实验报告

实验五：高斯混合模型参数估计（EM算法）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 封兆欣 | 学号 | 202011010083 | 专业 | 教育技术学 |

# 高斯混合模型参数估计（EM算法）

1. 知识回顾

请详细描述EM算法的原理和迭代过程：

模型假设：

假设X是由个高斯分布均匀混合而成的，这K个高斯分布有自己的均值、方差。

每一个高斯混合模型中的样本点都是通过下面的策略生成的：

1. 在K个高斯混合模型中按照概率P（Y=i）选择一个高斯分布。
2. 这个点在属于第i个高斯分布的条件下概率分布为X ~ N(μi,σi).

即：

P(X|Y=i)~ N(μi,σi).

P（X）=

要优化的目标：关于X的边缘概率密度的似然函数。

=

在选择参数的初始值之后开始进行迭代：

分为E步（Expectation）：计算P（Y|X,）(在给定观测数据X和当前参数的条件下隐变量Y的条件概率分布。

M步（Maximize）：求使优化目标函数最大的新一轮的迭代的参数

重复执行E步 ，M步，直到迭代收敛。

1. EM算法的实现：

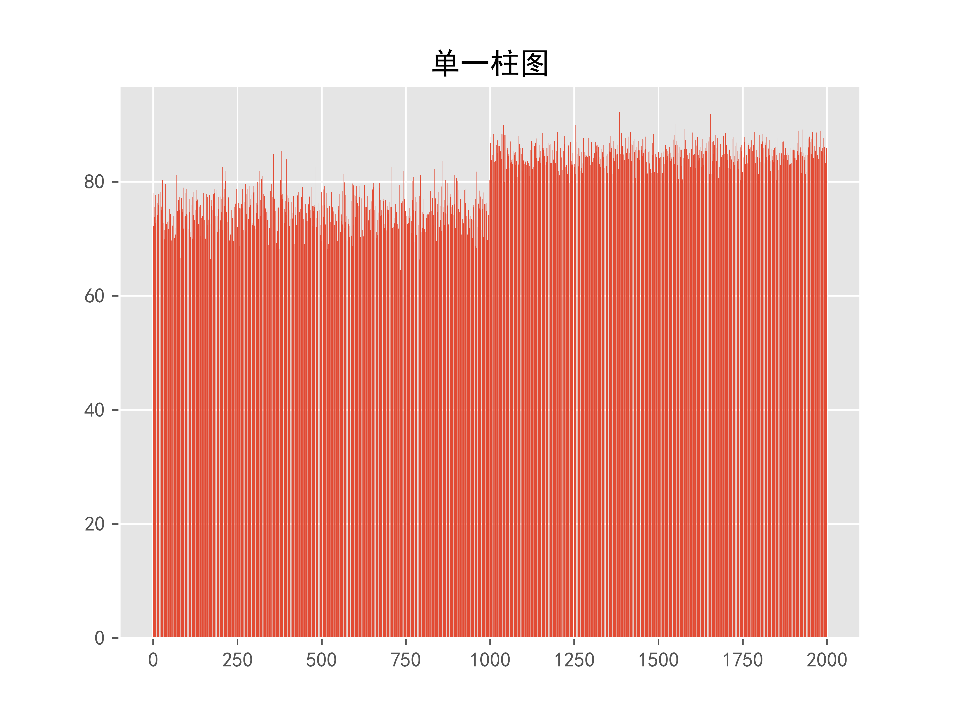
已经给出了主函数的程序，请完成函数em（）并加以注释：

包括E步，M步

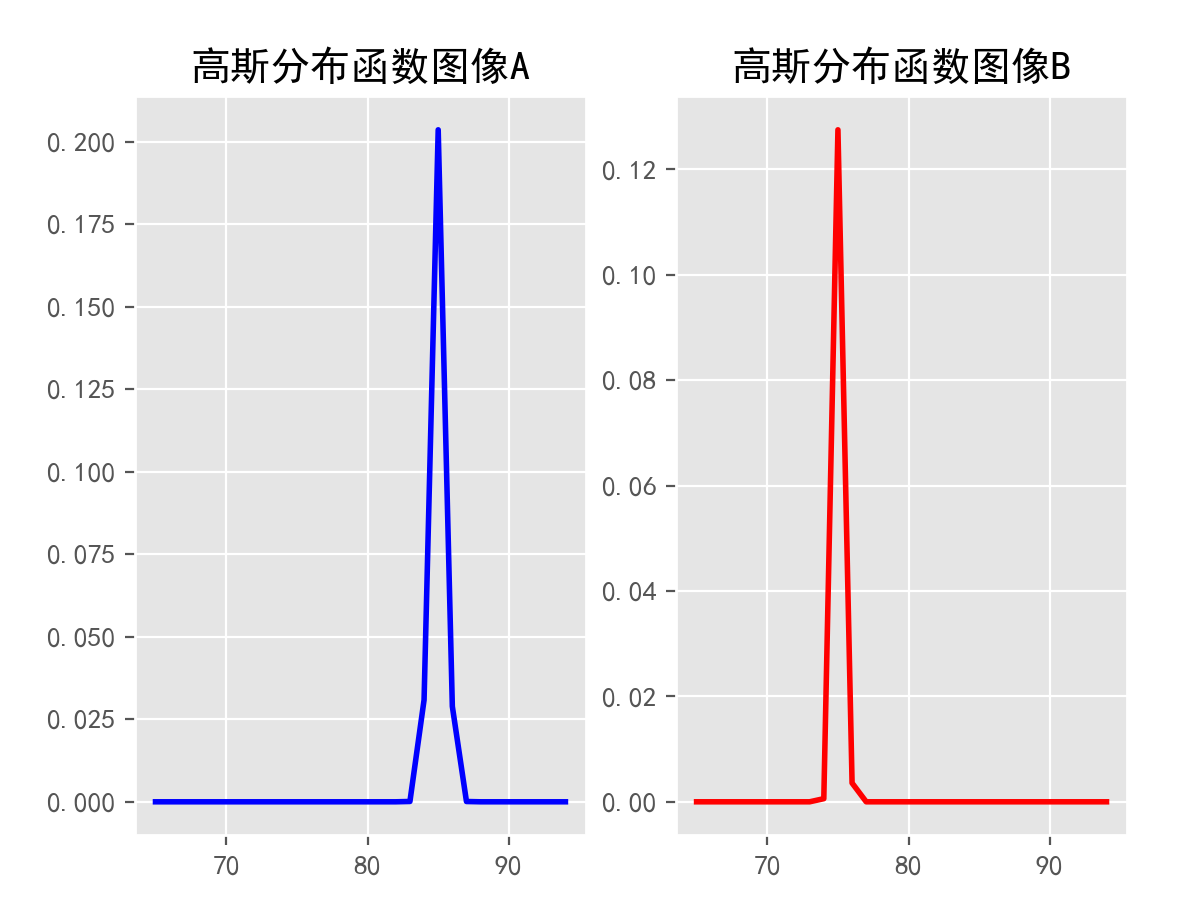
# EM算法的实现  
def em(h, mu1, sigma1, w1, mu2, sigma2, w2, time):  
 *'''  
 h为输入的数据  
 mu1，sigma1,w1分别为A的初始均值，方差，权重  
 mu2，sigma2,w2分别为B的初始均值，方差，权重  
 time代表迭代次数  
 '''* for i in range(time):  
 # 计算响应，E-step  
 r1, r2 = E(myData, mu1, sigma1, w1, mu2, sigma2, w2)  
