

Mixed clusters

*Sangyeon Kim

1. Aim

The goal of the experiments is to figure out that the phi variation in the mixture data set.

The observations in the previous data set, "faithful.csv" [1] are somewhat distributed in some distance. Thus, in this research, we'll generate some data set whose observations are mixed, and look into the phi variation.

2. Dataset

In this research, we created 'gmm' data set using mlpack library [2], mlpack is a fast, flexible machine learning library. The weights are [0.4, 0.6], the means are [1.5, 5], [4.5, 0.5], and the covariances are

$$C_1 = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 0.89 \end{pmatrix}, C_2 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1.01 \end{pmatrix}$$

From the parameters, we generated 1,000 observations.

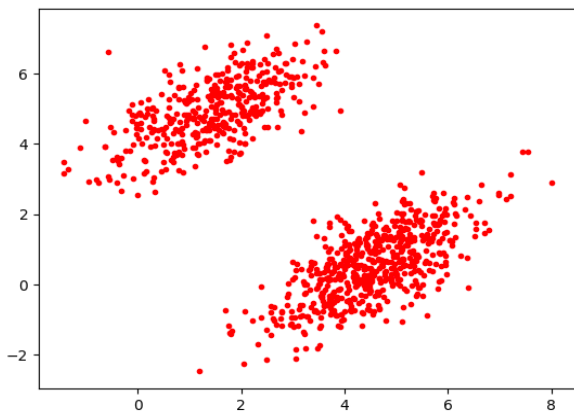


Figure 1. gmm data set that has 2 dimensionalities. There are 2 bunches of observations on the top-left and bottom-right.

To create mixture effects on the data set, we'll draw a circle. Besides, by changing a radius, we'll generate some different data sets for better view on phi variation.

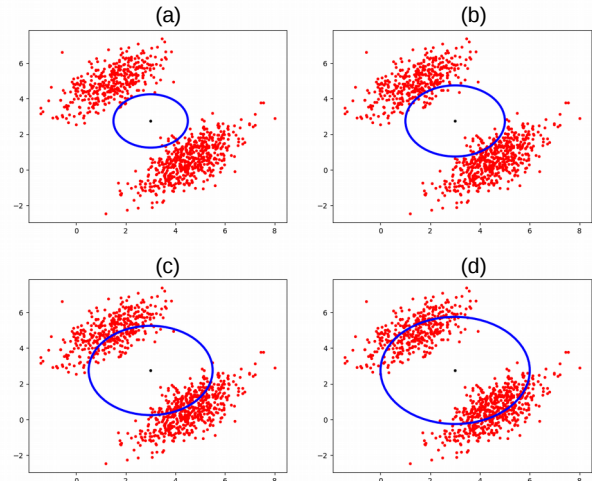


Figure 2. gmm data set with different circles. (a) Radius 1.5. (b) Radius 2. (c) Radius 2.5. (d) Radius 3.

After setting the circle in the middle of the piles of the observations, we'll generate $500 * 0.67$ observations within the circle to make an mixture effect between the two clusters.

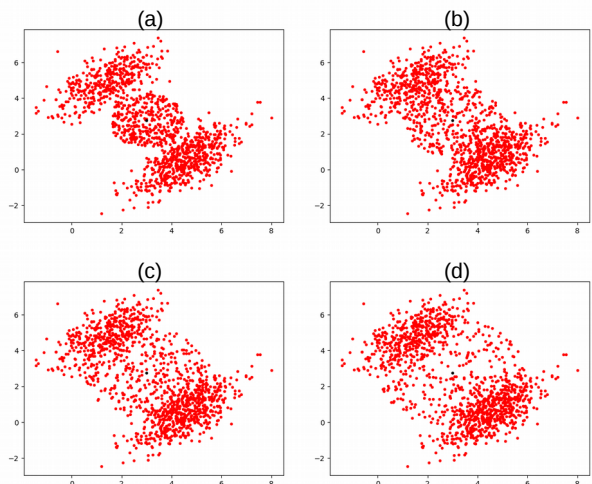


Figure 3. Observations generation with different circles. (a) Radius 1.5. (b) Radius 2. (c) Radius 2.5. (d) Radius 3.

