实验报告

实验五：高斯混合模型参数估计（EM算法）

姓名　姚冠宇　　　学号　202011260070　　专业　人工智能

一. 高斯混合模型参数估计（EM算法）

1. 知识回顾

原理：将数据视为由高斯混合模型采样而得到，我们的目的是求出这几个高斯分布的均值方差和权重
我们可以将我们的生成式模型的概率分布函数定义为：

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

$$0 \leqslant \pi_k \leqslant 1, \quad \sum_{k=1}^{K} \pi_k = 1,$$

那么我们的目标就是最大化由我们的高斯混合模型生成目标数据的概率。
因此损失函数可以写为：

$$\log p(\mathcal{X} \mid \boldsymbol{\theta}) = \sum_{n=1}^{N} \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta}) = \underbrace{\sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}_{=:\mathcal{L}}.$$

通过对损失函数求极值我们可以得到我们参数的目标值。对参数求偏导有：

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}_k} = \mathbf{0}^\top \iff \sum_{n=1}^{N} \frac{\partial \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} = \mathbf{0}^\top,$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Sigma}_k} = \mathbf{0} \iff \sum_{n=1}^{N} \frac{\partial \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} = \mathbf{0},$$

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = 0 \iff \sum_{n=1}^{N} \frac{\partial \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \pi_k} = 0.$$

且：

$$\frac{\partial \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{p(\boldsymbol{x}_n \mid \boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

$$\frac{1}{p(\boldsymbol{x}_n \mid \boldsymbol{\theta})} = \frac{1}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

故我们定义r_k来表示数据点由第k个高斯分布生成的概率：

$$r_{nk} := \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

首先计算μ
有：

$$\begin{aligned}\frac{\partial p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} &= \sum_{j=1}^{K} \pi_j \frac{\partial \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\partial \boldsymbol{\mu}_k} = \pi_k \frac{\partial \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\mu}_k} \\ &= \pi_k (\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),\end{aligned}$$

故：

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}_k} &= \sum_{n=1}^{N} \frac{\partial \log p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} = \sum_{n=1}^{N} \frac{1}{p(\boldsymbol{x}_n \mid \boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} \\ &= \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} \underbrace{\boxed{\frac{\pi_k \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}}}_{=r_{nk}} \\ &= \sum_{n=1}^{N} r_{nk} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}.\end{aligned}$$

令偏导为0，有：

$$\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n = \sum_{n=1}^{N} r_{nk}\boldsymbol{\mu}_k^{\text{new}} \iff \boldsymbol{\mu}_k^{\text{new}} = \frac{\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n}{\boxed{\sum_{n=1}^{N} r_{nk}}} = \frac{1}{\boxed{N_k}}\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n\,,$$

得到μ更新公式：

$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n}{\sum_{n=1}^{N} r_{nk}}\,,$$

再计算方差，同理令偏导为0有：

$$\begin{aligned}\frac{\partial p(\boldsymbol{x}_n \mid \boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_k} &= \pi_k \mathcal{N}\big(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big) \\ &\quad \cdot \left[-\tfrac{1}{2}\left(\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^{\top}\boldsymbol{\Sigma}_k^{-1}\right)\right].\end{aligned}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^{\top}\,,$$

最后计算权重
使用拉格朗日乘子法（因为权重有限制求和为1），损失函数为：

$$= \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}\big(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

求偏导：

$$\frac{\partial \mathfrak{L}}{\partial \lambda} = \sum_{k=1}^{K}\pi_k - 1\,.$$

$$
\begin{aligned}
\frac{\partial \mathfrak{L}}{\partial \pi_k} &= \sum_{n=1}^{N} \frac{\mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_j,\, \boldsymbol{\Sigma}_j\big)} + \lambda \\
&= \frac{1}{\pi_k} \underbrace{\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_j,\, \boldsymbol{\Sigma}_j\big)}}_{=N_k} + \lambda = \frac{N_k}{\pi_k} + \lambda\,,
\end{aligned}
$$

故EM的算法的流程可以写为：

1. Initialize $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$.
2. *E-step:* Evaluate responsibilities r_{nk} for every data point $\boldsymbol{x}_n$ using current parameters $\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$:

$$
r_{nk} = \frac{\pi_k \mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)}{\sum_j \pi_j \mathcal{N}\big(\boldsymbol{x}_n \,|\, \boldsymbol{\mu}_j,\, \boldsymbol{\Sigma}_j\big)}\,. \tag{11.53}
$$

