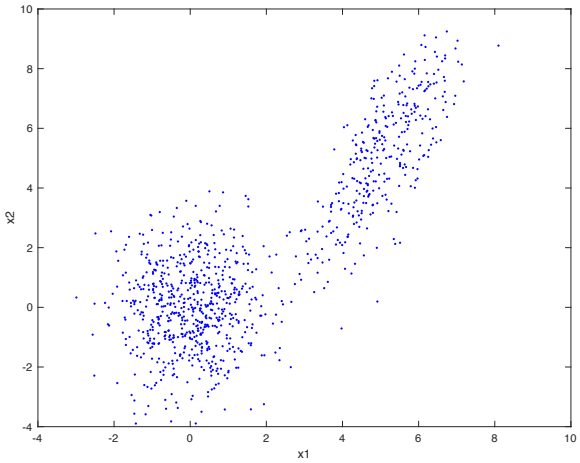
Unsupervised Learning Exercise

You have been given some unlabelled 2-D data in the file data.mat and you would like to analyse this data.



You believe that there are two clusters in the data

1. Perform K-means clustering on the data to find two cluster centroids
2. You are given a new data sample for $x_1 = 3$ and $x_2 = 3$. To which cluster does this new data sample belong?

You think that a Gaussian mixture model with two components is appropriate for modelling the data

3. Use the EM algorithm to find the model parameters
4. For the new data sample for $x_1 = 3$ and $x_2 = 3$. What are the probabilities that the sample belongs to each Gaussian component?