In data set (a), it seems to be 3 clusters case, however, we'll see how well QGMM is trained given the observations.

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

3. Experiments

we performed two experiments to find the possible improvements and good initial values in the training process for QGMM in 2-classes constraint environment.

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, the initial alphas (α_k) are $\alpha_1 = 0.5$, $\alpha_2 = 0.5$, the means are randomly generated, and the covariances (C_k) are that

$$C_1 = C_2 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

3.1 Phi variation

In this experiment, we manipulated the data set to make it more coupled, so we'll observe the phi variation as it is trained. In the existing data set, the distribution of the observations seem s to be less coupled. Thus, if we change it to coupled one more and more, we would see another property of phi. The training results can be seen in Appendix A, B, and C.

When we tested the test cases in Appendix A, we found the constraint was too jagged, so we increased the lambda from 1,500 to 10,000. and the result can be seen in Appendix C. As a result, the training processes in Appendix C were more stable than Appendix A.

From the results, we could find that the phi wasn't changed when we set the initial phi 0 or 90 in the mixed observations.

Although we couldn't observe the phi variation well in experiments A, B, and C, we figured out the effect of the initial value of phi in a mixed observations case, in the mixed observations cases, when the initial phi is 0, the clusters tend to be independent, not mixed, and when the initial phi is 90, the clusters tend to be mixed.

Conclusions

In this research, we figured out that the phi isn't changed well from the initial value in the mixed observations cases.

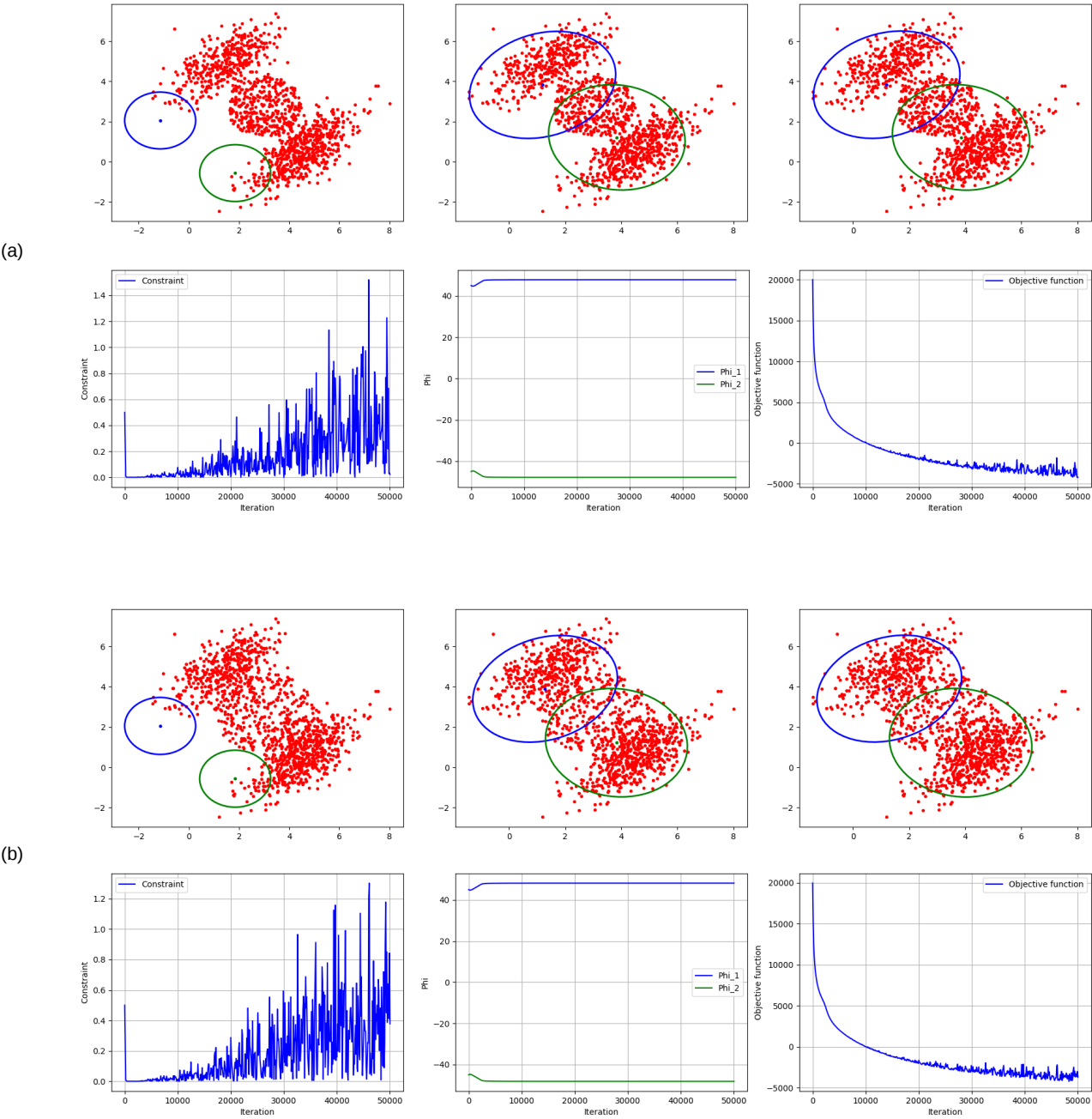
Also, in the mixed observations cases, when the value of phi is 0, the clusters tend to be independent, not mixed, and when the value of phi is 90, the clusters tend to be mixed.

