

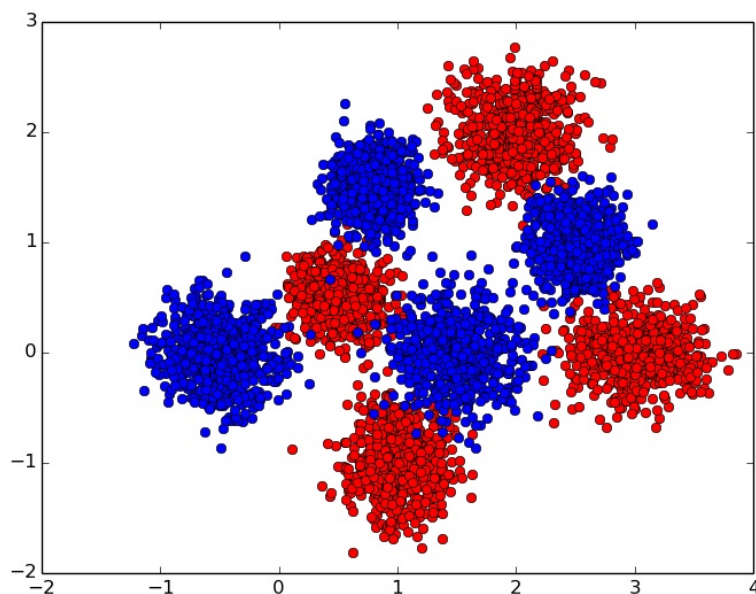
EM Algorithm for GMM

Qi Liu

November 4, 2016

1 Data analysis

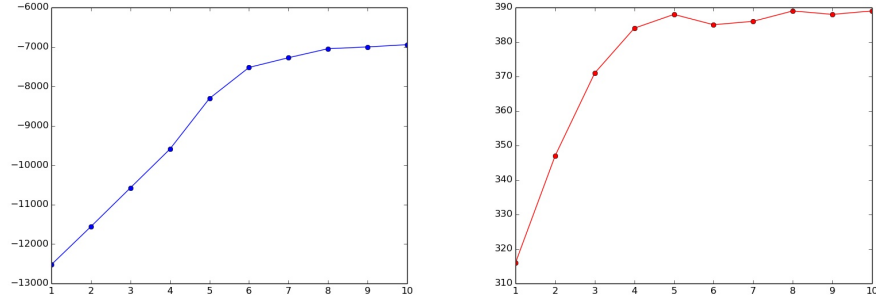
There are 4800 samples in the training data which are divided into 2 groups. The details can be found in the figure below. It is can be observed that there are 4 Gaussian component for each group.



2 EM training

The GMM model for each group has been trained by 10 iterations. The initial values of c are the uniform distribution. The initial value of μ are $(-1, 0)$, $(1, 0)$, $(0, -1)$ and $(0, 1)$. The initial values of σ are all the identity matrix. The log likelihood and the correct predict number of developing set have been calculated

for each iteration. The left figure below correspond to the log likelihood and the right one is for correct predict number of developing set. It is can be found that EM algorithm converge very quickly due to the predict accuracy of developing set become very high at the 4th iteration.



3 Test set evaluation

The evaluation result for the test set can be found in the file *result/result.txt*. The figure about grouping of test set is shown below.

