CS420 Machine Learning Homework 1

Xiaoyu Shi, 516030910567

1 k-mean vs GMM

We adopt an E-M loop for the new variant of k-mean algoritm. Instead of decide which cluster the data point belongs to, we calculate a probility to each of the data point to represent the possibility of it to be in a cluster, and use posterior probability to assign each point to a cluster.

Suppose there are K clusters and N data points, we define the E step and M step as follows, and the pseudo-code1 is presented below.

$$\textbf{E-Step} \qquad p(Z_{nk}) = \frac{||x_n - \mu_k||^2}{\Sigma_{i=1}^{K} ||x_n - \mu_i||^2}$$
$$\textbf{M-Step} \qquad \mu_k = \frac{\Sigma_{i=1}^{N} p_{ik} x_i}{\Sigma_{i=1}^{N} p_{ik}}$$

```
Sum of clusters: K;
for i from 1 to K do
    Init(μ_i);
end
while not convergent do
    E step;
    M step;
end
```

Algorithm 1: a variant of k-mean algorithm

The limitations is that, compared with k-means, we have to do more calculations and the model becomes more complicated, and compared with GMM, due to the difference in distance calculation, the classification is somewhat inaccurate. Also, the same as K-means and EM, classification is affected by the initial value and may not find the global optimum, and the K has to be set.

The advantages, however, is that, for GMM, we do not calculate so much and the model complexity is lowered, and for k-means, we do not assign points to a cluster directly, but use a soft assignment to reduce mistakes.

2 k-mean vs CL

The greatest difference between CL and k-means is that CL algorithm is online, it can make changes to improve the "bad initializations" situation, while k-means is offline.

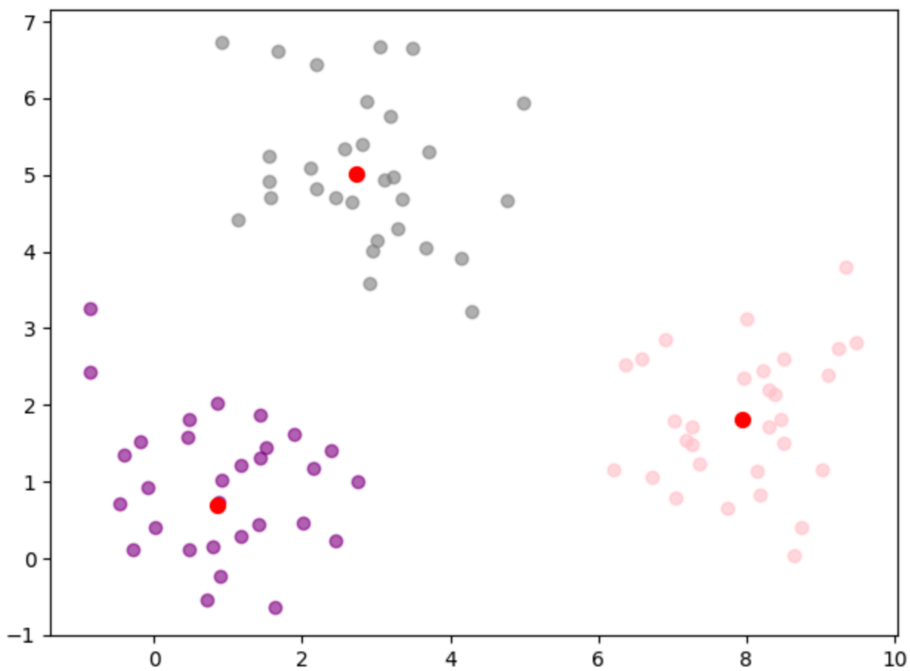

图 1: k-means improved by CL

We add the concept of penalty into K-means. For each new data the algorithm is processing, we calculate the distance between the data point and each center, if a center is the second close to the data point, it gets penalied.

From the figure above we know that the algorithm classified the data points well and the centers are in the right place.

3 model selection of GMM

AIC, BIC: From the figures below we can know that the performances of AIC and BIC are not far-off when facing different number of dimensions and samples.