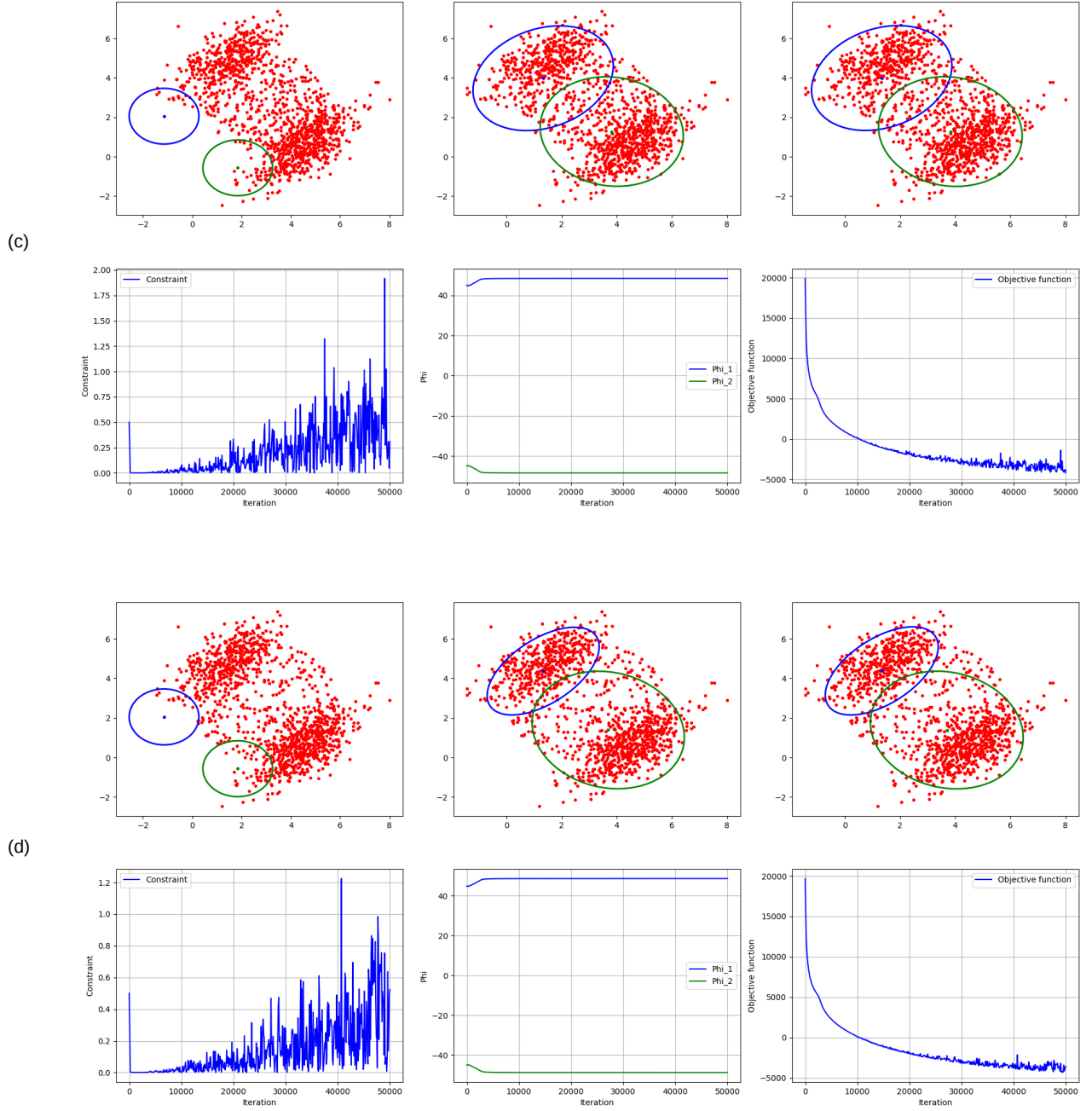
Therefore, we figured out that according to the initial observations distribution, setting the proper initial phi is important on the view of the training performance.

