First of all, from the shape of the data, it seems that it consists of 3 or 4 clusters, so I used 4 mixtures first. , for finding the exact number of mixtures one should try different number of mixtures and calculate the log-likelihood with respect to each mixture. We are maximizing the log likelihood, so number of mixtures which maximizes the log likelihood is the best one for out data.

Here is the plot for data on its features x1 and x2:

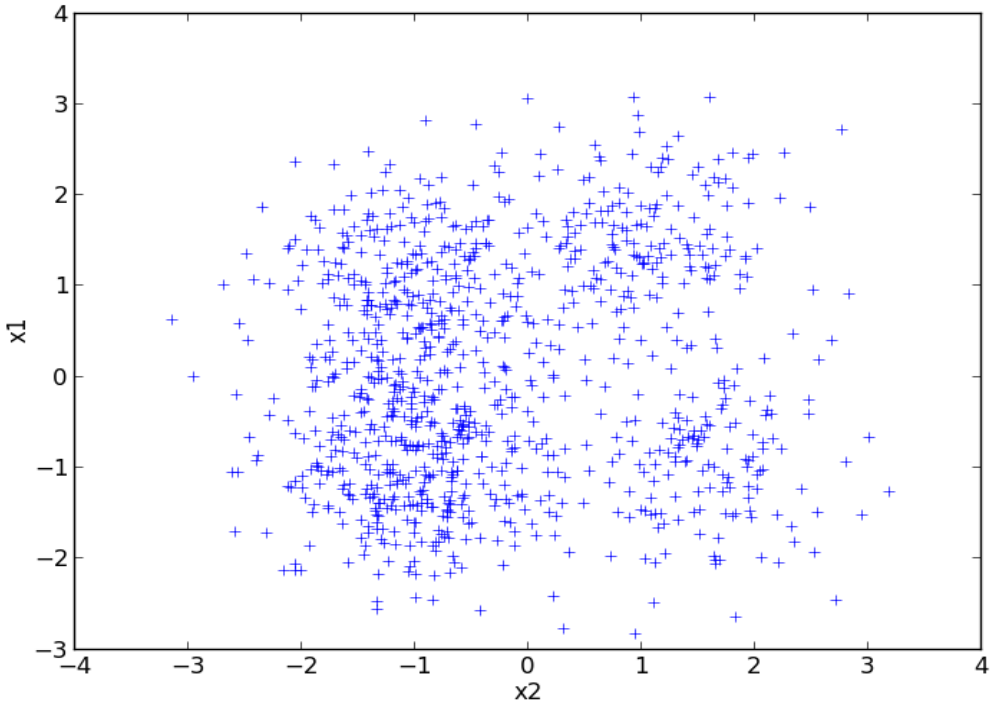

Figure 1-plot data in x1 , x2

Figure below, shows likelihood vs iterations when the number of mixtures is equal to 4:

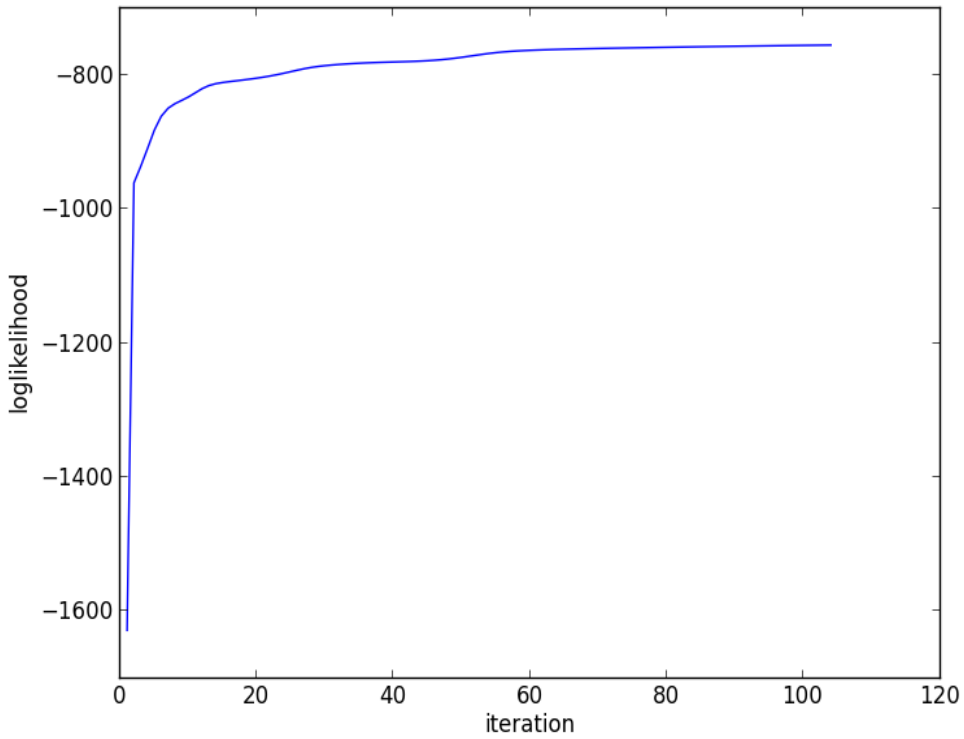

Figure 2-log likelihood vs iterations when number of mixtures equals to 4

Figure3 shows the log likelihood vs iteration when the number of mixtures is 2 :