3. *M-step:* Reestimate parameters $\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ using the current responsibilities r_{nk} (from E-step):

$$
\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} \boldsymbol{x}_n\,, \tag{11.54}
$$

$$
\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top\,, \tag{11.55}
$$

$$
\pi_k = \frac{N_k}{N}\,. \tag{11.56}
$$

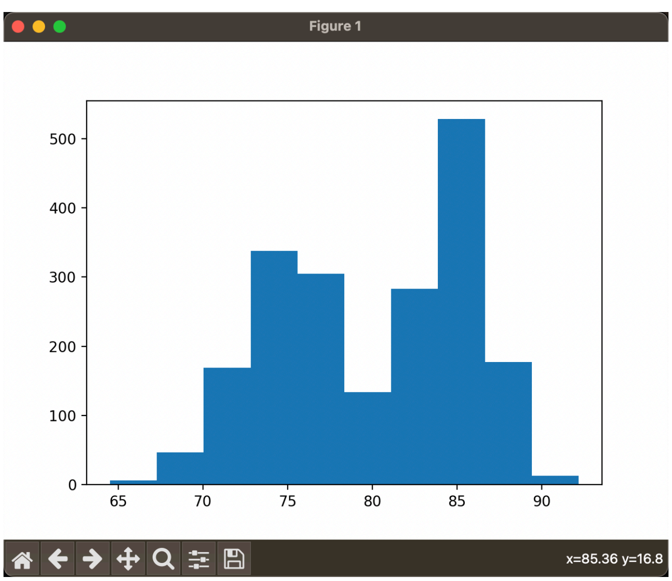
2. EM算法的实现：

已经给出了主函数的程序，请完成函数em（）并加以注释

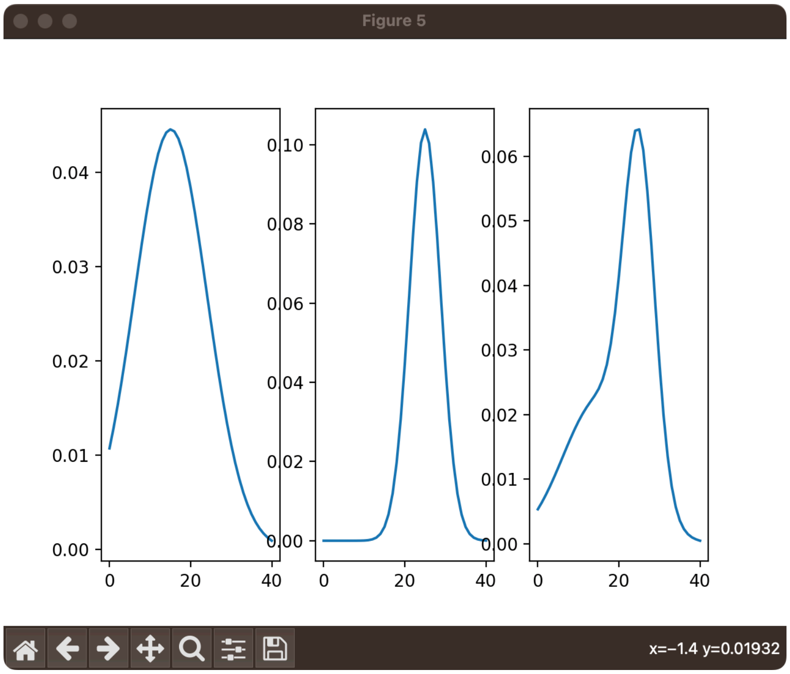
```python
def em(iter):
    '''
    h为输入的数据
    mu1，sigma1,w1分别为A的初始均值，方差，权重
    mu2，sigma2,w2分别为B的初始均值，方差，权重
    '''
    #E-step,将中间值计算出来加速计算
    temp1=w1m[iter-1]*stats.norm(mu1m[iter-1],np.sqrt(sigma1m[iter-1])).pdf(h)
    temp2=w2m[iter-1]*stats.norm(mu2m[iter-1],np.sqrt(sigma2m[iter-1])).pdf(h)
    temp3=temp1+temp2
    r1=temp1/temp3
    r2=temp2/temp3
    #更新均值
    mu1m[iter]=np.sum(r1*h)/sum(r1)
    mu2m[iter]=np.sum(r2*h)/sum(r2)
    #更新方差
    #pdb.set_trace()
    sigma1m[iter]=np.sum(r1*(h-mu1m[iter])**2)/sum(r1)
    sigma2m[iter]=np.sum(r2*(h-mu2m[iter])**2)/sum(r2)
    #更新权重
    w1m[iter]=sum(r1)/lenth
    w2m[iter]=sum(r2)/lenth
```

