Programming 1：

（1）m=2,3,4,5 时 EM 算法结果如下图 1，2，3，4 所示。最大似然估计值分别为-4.327，-4.136、-4.137、-4.091，差别不大。但是因为 GMM 个数 m 不同，所以结果有很大的区别。

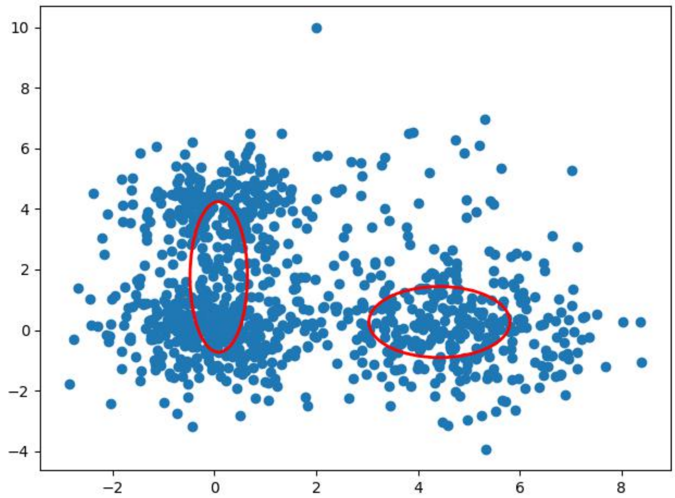

图 1： m=2

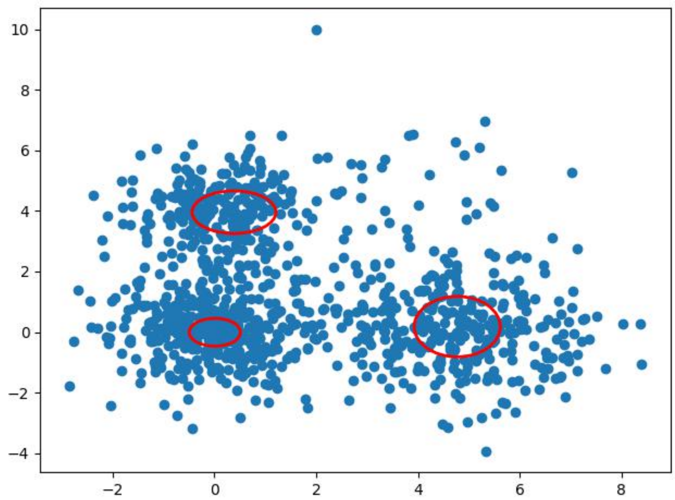

图 2： m=3

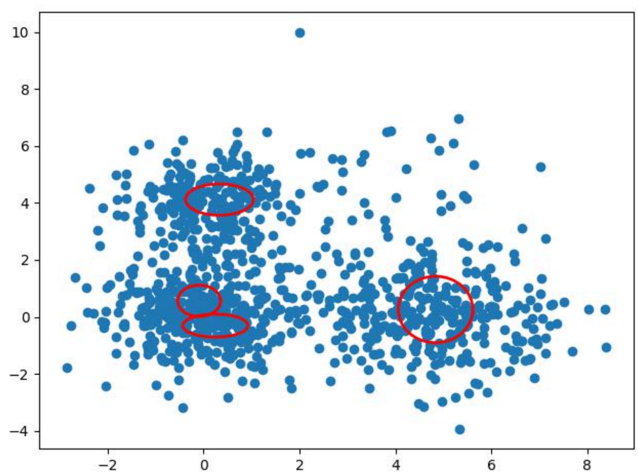

图 3： m=4

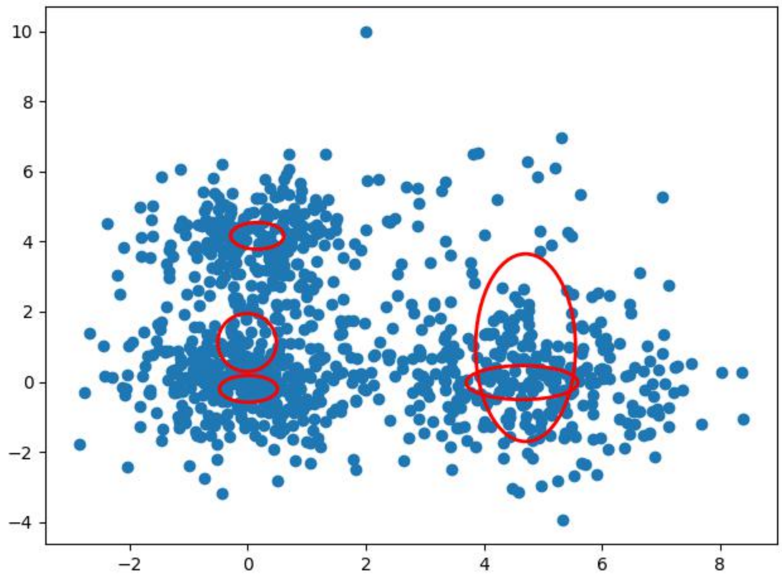

图 4： m=5

根据最大似然估计值可知 m=5 时最好。 肉眼观测可知数据有明显的 3 个密集分布，m=3 时效果最好。另外根据贝叶斯信息准则（Bayes Information Criterion, BIC）$BIC = -ln(n)k + 2ln(L)$，倾向于选择参数更少的模型。m=2, 3, 4, 5 时 BIC 分别为-22.27, -28.99, -35.89, -42.76，所以 m=2 的模型更优。综上，在不同标准下模型选择有不同的结果，要根据实际问题处理。

（2）修改 EM 算法中 M-step 使得不同高斯分布的方差相同，具体推导见手写 pdf。重新运行程序，m=2,3,4,5 时 EM 算法结果如下图 5，6，7，8 所示。最大似然估计值分别为-4.261, -4.228、-4.153、-4.177，差别不大。同样肉眼观测可知数据有明显的 3 个密集分布，m=3 时效果最好。另外根据贝叶斯信息准则（Bayes Information Criterion, BIC）$BIC = -ln(n)k + 2ln(L)$，倾向于选择参数更少的模型，m=2 的模型更优。