References

- [1] <https://www.kaggle.com/janithwanni/old-faithful>
- [2] 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]

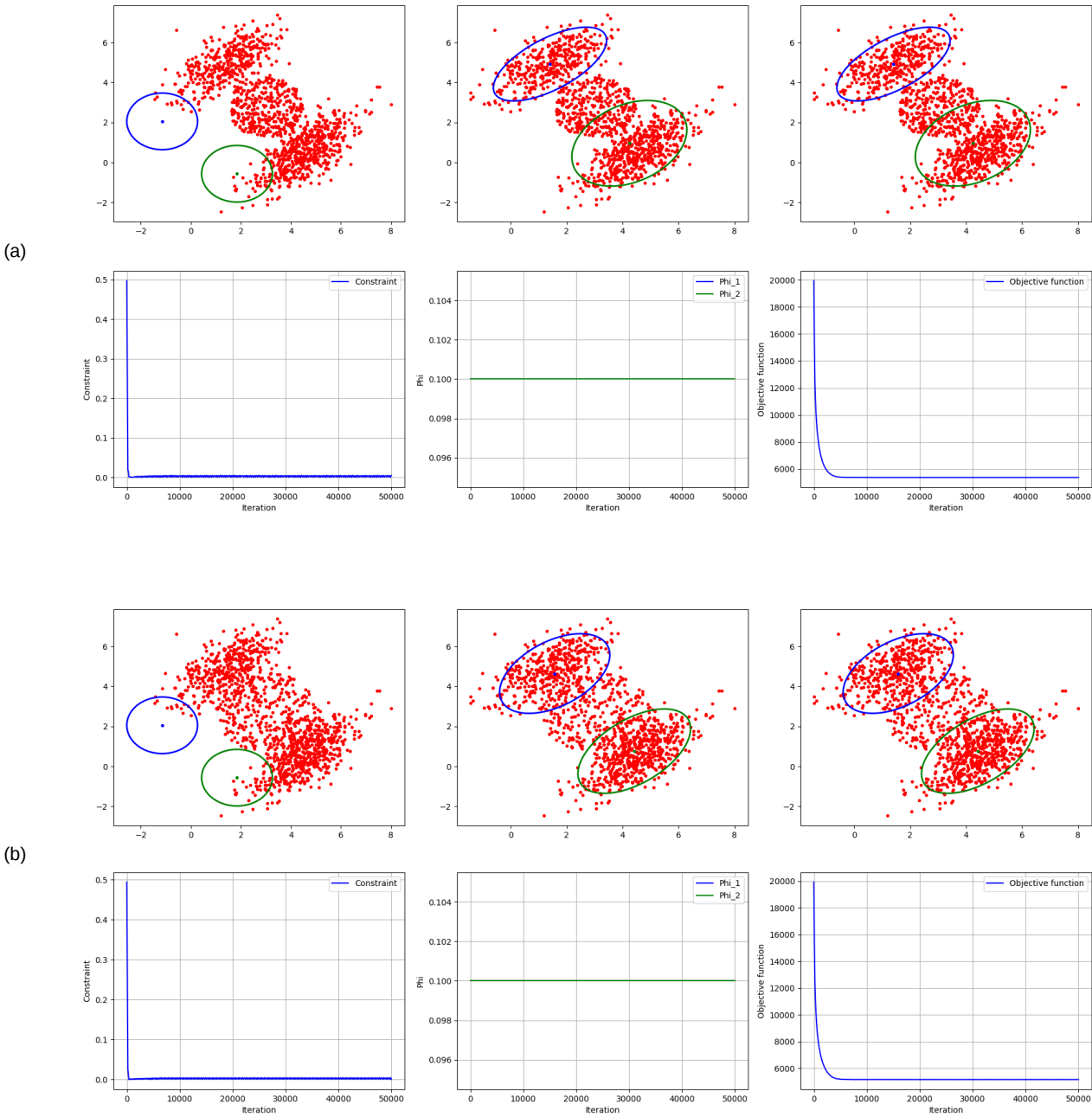
Appendix A-1. Training results

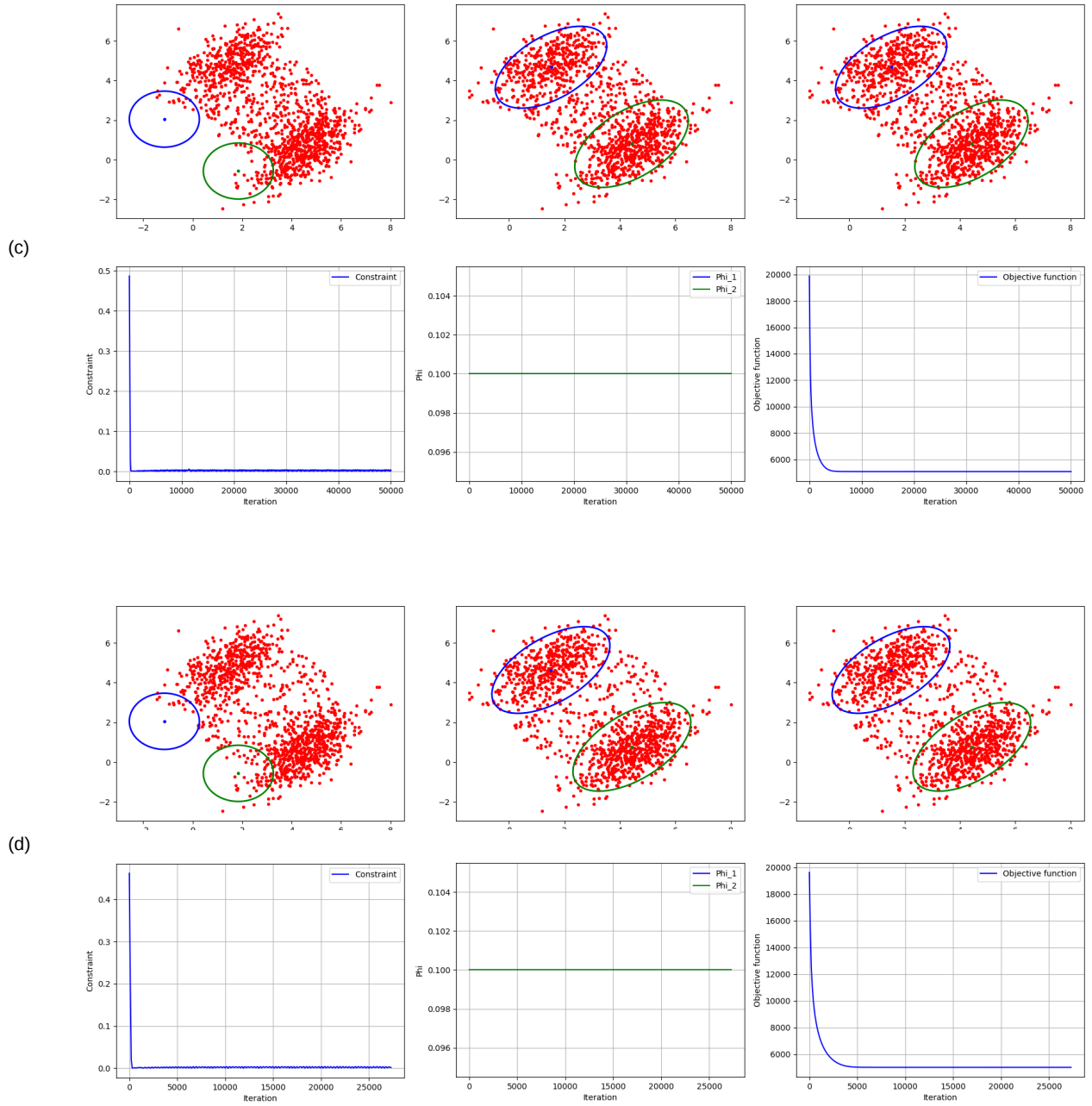




Appendix A-1. Training results with the initial ϕ 90 (45 - (- 45)), λ 1,500, and the initial means are [-1.1431749105000186, 