 # mu、sigmal、w的更新，M-step  
 mu1, sigma1, w1, mu2, sigma2, w2 = M(myData, mu1, mu2, r1, r2)  
 return mu1, sigma1, w1, mu2, sigma2, w2  
  
def Gauss(myData, mu, sigma):  
 P = 1 / (sigma \* np.sqrt(2 \* np.pi)) \* np.exp(-1 \* np.square(myData - mu) / (2 \* np.square(sigma)))  
 return P  
  
def E(myData, mu1, sigma1, w1, mu2, sigma2, w2):  
 # 计算响应度  
 r1 = w1 \* Gauss(myData, mu1, sigma1)  
 r2 = w2 \* Gauss(myData, mu2, sigma2)  
 # 归一化  
 sum = r1 + r2  
 r1 = r1 / sum  
 r2 = r2 / sum  
 return r1, r2  
  
def M(myData, mu1\_old, mu2\_old, r1, r2):  
 # 更新参数mu  
 mu1\_new = np.dot(r1, myData) / np.sum(r1)  
 mu2\_new = np.dot(r2, myData) / np.sum(r2)  
 # 更新参数sigma  
 sigma1\_new = np.sqrt(np.dot(r1, np.square(myData - mu1\_old)) / np.sum(r1))  
 sigma2\_new = np.sqrt(np.dot(r2, np.square(myData - mu2\_old)) / np.sum(r2))  
 # 更新参数w  
 m = len(myData)  
 w1\_new = np.sum(r1) / m  
 w2\_new = np.sum(r2) / m  
 return mu1\_new, sigma1\_new, w1\_new, mu2\_new, sigma2\_new, w2\_new

