实验介绍

实验五：高斯混合模型参数估计（EM算法）

### 实验背景：

期望最大化方法（EM）算法是一种重要的用于解决含有隐变量（latent variable）问题的参数估计方法。在高斯混合模型（GMM）中，EM算法是估计高斯混合模型参数的有力工具。

假设有两个班级（班级A和班级B），得到两个班级中每个人的语文期末考试成绩（满分100），却不知道这些成绩来自哪个班级。假设两个班级的成绩均符合高斯分布，请利用EM算法，估算出混合高斯模型的相关参数。

# 实验目的

理解并掌握使用EM算法估计高斯混合模型参数的方法。

### 实验内容及要求:

共有2000个语文成绩的数据，数据以.mat的数据格式给出，请完成以下任务：

1. 自行编程实现EM算法；
2. 完成EM算法编程后，根据提供的程序，完成实验报告。实验报告要求包括：知识回顾，EM算法函数的实现，结果输出等；

本次实验安排为一次实验课，需要在本次实验课结束的一周内完成自己的实验报告，并提交实验报告以及实验代码（以.py文件的格式提交）

提交链接： <https://pan.bnu.edu.cn/l/8FDV2k>

程序：

# 导入必要的库  
import numpy as np  
import pandas as pd  
import matplotlib.pylab as plt  
from scipy import stats  
import scipy.io as scio  
  
# EM算法的实现  
def em(h, mu1, sigma1, w1, mu2, sigma2, w2):  
 *'''  
 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  
 # 计算响应  
  
 # p1, p2权重 \*A B的后验概率  
  
  
 # M-step  
 # mu的更新  
  
 # sigmal的更新  
 # w的更新  
 return mu1, sigma1, w1, mu2, sigma2, w2  
  
  
# 加载数据  
dataFile = 'data.mat'  
h = scio.loadmat(dataFile)  
  
  
# GMM的构造  
# Step 1.首先根据经验来分别对A,B班级的均值、方差和权值进行初始化,如：  
'''  
mu1 = ???  
sigma1 = ???  
w1 = ??? # A  
mu2 = ???  
sigma2 = ???  
w2 = ??? # B  
'''  
  
n = len(h) # 样本长度  
mu1m = np.zeros(50)  
mu2m = np.zeros(50)  
w1m = np.zeros(50)  
w2m = np.zeros(50)  
sigma1m = np.zeros(50)  
sigma2m = np.zeros(50)  
w1m[0] = w1  
w2m[0] = w2  
sigma1m[0] = sigma1  
sigma2m[0] = sigma2  
# 开始EM算法的主循环  
for i in range(1, 50):  
 mu1, sigma1, w1, mu2, sigma2, w2 = em(h, mu1, sigma1, w1, mu2, sigma2, w2)  
   
# 作图