2.0529438865919479], [1.846096025789632, -0.5656381542446731]. (a) Radius 1.5 and the final ϕ is [47.822952, -47.822952]. (b) Radius 2 and the final ϕ is [48.167175, -48.167175]. (c) Radius 2.5 and the final ϕ is [48.460873, -48.460873]. (d) Radius 3 and the final ϕ is [48.721973, -48.721973].

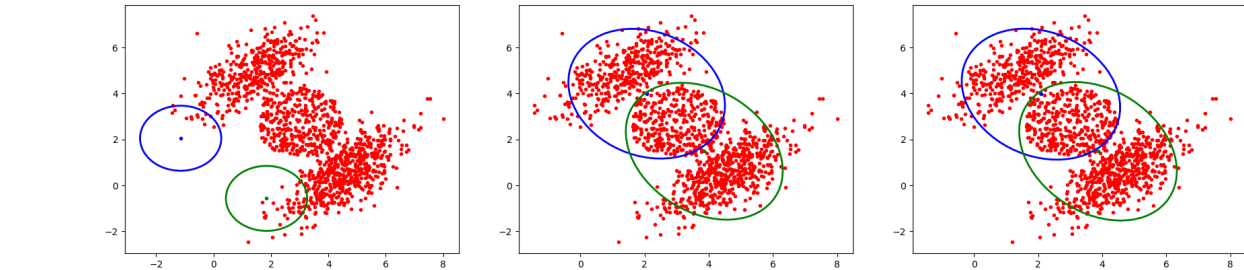
Appendix B-1. Training results



