Multiple clusters

*Sangyeon Kim

1. Aim

The goal of the experiments is to check if QGMM can be trained well given the data set that has multiple clusters more than 2 clusters.

2. Dataset

In this research, we generated a data set "multiple_5.csv" that has 5 clusters.

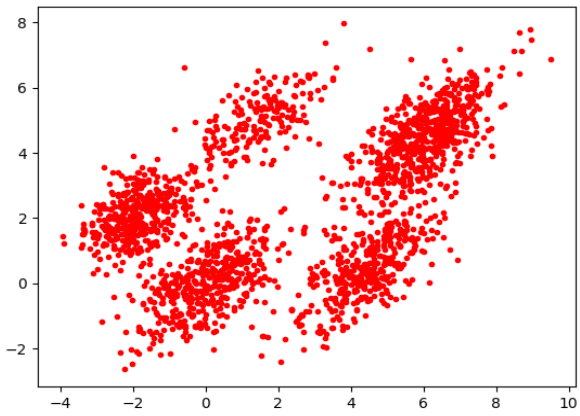


Figure 1. multiple_5 data set that has 5 clusters and 2 dimensionalities.

To generate the data set, we used 5 Gaussian distributions using mlpack library, fast flexible C++ machine learning library [1]. The weights are [0.1, 0.2, 0.2, 0.3, 0.2], the means are [1.5, 5], [4.5, 0.5], [0, 0], [6, 4.5], and [-2, 2], and the covariances are

$$C_1 = C_3 = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 0.89 \end{pmatrix}$$

$$C_2 = C_4 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1.01 \end{pmatrix}$$

$$C_5 = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}$$

Using the parameters, we generated 2,000 observations.

3. Experiments

In these experiments, the maximum iteration of training process is 50,000 and the tolerance is 1e-3. we used Adam optimizer with 0.001 learning rate (η). Also, as the training parameters, all the initial alphas (α_k) are 0.5, the means are randomly or arbitrarily generated, and all the initial covariances (C_k) are that

$$C_k = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

3.1 Training results

In this experiment, we'll check the training accuracy of QGMM in the "multiple_5" data set. In Appendix A, there are 1 and 2 experiments and the difference between them is the initial phi and the difference between each experiment (a) and (b) is the initial lambda.

Actually, we wanted to set the initial phi to 90 equally, however, because of the property of the subtraction, we couldn't, so we set the value to make sure that the adjacent clusters' phi is 90.

Through the experiments, we figured out that when the initial phi is 0, the constraint works better than initial phi 90. Besides, although the higher lambda has an effect on the constraint, the objective function has been trembled more.

In Appendix B, the difference with Appendix is the initial means. The initial means were distributed better than Appendix A.

Likewise, when the initial phi is 0, the constraint works better than initial phi 90. Also, when the initial phi 0, the final clusters were independent while the final clusters were mixed a bit in the case of the initial phi 90.

As we saw in the previous researches like 04. Phi modeling and 05. Mixture case, when

*This research was performed during Google Summer of Code 2019 for mlpack mentored by Sumedh Ghaisas.

the phi is 0, the clusters tend to be independent more and more, and the initial phi almost wasn't changed.

4. Conclusions

From the experiments, we figured out that when the phi difference between the clusters is close to 0, the clusters tend to be independent, not mixed, when the phi difference is close to 90, the clusters tend to be mixed.

Besides, the initial phi is almost not changed, therefore, the initial phi matters on the training performance side.

However, from the experiments, we also figured out that there are some drawbacks on the multiple clusters case that we found through the experiments.

First, it is hard to set the equal initial phi difference, except for zero. Because $\phi_{l,k}$ between the two clusters is calculated from the subtraction $\phi_k - \phi_l$, it's hard to set the initial ϕ_k to make $\phi_{l,k}$ to have a equal value for all the clusters, except for the subtraction to zero

Second, it take more time to calculate the probability because the time complexity of general GMM is $O(n)$, while QGMM is $O(n^2)$.

Therefore, the more clusters are, the more time it take.

