

Matlab 习题 1

1. 设有 GMM 模型的 PDF 写为

$$p(y_i) = \sum_{k=1}^K w_k N(y_i | \mu_k, C_k)$$

设一个具体模型为 $K=3$ ，且参数为

$$\mu_1 = [10, 3]^T, \quad \mu_2 = [1, 1]^T, \quad \mu_3 = [5, 4]^T$$

$$C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}, \quad C_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$w_1 = w_2 = w_3 = 1/3$$

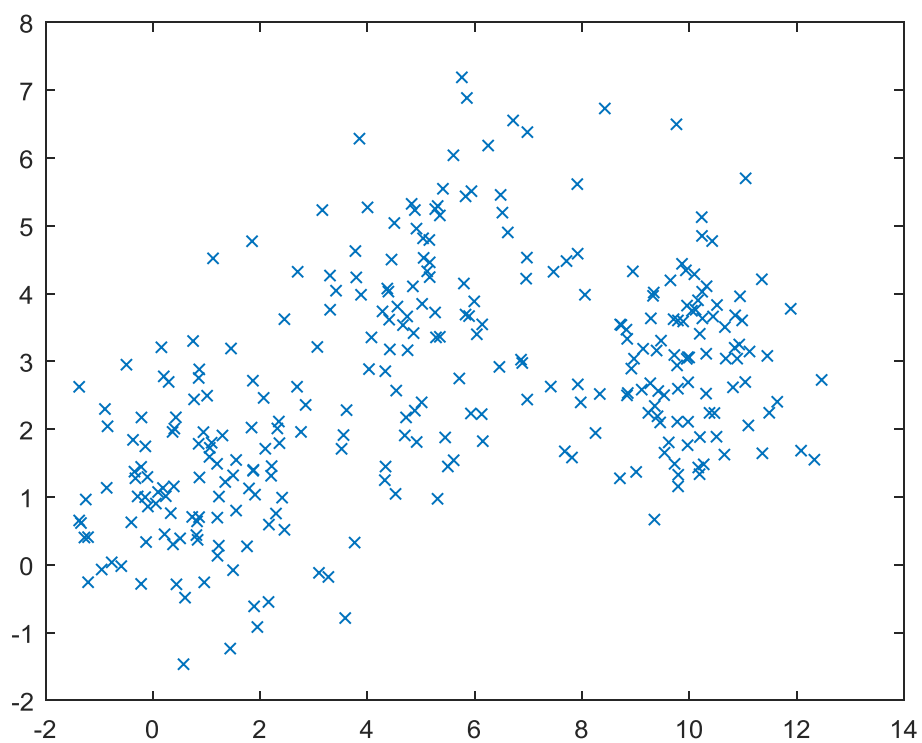
对该模型，随机产生 300 个样本，以产生的样本估计该模型的参数，与实际参数进行比较。为了直观，画出如下两个图

- (1) 按每一个样本点的两维取值，在二维平面画一个“x”点，显示出 300 个点。
- (2) 给出初始值，画出随着迭代似然函数的变化，给出一些不同的初始值，对比估计结果。