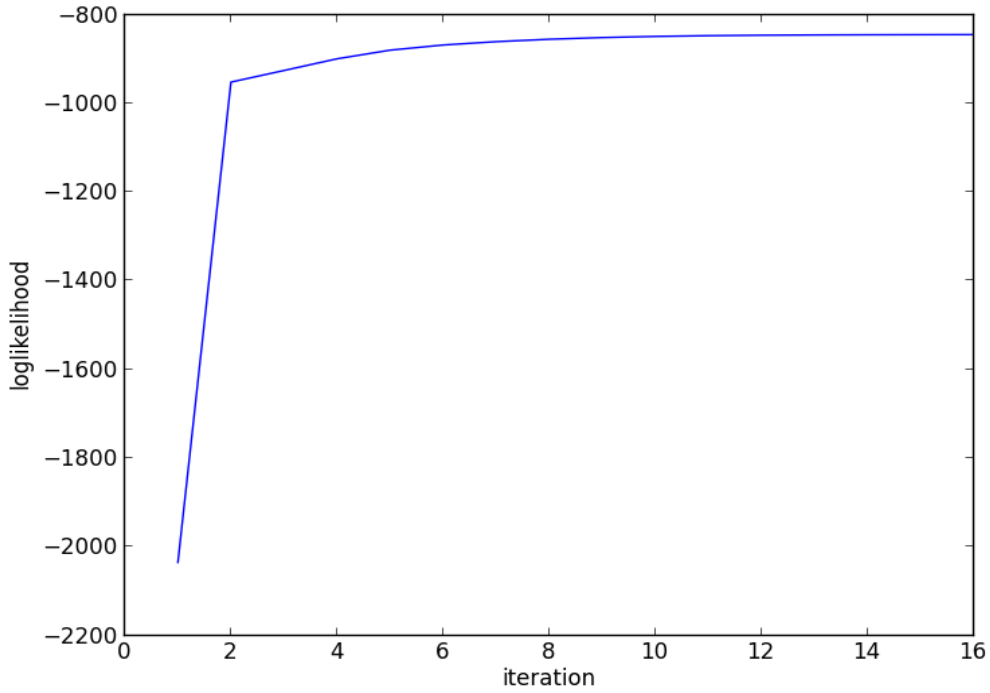

Figure 3-log likelihood vs iterations when the number of clusters equals to 2

Figure 4 shows the log likelihood vs iterations when the number of mixtures is equal to 3 .

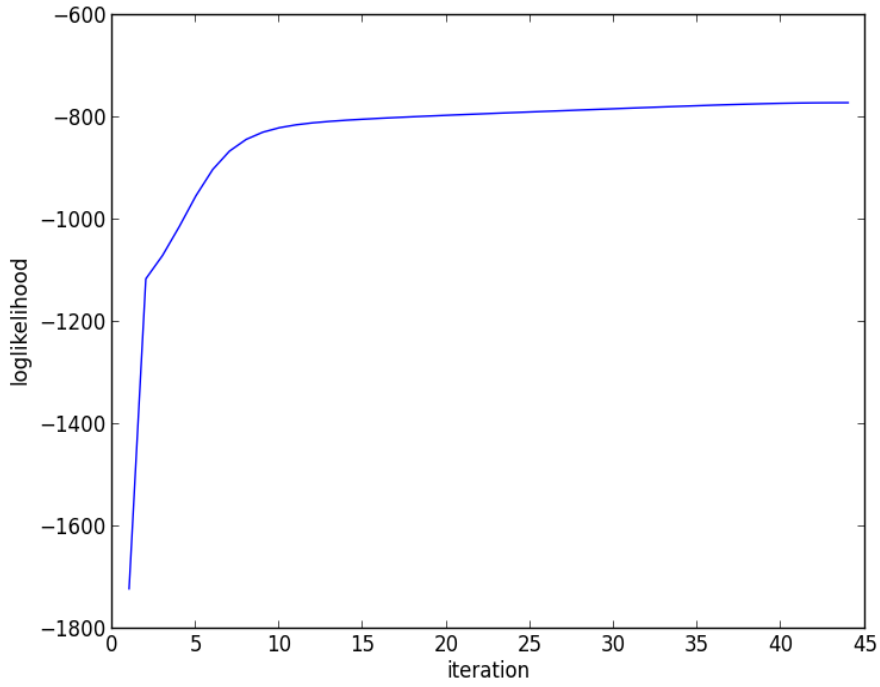

Figure 4--log likelihood vs iterations when the number of clusters equals to 3

When the number of mixtures is equal to 4, th log likelihood roughly converges to -755, and when its 3 it converges to roughly -772 . but when the number of mixtures is equal to 2 it converges to a higher number. So I will choose 4, as the best number of mixtures, however 3 is also good.