1. 结果输出

画出数据的直方图：



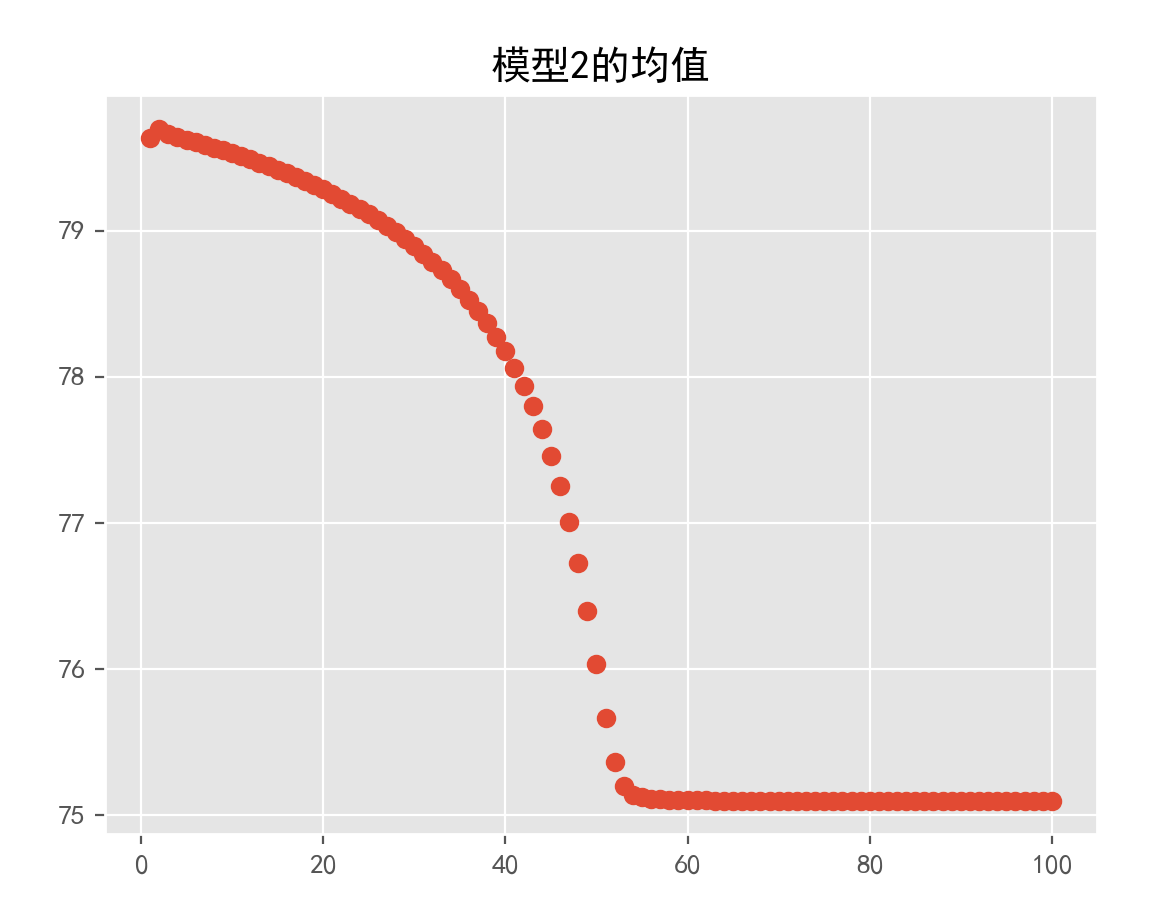