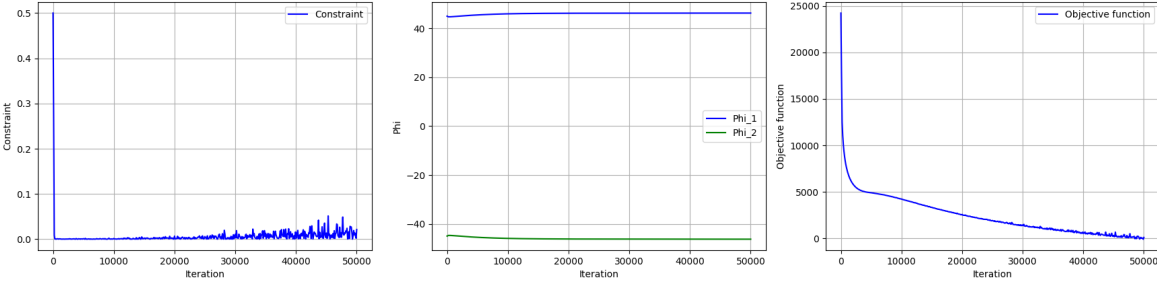


Appendix B-1. Training results with the initial phi 0 (0.1 - (0.1)), lambda 1,500, and the initial means are [-1.1431749105000186, 2.0529438865919479], [1.846096025789632, -0.5656381542446731], and the final phis of all the cases are [0.1, 0.1] (not changed). (a) Radius 1.5. (b) Radius 2. (c) Radius 2.5. (d) Radius 3.