解答：

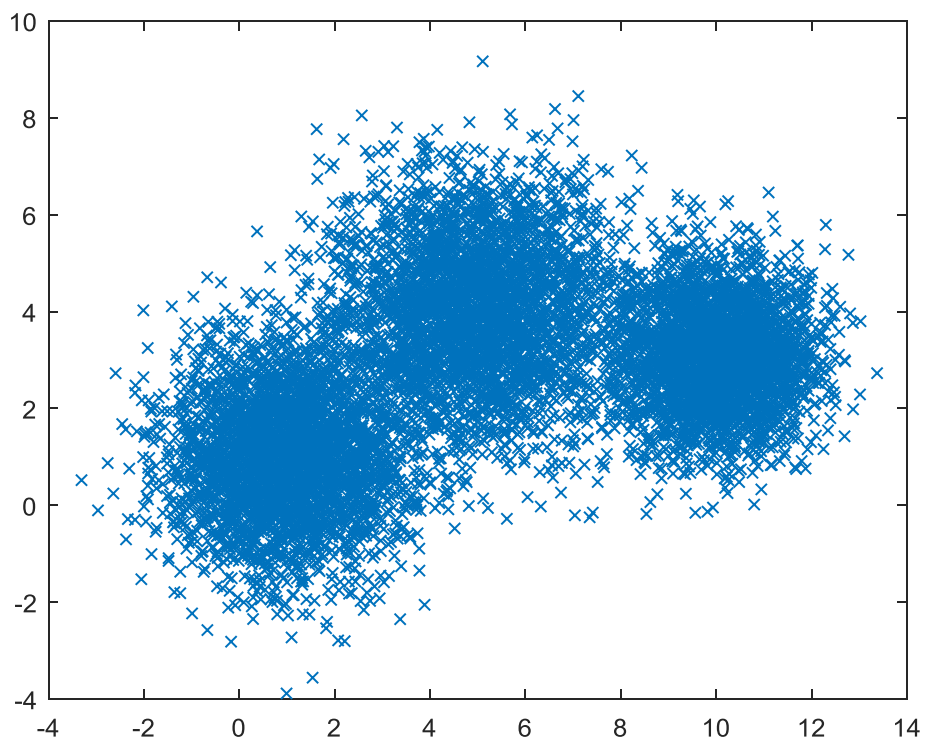
1. 对生成的样本进行显示

- a) 生成 300 个样本的结果：



感觉分布看起来不是很明显。

b) 生成 10000 个点:



由于高斯分布的加权值为均等的，三个分布簇的中心应该是在均值点附近，点的分布情况和实际分布情况相近。

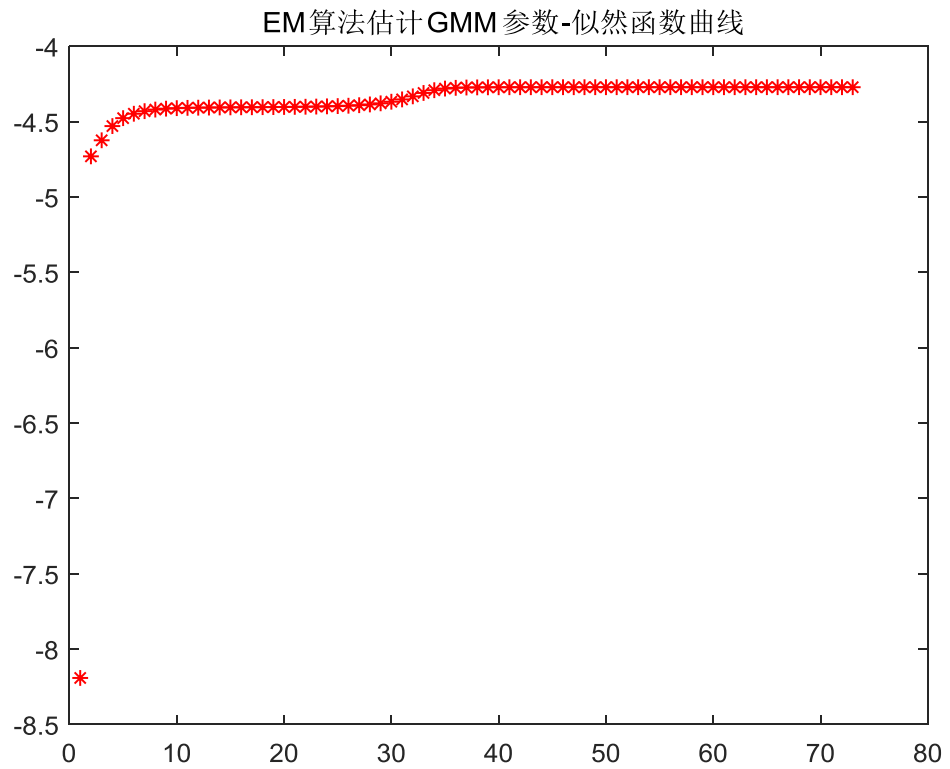
2. 对比估计结果

a) 手动设置初始值

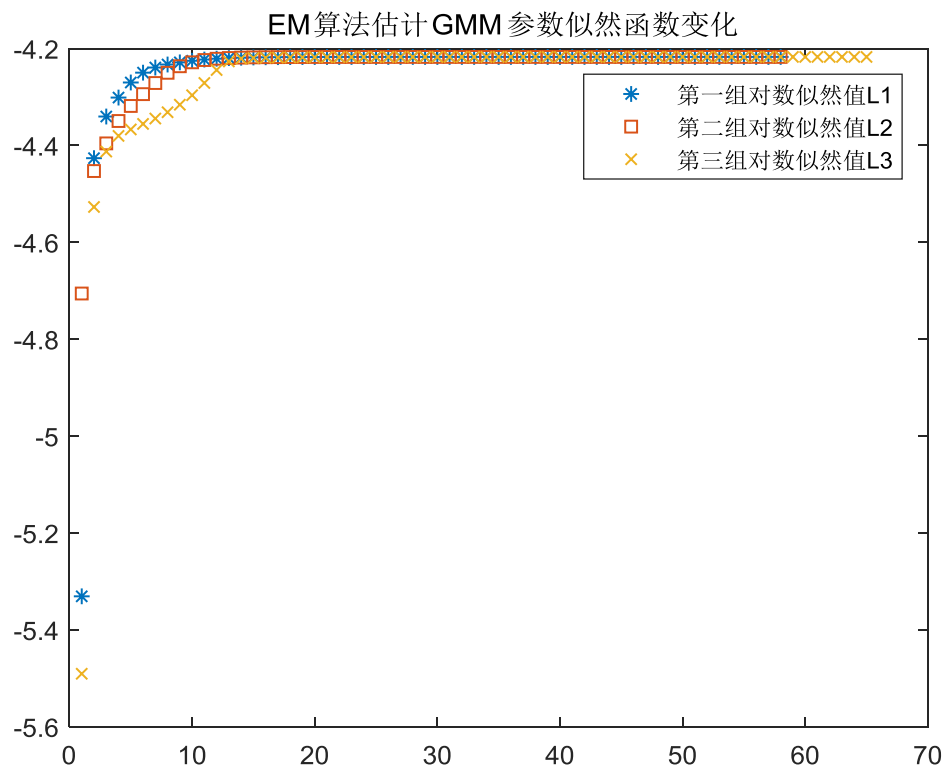
```
-----原始参数值-----
mu_ori =
    10     1     5
     3     1     4
sigma_ori =
     1     0     3     0     2     0
     0     1     0     3     0     2
w =
    0.3333    0.3333    0.3333
-----手动初始化-----
mu =
     3     2     0
     5     1     3
sigma =
     1     0     3     0     2     0
     0     1     0     3     0     2
phi =
    0.5000    0.4000    0.1000
```

```
-----估计值-----
mu_1 =
    5.1522    9.9966    1.1606
    3.9780    2.9918    0.9562
sigma_1 =
    1.6863         0    0.9931         0    1.7687         0
         0    1.9429         0    0.9205         0    1.7600
weight_1 =
    0.3215    0.3462    0.3323
```

似然函数变化:

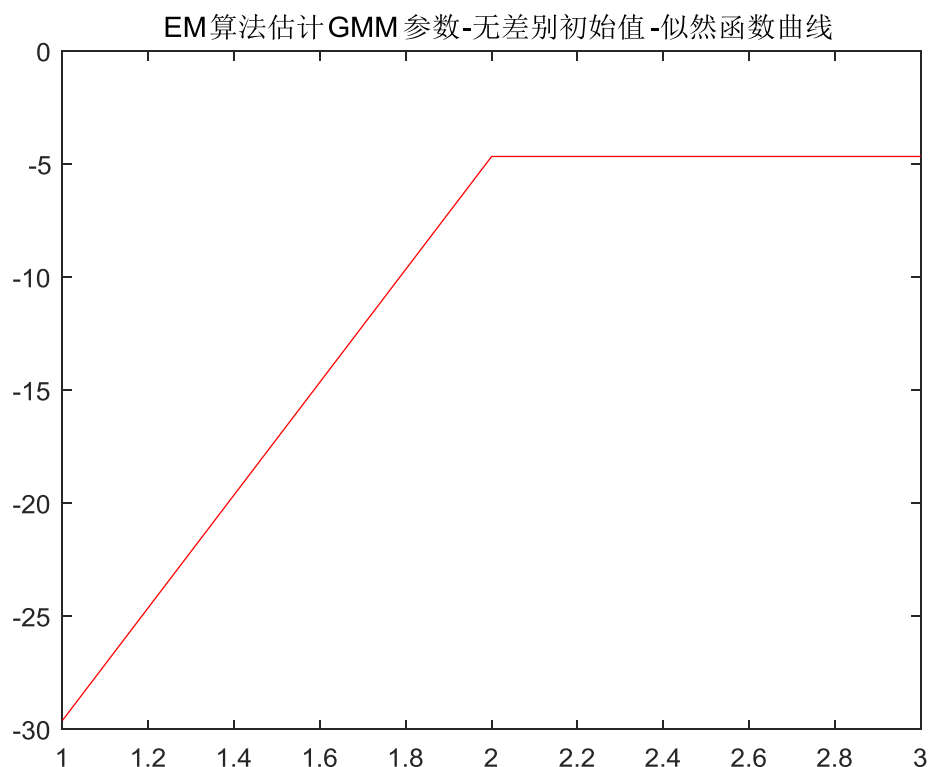


b) 使用随机初始化的方式进行计算



对比结论:

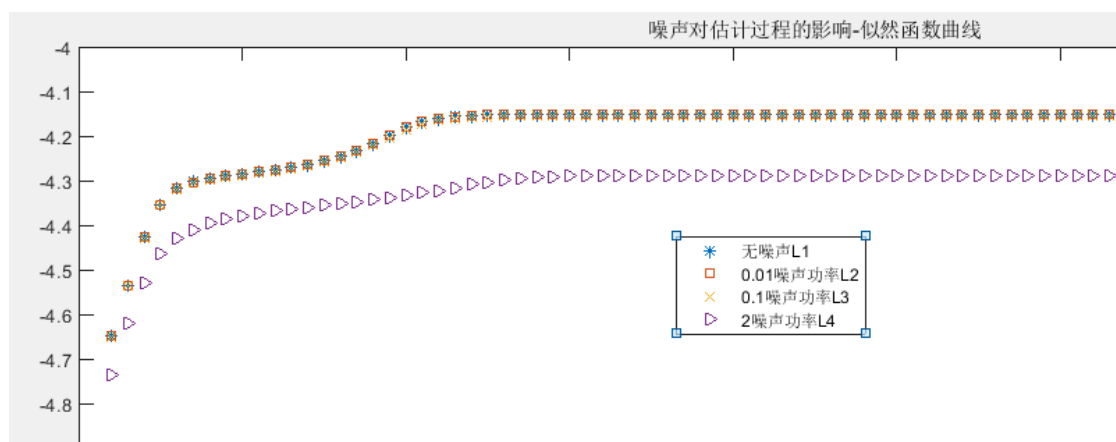
显然，如果是完全随机初始化，通常情况下，性能都差不多，基本上都能在20次以内收敛。如果初始值设置的离原始值比较远，相对来说需要的收敛次数就多一点。如果设置成各组完全没有差别的初始值，收敛的情况最差，其似然函数曲线为：