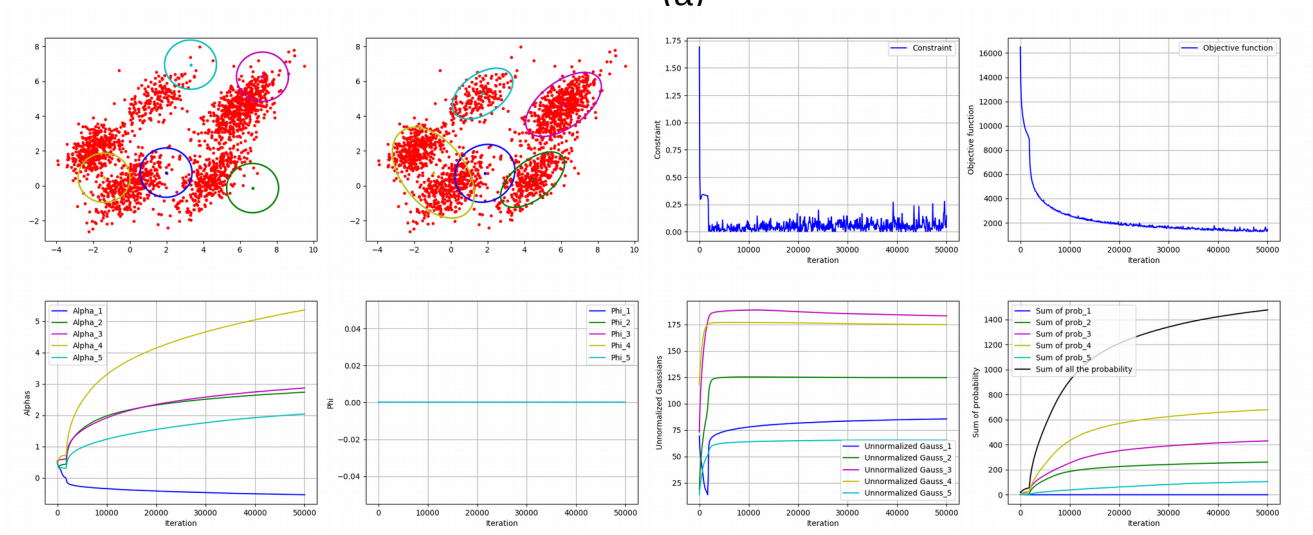
Consequently, QGMM can be trained well in multiple clusters cases when we find the hyper-parameters properly, for example, the initial means, lambda, and phi.

References

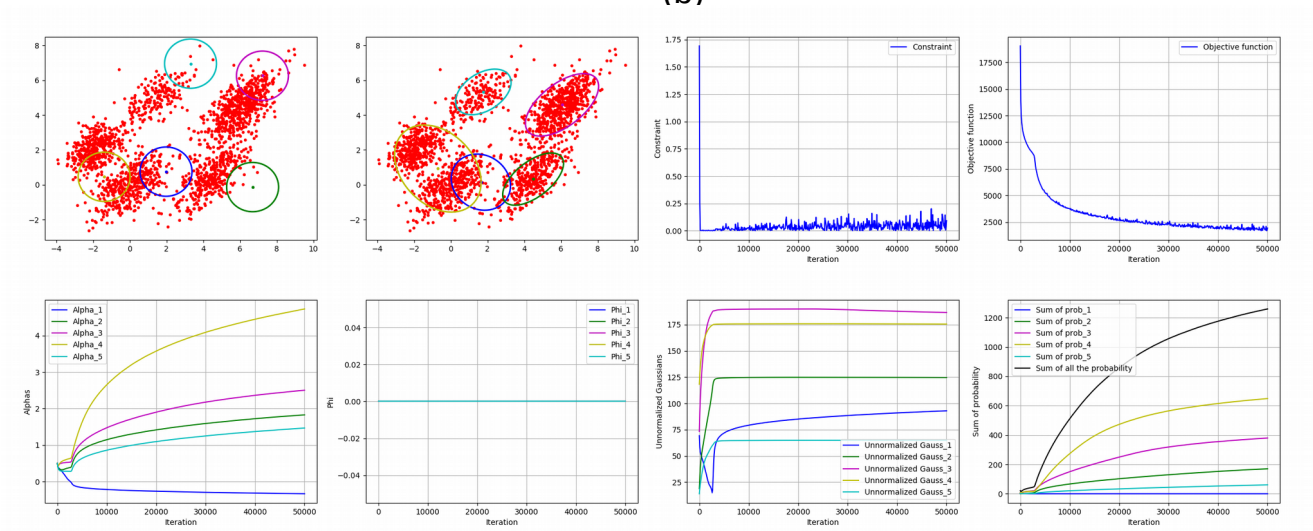
[1] Curtin, R.; Edel, M.; Lozhnikov, M.; Mentekidis, Y.; Ghaisas, S.; Zhang, S. mlpack 3: A fast, flexible machine learning library. J. Open Source Softw. 2018, 3, 726 [Ref]

Appendix A-1. Training results

(a)