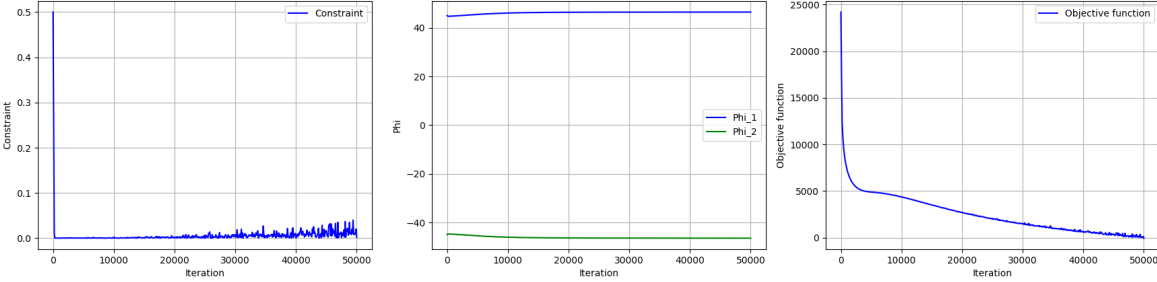
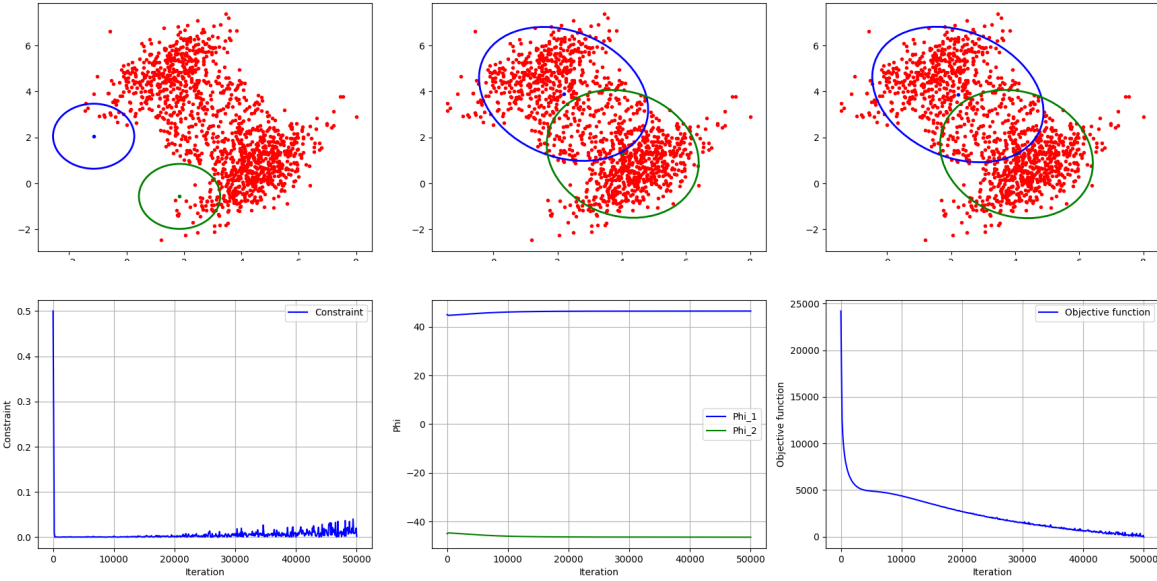
Appendix C-1. Training results