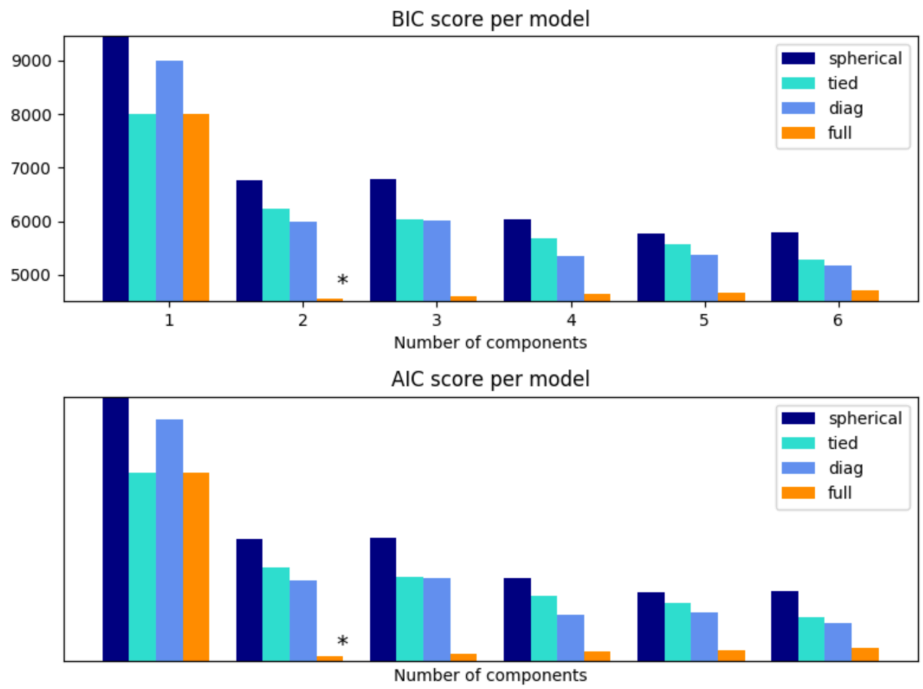

图 2: 500 samples per cluster, 2 dim

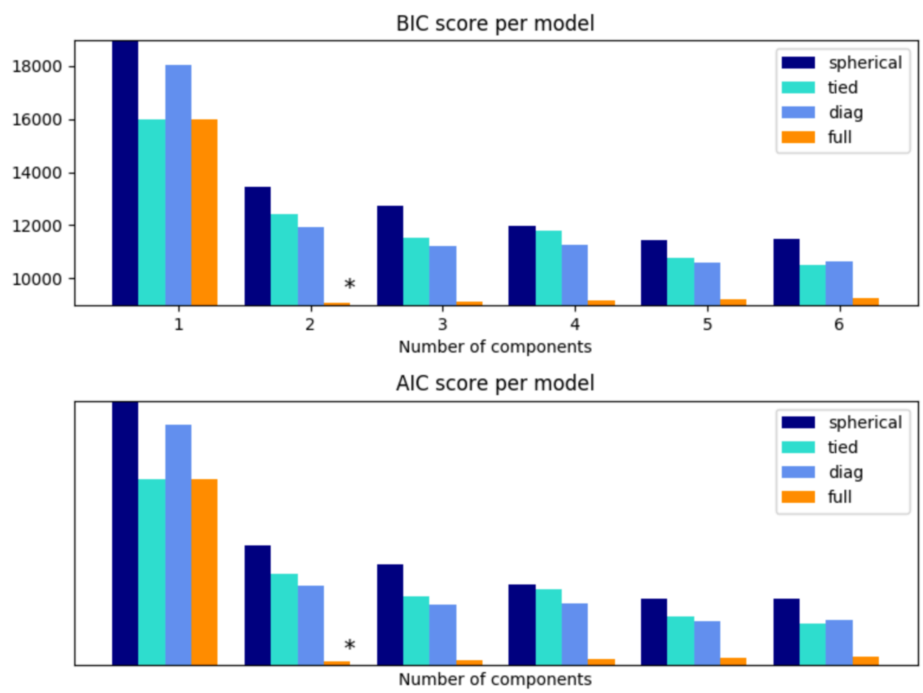

图 3: 1000 samples per cluster, 2 dim

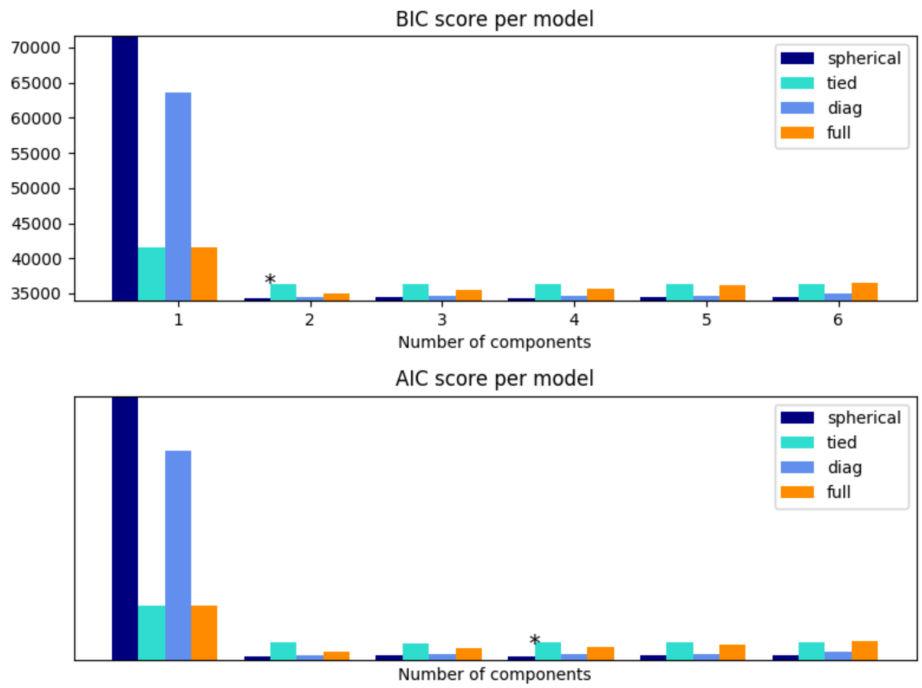

图 4: 500 samples per cluster, 10 dim