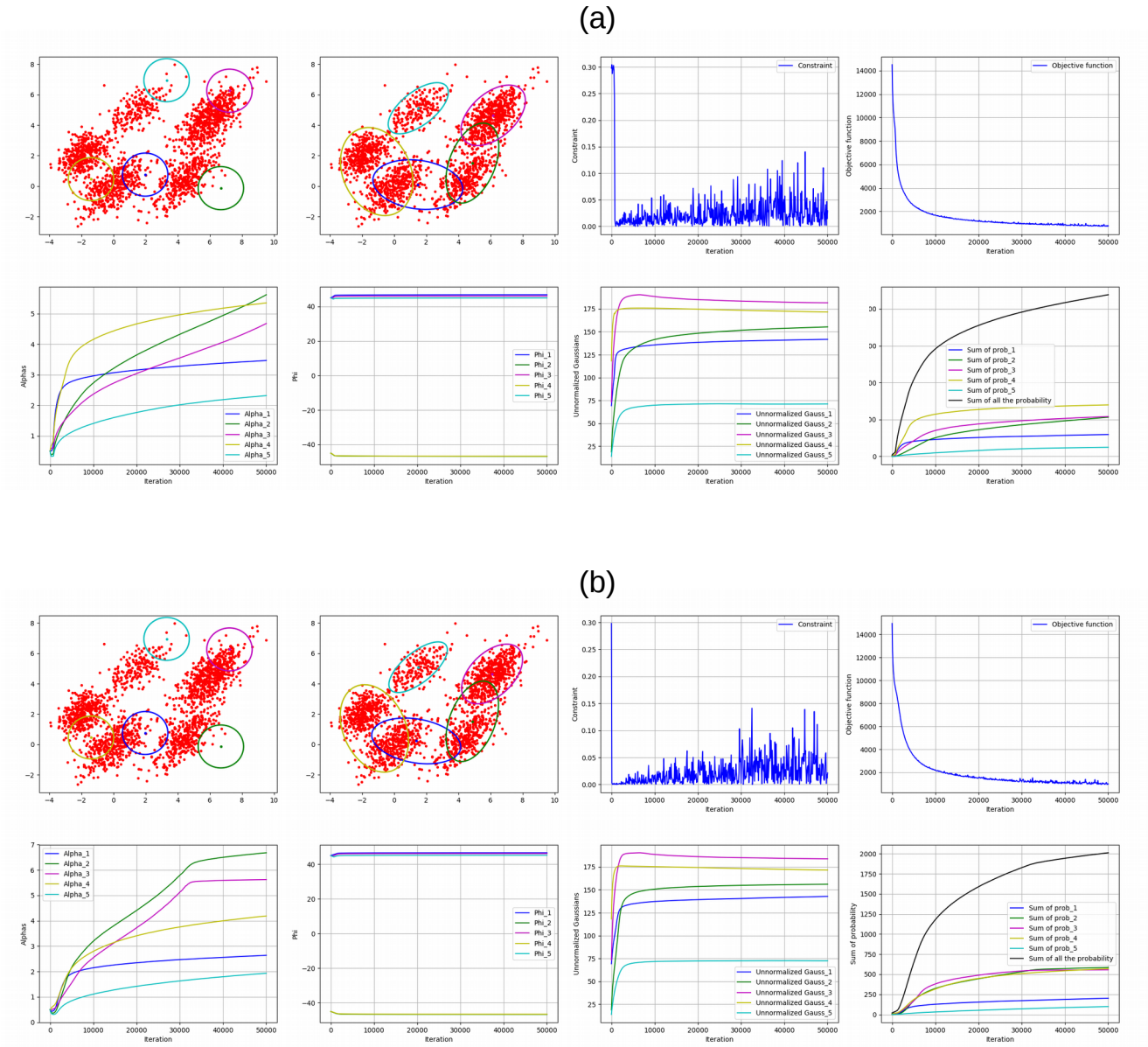


(b)