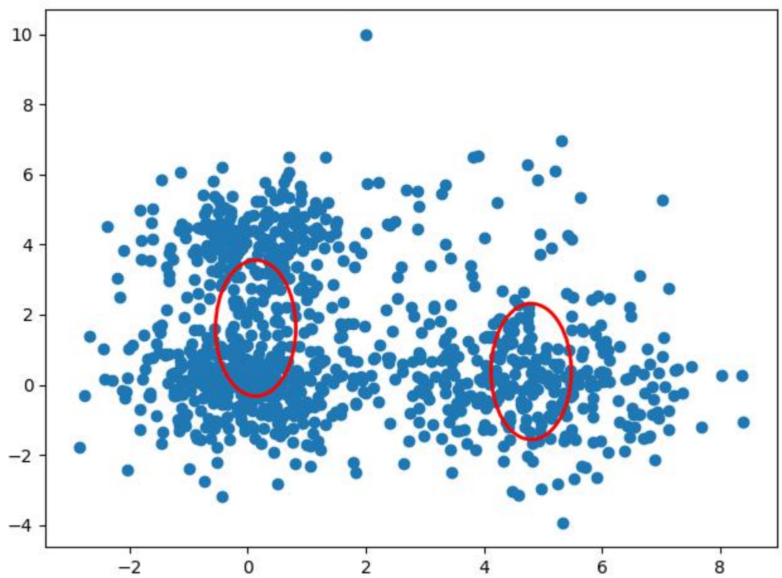

图 5： m=2

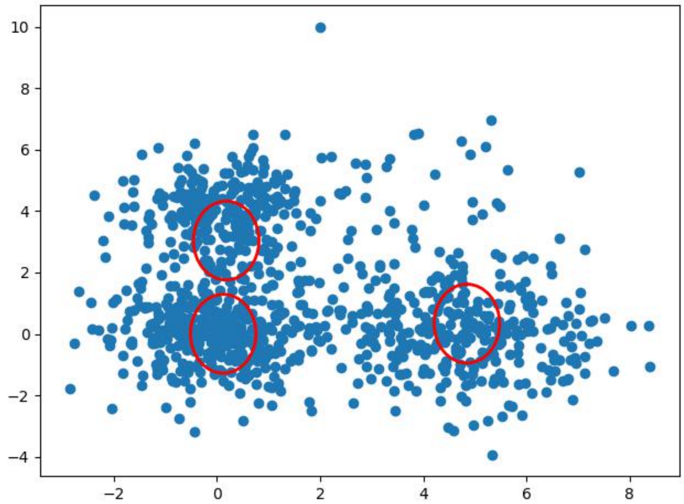

图 6： m=3

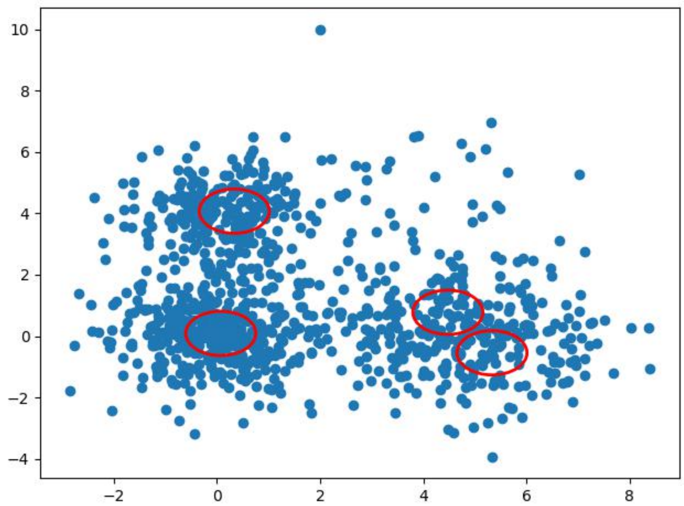

图 7： m=4

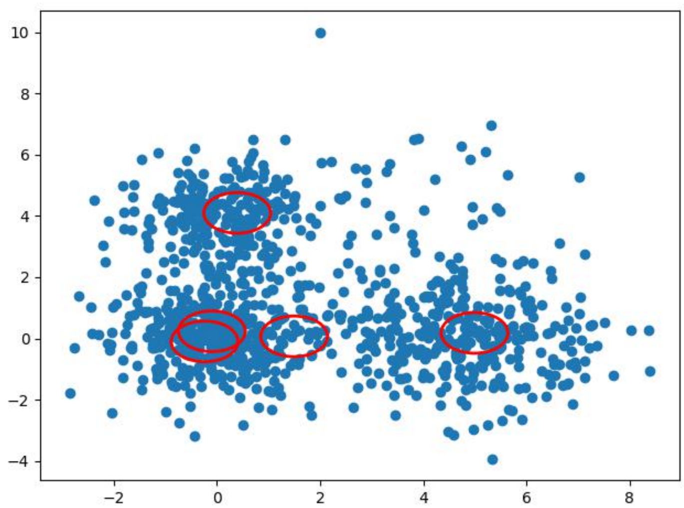

图 8： m=5

Programming2：

具体推导见手写 pdf，程序结果见下图 9。

```
在x3信息缺失的情况下：
估计的均值为[-0.0709     -0.6047      0.77279886]
协方差矩阵为
[[0.90617729 0.56778177 0.88143074]
 [0.56778177 4.20071481 0.46210902]
 [0.88143074 0.46210902 1.78280017]]
在信息完整的情况下：
估计的均值为[-0.0709 -0.6047 -0.911 ]
协方差矩阵为
[[0.90617729 0.56778177 0.3940801 ]
 [0.56778177 4.20071481 0.7337023 ]
 [0.3940801  0.7337023  4.541949  ]]
```

图 9：信息缺失和完整情况下估计得到的均值和协方差

可以看到数值上差别不大，说明 EM 算法在信息缺失的情况下也能起到一定的估计作用。