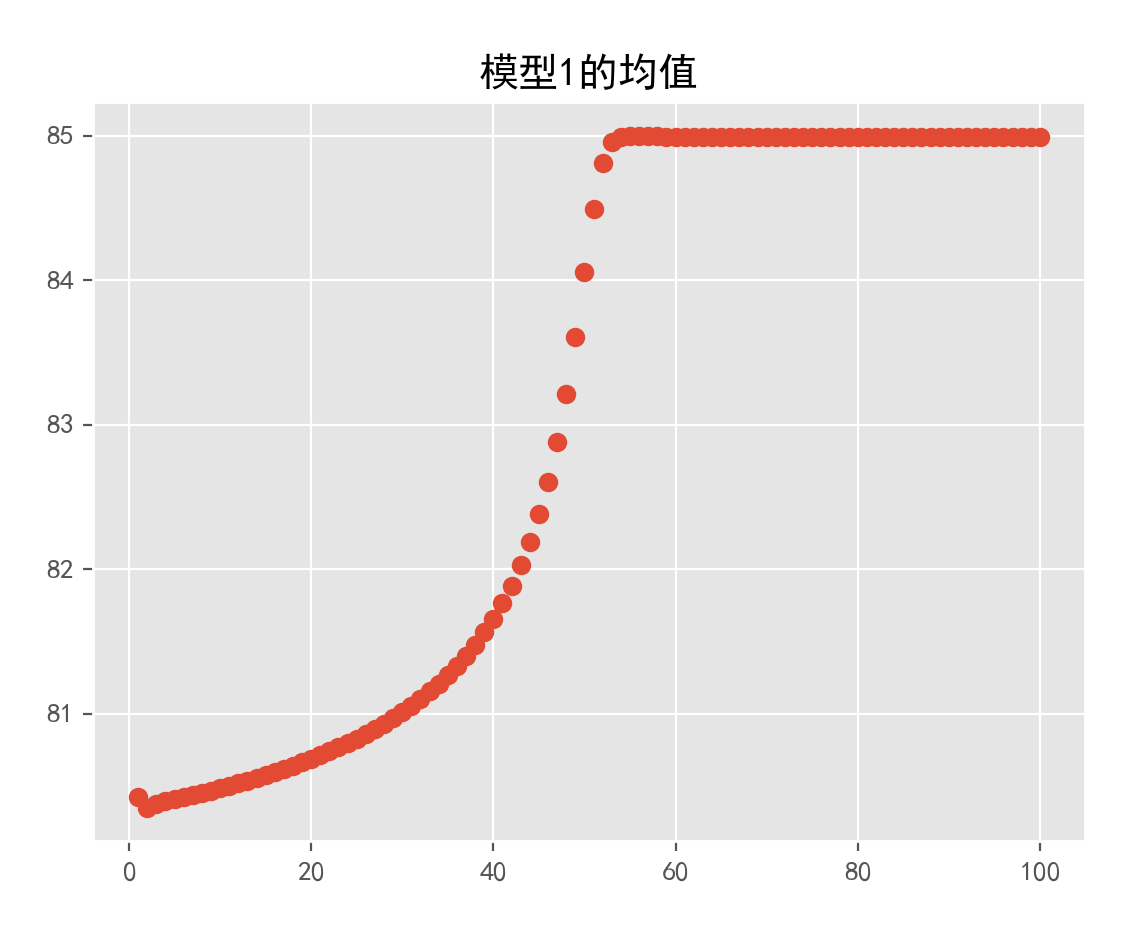
在一张图上画出A和B的高斯估计和混合高斯估计得到的概率密度图：