VBEM: From the figure below we can see that the algorithm select the optimal $k = 2$ for the dataset automatically, with is much better than the guessed k.

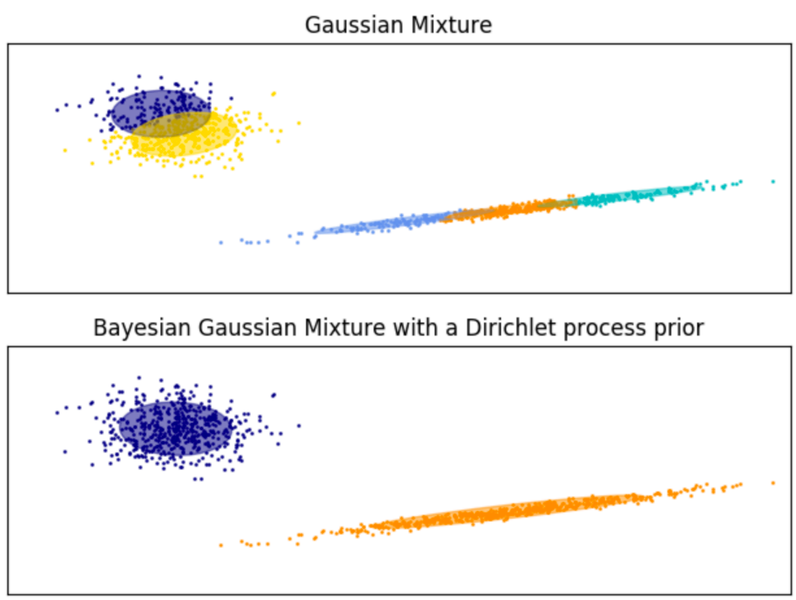

图 5: VBEM