3. 结果输出

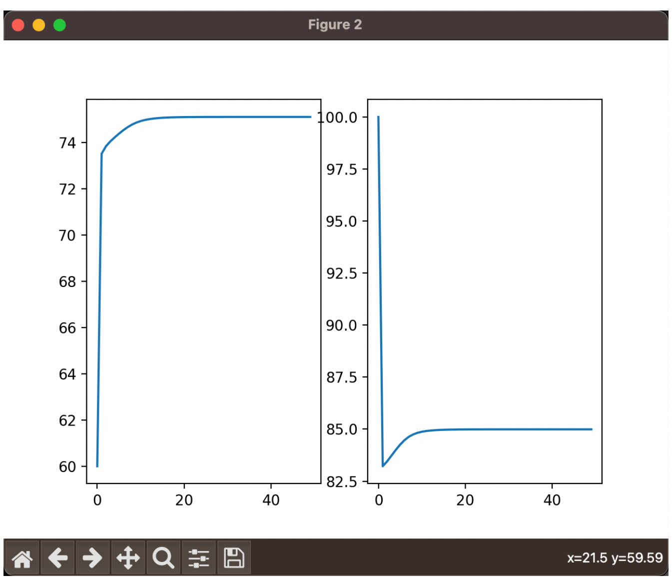
画出数据的直方图：

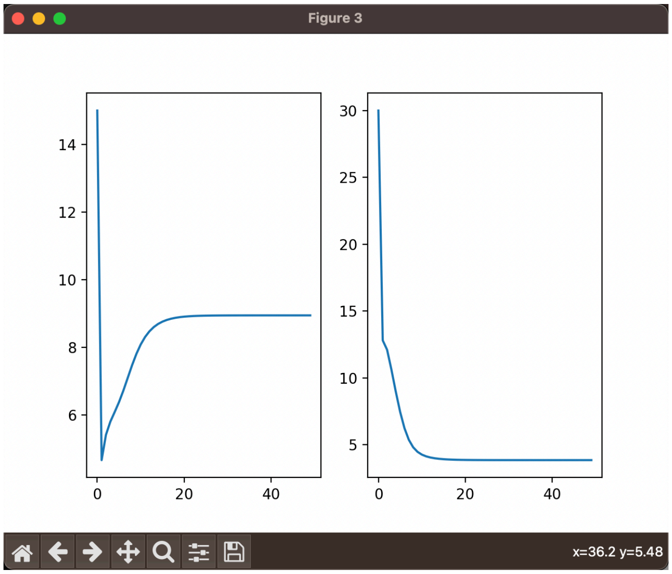


在一张图上画出A和B的高斯估计和混合高斯估计得到的概率密度图：

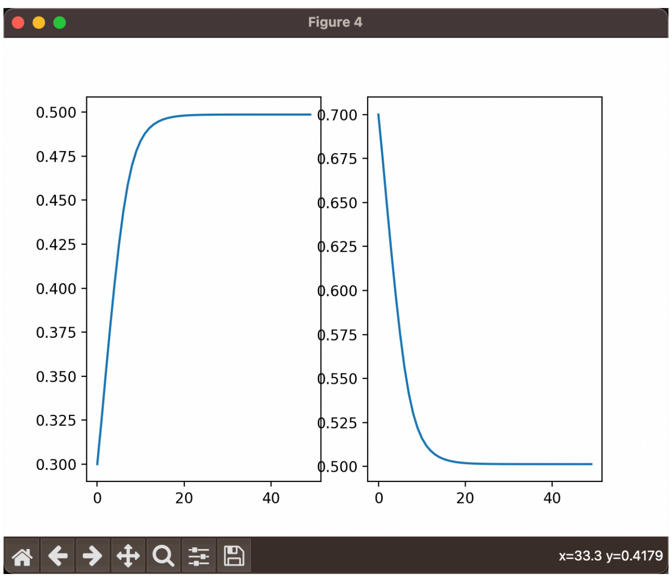


分别给出均值，方差，权重随迭代次数的变化图：
均值：





方差：



权重：

4. 思考题（选做）

如何确定合适的初始值，以加快收敛?

观察数据的分布，合理估计高斯分布个数（峰值个数），均值（峰值所在处），权重（峰值比例），方差

如何实现从高斯分布中随机抽样?

先从标准高斯分布中抽样，之后通过目标分布的均值和方差对抽样结果进行线性变换得到我们想要的抽样结果。需要注意的是如何从混合高斯模型里抽样，先从各个高斯分布抽样得到样本，最后将结果加权相加得到混合高斯模型的抽样结果。

如何实现多元高斯混合模型的参数估计?

第一部分已经详细解释

二、实验总结

本实验实现了em算法估计，但是由于75和85有些接近，不合适的参数会让结果收敛到均值80，初始化的参数应当多加注意。
得到混合高斯模型的参数之后，我们既可以通过采样来生成数据，也可以应用贝叶斯定理来判断数据的来源。具体公式为：

$$
\begin{aligned}
p(z_{nk} = 1 \mid \boldsymbol{x}_n) &= \frac{p(\boldsymbol{x}_n \mid z_{nk} = 1) p(z_{nk} = 1)}{\sum_{j=1}^{K} p(\boldsymbol{x}_n \mid z_{nj} = 1) p(z_{nj} = 1)} \\
&= \frac{\pi_k \mathcal{N}\left(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\left(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)} = r_{nk} \,.
\end{aligned}
$$

可以得到数据点由每一类生成的概率。可以应用在归因分析之中。