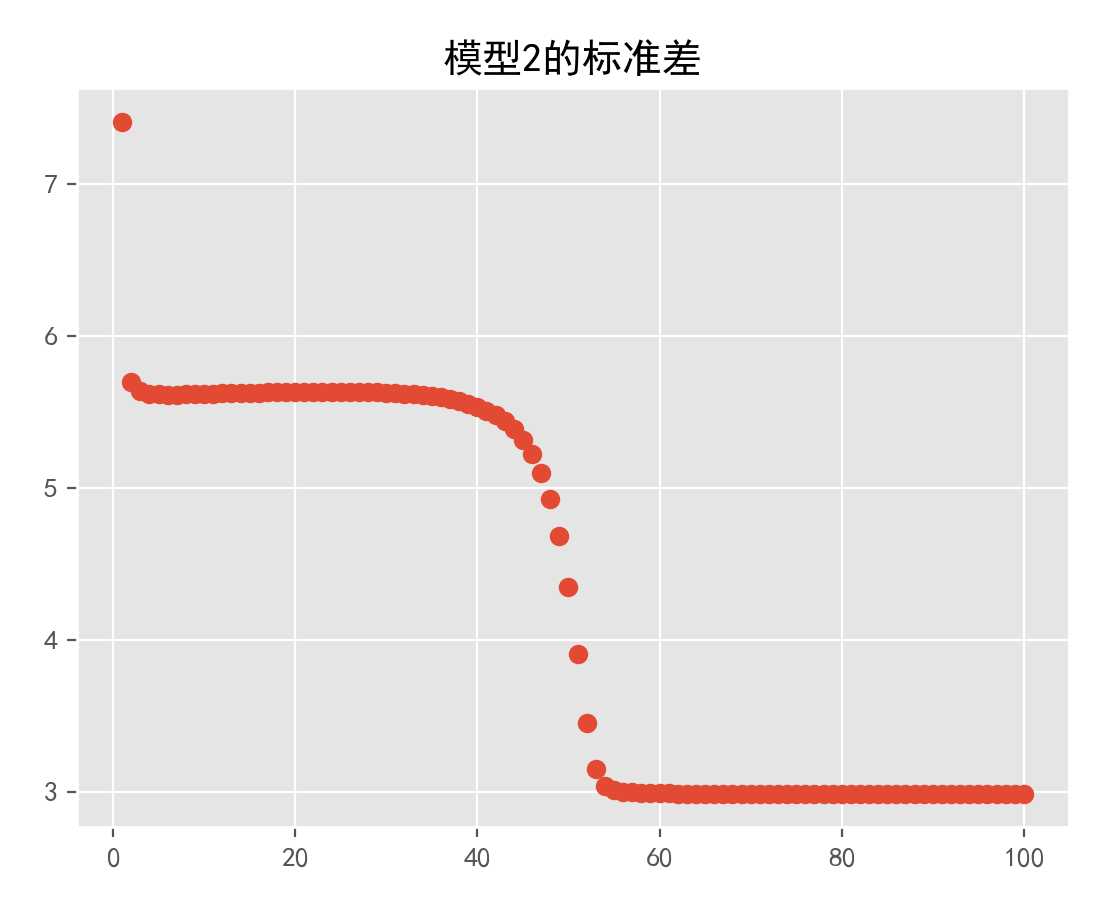
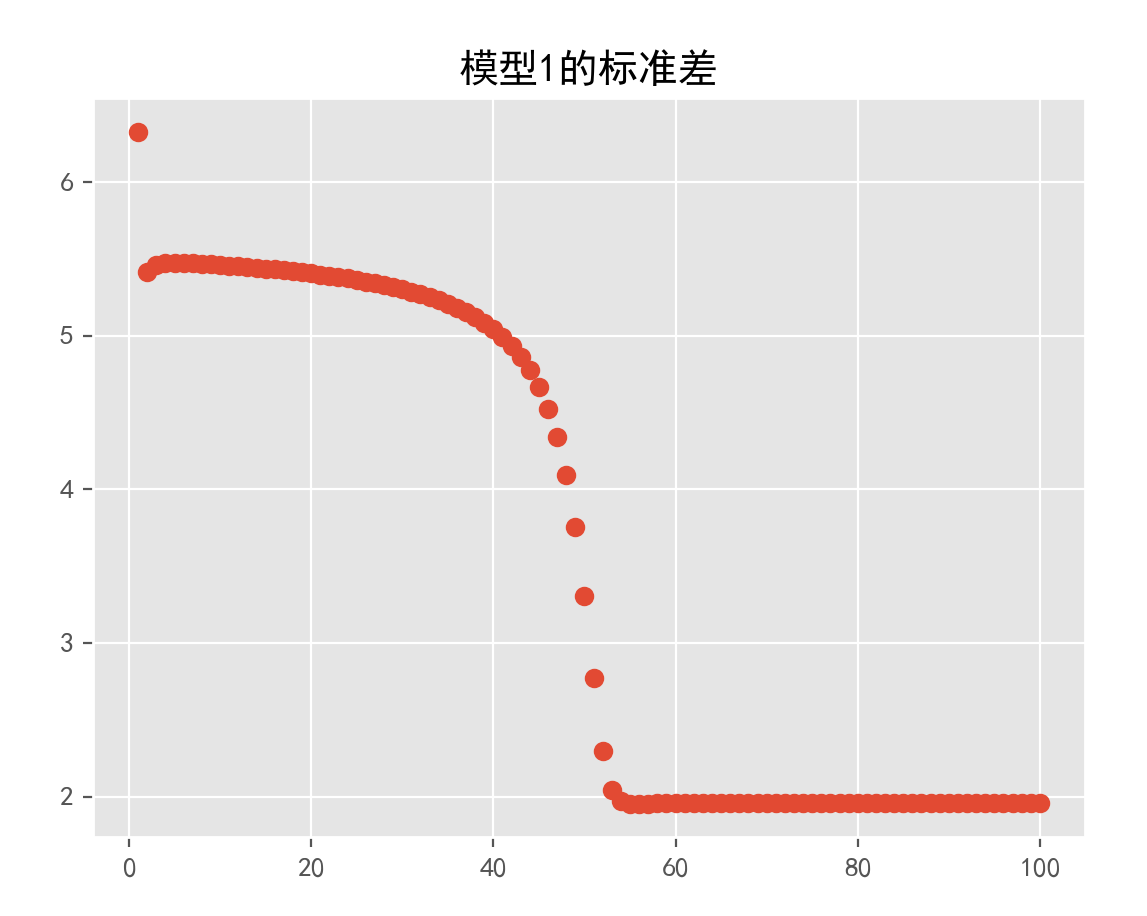