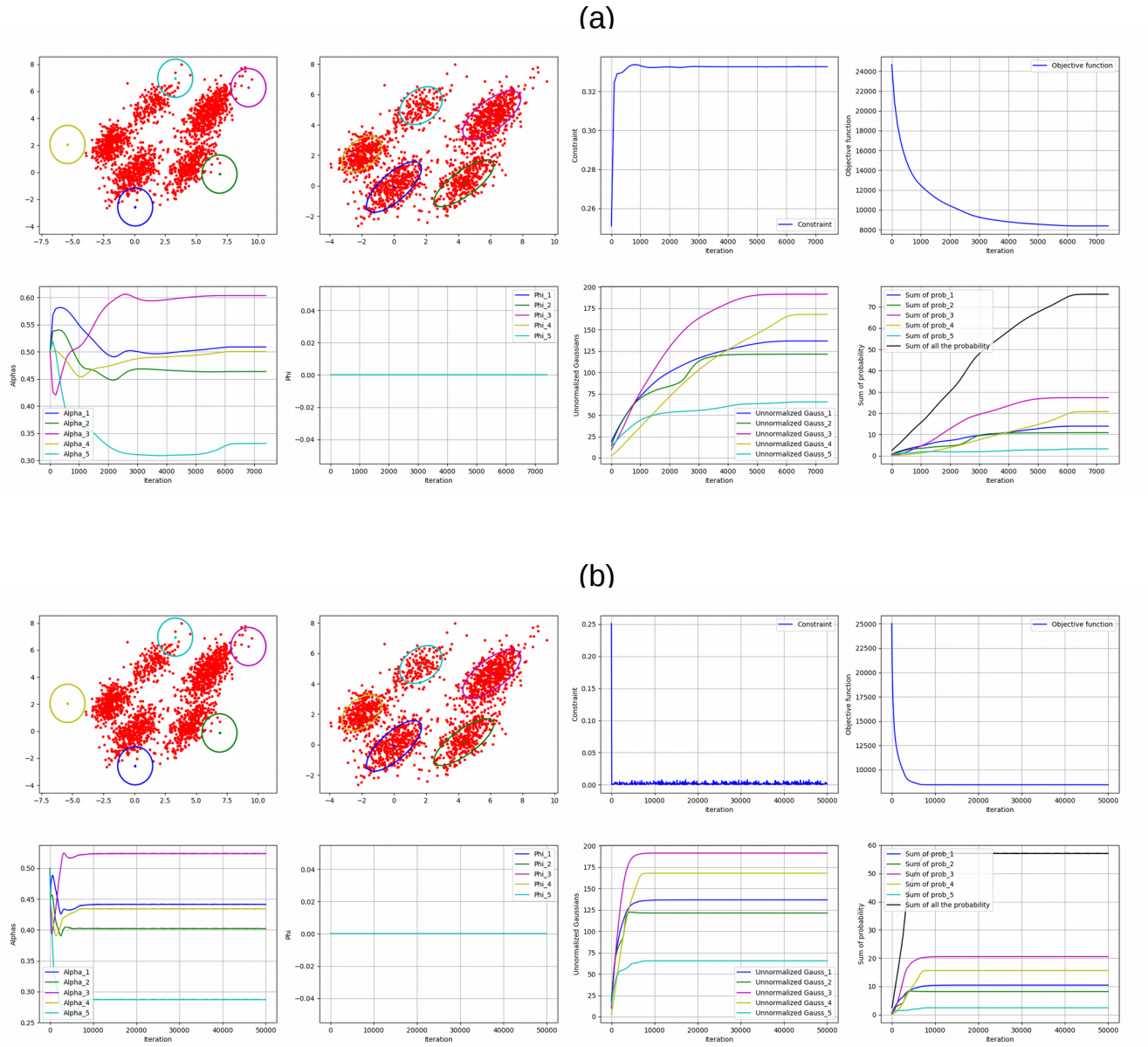
Appendix A-1. Training results. The sequence of the clusters are blue, green, magenta, yellow, and cyan. The means are [1.972979456933269, 0.7560544338372392], [6.686047548070453, -0.1307013280230782], [7.221786450205375, 6.262448613633105], [-1.3977125108429398, 0.4590524145870889], and [3.307847427018068, 6.954851796948616]. The phis are [0, 0, 0, 0, 0]. (a) Lambda 1500, (b) Lambda 3000