最终的估计各组的值都一样，估计基本失效。

C) 考虑有噪声情况

分别设置噪声功率为 0, 0.01, 0.1, 2。其结果为：



可以看出，噪声较小的时候对收敛曲线没有影响，收敛结果不变。如果点的噪声较大，会对收敛的最终值产生影响。

d) 其他分析:

在实际解决问题当中,经常是不知道有多少个分布的,首先要对分布的个数进行估计。一种很好的方法是进行数据的可视化,从而对数据的分布有一个简单的认识,另一种方式是先对数据进行聚类,看会有几个聚类中心,然后对 k 值进行设置。

附录 1:

对 GMM 模型使用 EM 算法进行估计的基本过程:

循环下面步骤,直到收敛:{
(E步) 对于每一个 i 和 j 计算

$$w_i(j) = \frac{\frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right) \cdot \phi_j}{\sum_{k=1}^K \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \cdot \phi_k}$$

(M步),更新参数:

$$\begin{aligned}\phi_j &= \frac{1}{N} \sum_{i=1}^N w_i(j) \\ \mu_j &= \frac{\sum_{i=1}^N w_i(j) x_i}{\sum_{i=1}^N w_i(j)} \\ \sigma_j^2 &= \frac{\sum_{i=1}^N w_i(j) (x_i - \mu_j)^2}{\sum_{i=1}^N w_i(j)}\end{aligned}$$

附录 2: 程序运行数据结果

```

-----原始参数值-----
mu_ori =
    10     1     5
     3     1     4
sigma_ori =
     1     0     3     0     2     0
     0     1     0     3     0     2
w =
    0.3333    0.3333    0.3333
-----手动初始化-----
mu =
     3     2     0
     5     1     3
sigma =
     1     0     3     0     2     0
     0     1     0     3     0     2
phi =
    0.5000    0.4000    0.1000
-----估计值-----
mu_l =
    5.2245    9.7941    1.0606
    3.9833    2.9184    1.0585
sigma_l =
    1.8034         0    0.8474         0    1.2804         0
         0    1.6542         0    1.0756         0    1.7680
weight_l =
    0.3239    0.3575    0.3186
-----完全无差别的初始值设置-----
mu =
     0     0     0
     0     0     0
sigma =
     1     0     1     0     1     0
     0     1     0     1     0     1
phi =
    0.3333    0.3333    0.3333
-----无差别初始值时的估计值-----
mu_l =
    5.5314    5.5314    5.5314
    2.6707    2.6707    2.6707
sigma_l =
    14.1901         0    14.1901         0    14.1901         0
         0    2.8918         0    2.8918         0    2.8918
weight_l =
    0.3333    0.3333    0.3333
mu_l =
    1.0606    9.7941    5.2245
    1.0585    2.9184    3.9833
sigma_l =
    1.2804         0    0.8474         0    1.8034         0

```

```

0      1.7680      0      1.0756      0      1.6542
weight_1 =
0.3186      0.3575      0.3239
mu =
2      3      6
2      6      3
sigma =
7      0      3      0      10      0
0      7      0      3      0      10
weight =
0.0190      0.5951      0.3860
mu_2 =
1.0606      5.2245      9.7941
1.0585      3.9833      2.9184
sigma_2 =
1.2804      0      1.8034      0      0.8474      0
0      1.7680      0      1.6542      0      1.0756
weight_2 =
0.3186      0.3239      0.3575
mu =
8      5      5
4      4      6
sigma =
1      0      2      0      9      0
0      1      0      2      0      9
weight =
0.3931      0.4254      0.1815
mu_3 =
9.7941      5.2245      1.0606
2.9184      3.9833      1.0585
sigma_3 =
0.8474      0      1.8034      0      1.2804      0
0      1.0756      0      1.6542      0      1.7680
weight_3 =
0.3575      0.3239      0.3186
-----噪声对估计的影响-----
mu =
3      2      0
5      1      3
sigma =
1      0      3      0      2      0
0      1      0      3      0      2
phi =
0.5000      0.4000      0.1000
>>

```