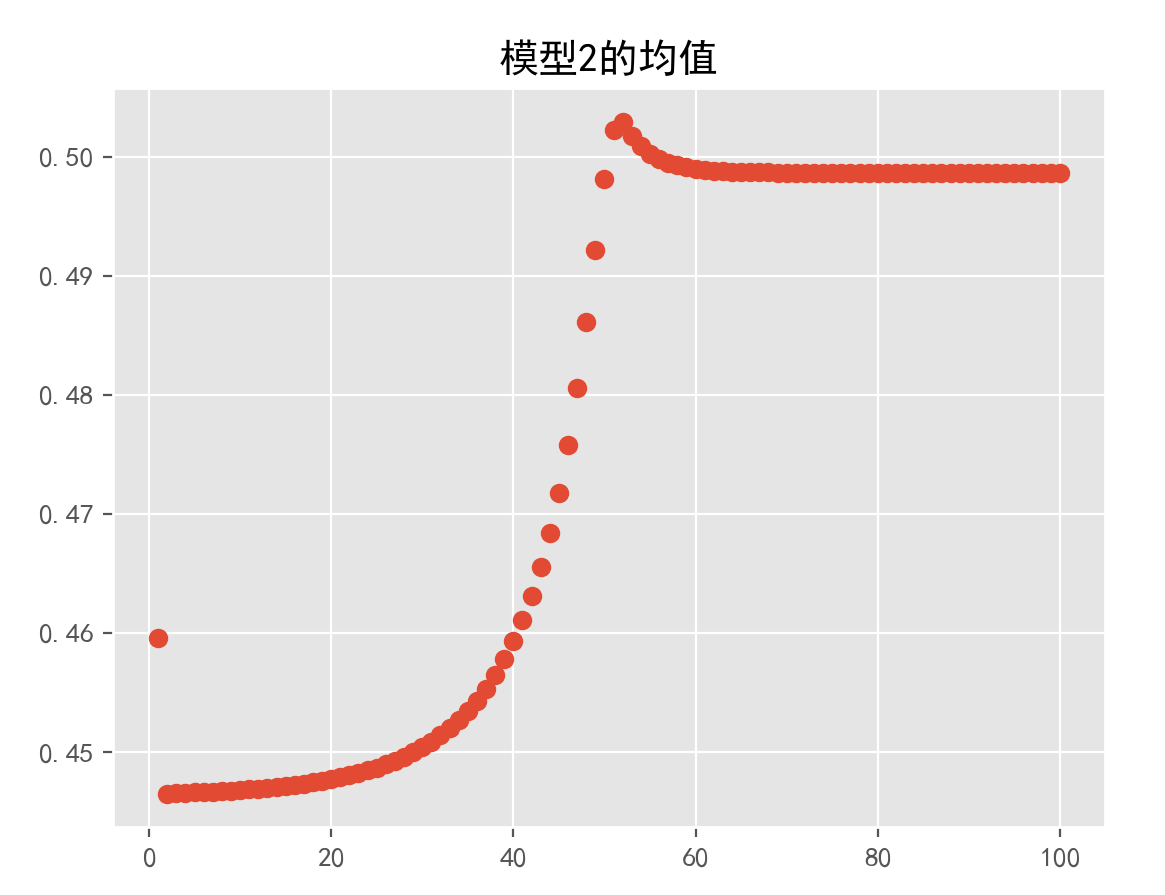
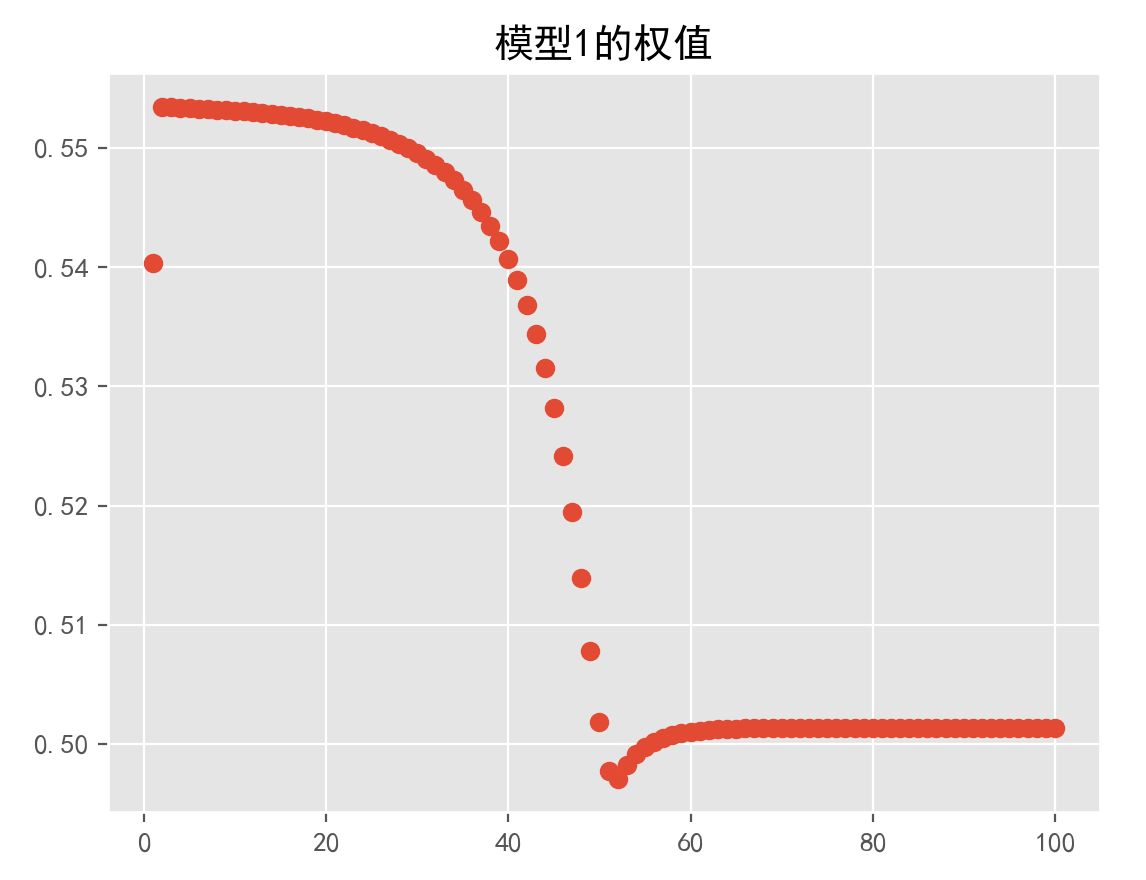
分别给出均值，方差，权重随迭代次数的变化图：

初始参数：mu1 = 77; sigma1 = 5; w1 = 0.5; mu2 = 75; sigma2 = 6; w2 = 0.5

迭代次数：100次（因为50次还看不出明显的收敛，所以改成了50次）







1. 思考题（选做）

如何确定合适的初始值，以加快收敛？

如何实现从高斯分布中随机抽样？

如何实现多元高斯混合模型的参数估计？

# 实验总结

初始值的设置非常重要，决定着是否能够尽快达到收敛。

打印出每一步迭代中参数的更新结果后发现每一次迭代对参数更新的程度很大，但是对优化目标函数的影响却较小。说明EM方法对参数优化的速度非常快，与普通的梯度方法的速度形成鲜明对比。