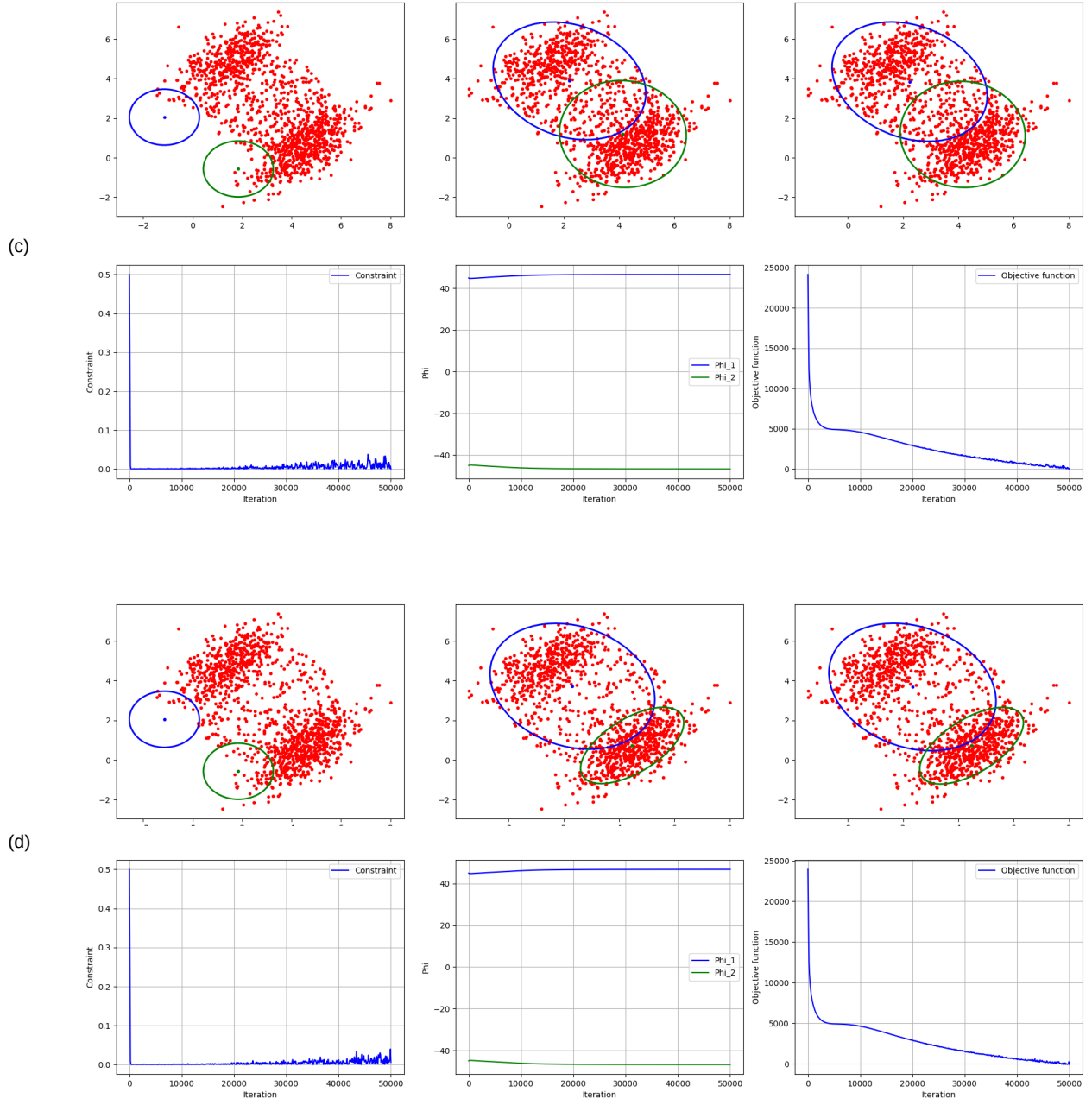


(a)



(b)





Appendix C-1. Training results with the initial ϕ 90 (45 - (-45)), λ 10,000, and the initial means are [-1.1431749105000186, 2.0529438865919479], [1.846096025789632, -0.5656381542446731]. (a) Radius 1.5 and the final ϕ is [46.26429, -46.26429]. (b) Radius 2 and the final ϕ is [46.47259, -46.47259]. (c) Radius 2.5 and the final ϕ is [46.68909, -46.68909]. (d) Radius 3 and the final ϕ is [46.78338, -46.78338].