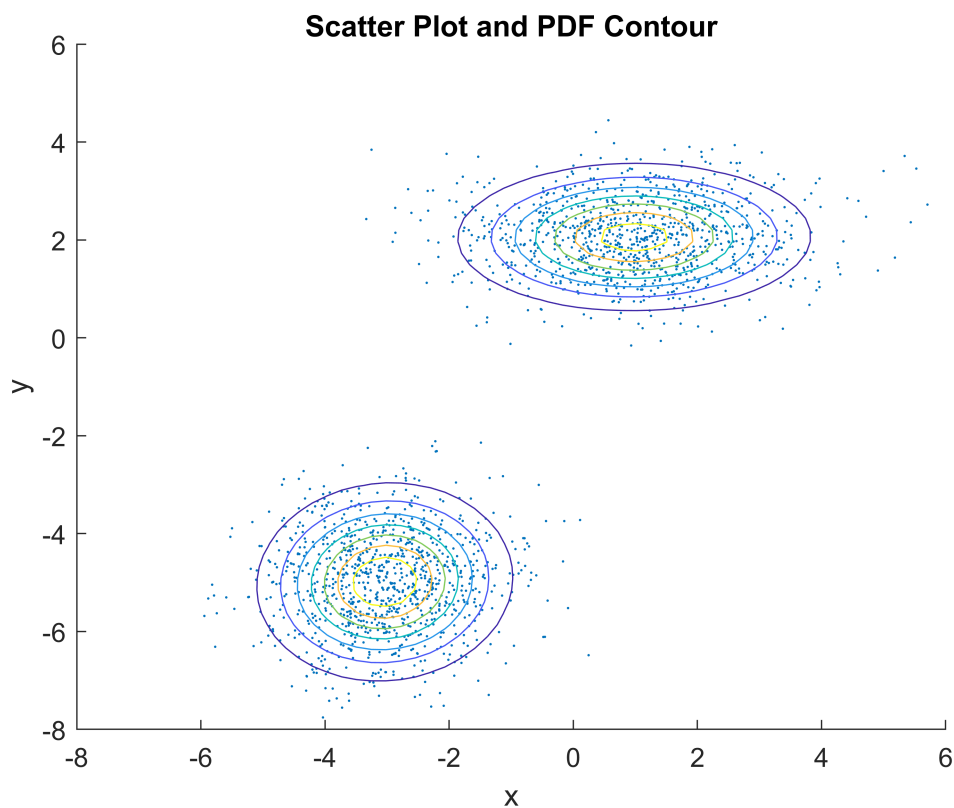


Fit a two-component GMM. Plot the pdf of the fitted GMM.

```
options = statset('Display','final');  
gm = fitgmdist(X,2,'Options',options);
```

5 iterations, log-likelihood = -7105.71

```
gmPDF = @(x,y)pdf(gm,[x y]);  
  
hold on  
h = ezcontour(gmPDF,[-8 6],[-8 6]);  
title('Scatter Plot and PDF Contour')  
hold off
```



Display the estimates for mu, sigma, and mixture proportions

```
ComponentMeans = gm.mu
```

```
ComponentMeans =  
-3.0377 -4.9859  
 0.9812  2.0563
```

```
ComponentCovariances = gm.Sigma
```

```
ComponentCovariances =
```

```
ComponentCovariances(:,:,1) =
```

```
1.0132    0.0482
0.0482    0.9796
```

```
ComponentCovariances(:,:,2) =
```

```
1.9919    0.0127
0.0127    0.5533
```

```
MixtureProportions = gm.PComponents
```

```
MixtureProportions =
0.5000    0.5000
```

Fit four models to the data, each with an increasing number of components.

```
AIC = zeros(1,4);
gm = cell(1,4);
for k = 1:4
    gm{k} = fitgmdist(X,k);
    AIC(k) = gm{k}.AIC;
end
```

Display the number of components that minimizes the AIC.

```
[minAIC,numComponents] = min(AIC);
numComponents
```

```
numComponents = 2
```

The two-component model minimizes the AIC.

Display the two-component GMM.

```
gm2 = gm{numComponents}
```

```
gm2 =
```

```
Gaussian mixture distribution with 2 components in 2 dimensions
```

```
Component 1:
```

```
Mixing proportion: 0.500000
```

```
Mean:    -3.0377    -4.9859
```

```
Component 2:
```

```
Mixing proportion: 0.500000
```

```
Mean:     0.9812     2.0563
```

Both the Akaike and Bayes information are negative log-likelihoods for the data with penalty terms for the number of estimated parameters. You can use them to determine an appropriate number of components for a model when the number of components is unspecified.