Appendix A-2. Training results



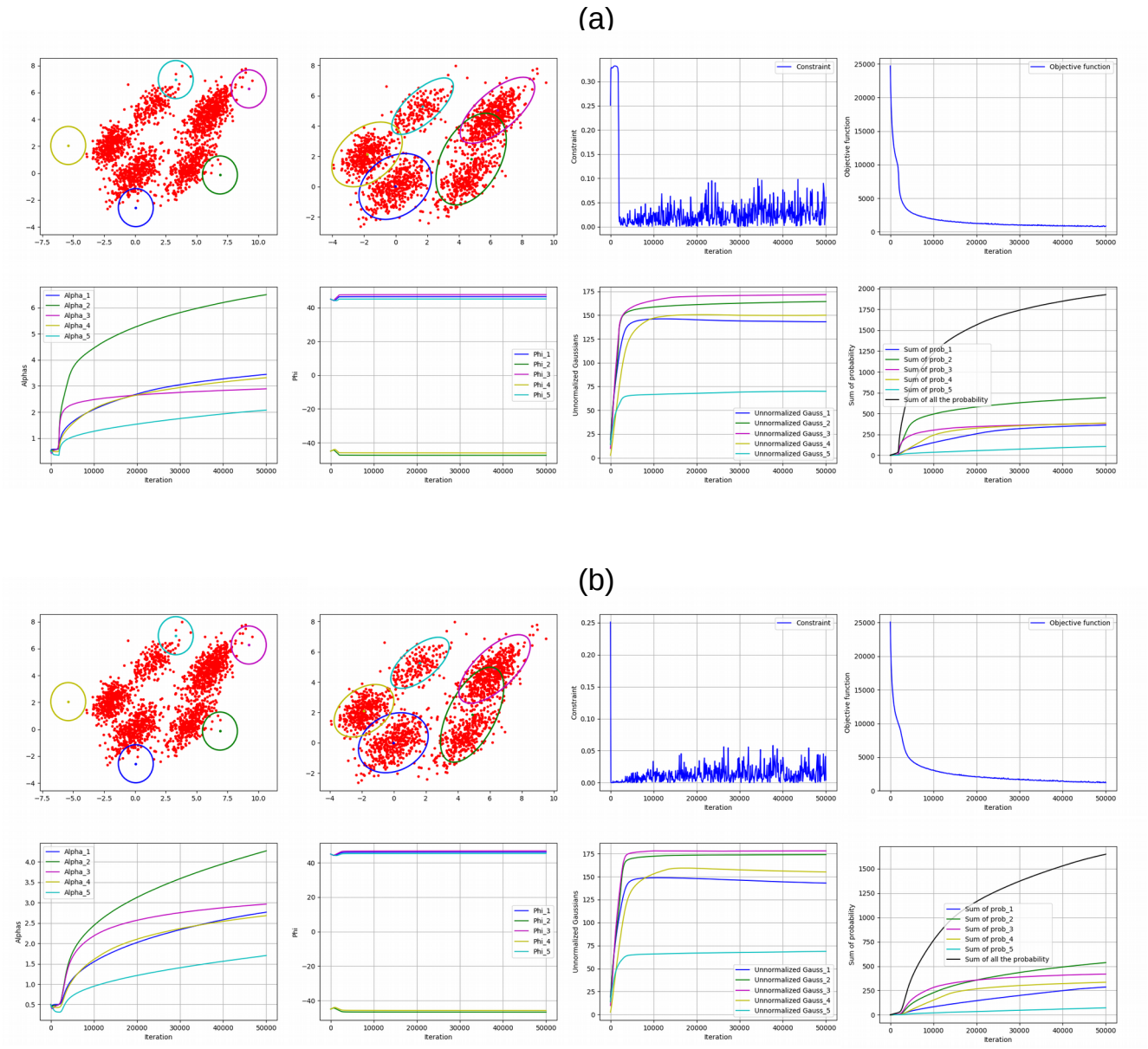
Appendix A-2. Training results. The sequence of the clusters are blue, green, magenta, yellow, and cyan. The means are [1.972979456933269, 0.7560544338372392], [6.686047548070453, -0.1307013280230782], [7.221786450205375, 6.262448613633105], [-1.3977125108429398, 0.4590524145870889], and [3.307847427018068, 6.954851796948616]. The phis are [45, -45, 45, -45, 45]. (a) Lambda 1500, (b) Lambda 3000

Appendix B-1. Training results



Appendix A-1. Training results. The sequence of the clusters are blue, green, magenta, yellow, and cyan. The means are [0.072979456933269, -2.5560544338372392], [6.886047548070453, -0.1307013280230782], [9.221786450205375, 6.262448613633105], [-5.3977125108429398, 2.0590524145870889], and [3.307847427018068, 6.954851796948616]. The phis are [0, 0, 0, 0, 0]. (a) Lambda 1500, (b) Lambda 3000

Appendix B-2. Training results



Appendix A-2. Training results. The sequence of the clusters are blue, green, magenta, yellow, and cyan. The means are [0.072979456933269, -2.5560544338372392], [6.886047548070453, -0.1307013280230782], [9.221786450205375, 6.262448613633105], [-5.3977125108429398, 2.0590524145870889], and [3.307847427018068, 6.954851796948616]. The phis are [45, -45, 45, -45, 45]. (a) Lambda 1500, (b) Lambda 3000