作业三：2D MOG

实验目的：用 EM 方法生成二维高斯分布。

实验过程：

高斯混合模型就是假设数据服从高斯混合分布，即数据可以看作是从数个高斯分布中生成出来的。每个 MOG 由 K 个高斯分布组成，每个高斯称为一个 component，这些 component 线性相加组成 MOG 的概率密度函数，即：

$$\begin{aligned} p(x) &= \sum_{k=1}^{K} p(k)p(x|k) \\ &= \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \end{aligned}$$

D 维高斯分布可以表示为：

$$\mathrm{N}(\mathrm{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathrm{x}-\mu)^T \Sigma^{-1} (\mathrm{x}-\mu)\right\}$$

其中，μ是 D 维的均值向量，Σ 是 D*D 的协方差矩阵，$|\Sigma|$ 表示 Σ 的行列式。具体的求解过程分为两个进行：

1. 估计数据点由每个高斯模型生成的概率：对于每个数据 x_i 来说，它由第 k 个 Component 生成的概率为

$$\gamma(i,k) = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}$$

由于式子里的 μ_k 和 Σ_k 也是需要我们估计的值，我们采用迭代法，在计算 $\gamma(i,k)$ 的时候我们假定 μ_k 和 Σ_k 均已知，我们将取上一次迭代所得的值（或者初始值）。

```
#E-step 计算第 i 个数据由滴第 k 个分布生成的概率

for i in range(datanum):

    for k in range(classnum):

          postpro = posterior_probability(x[i],pmu[k],detpsigma[k],invpsigma[k])

          gamma[i,k] = pi[k]*postpro

   gamma[i] = gamma[i]/sum(gamma[i])
```

```
def posterior_probability(x,pmu,detpsigma,invpsigma): #计算 N（xi | μk,Σk）
    D = len(x)
    #-1/2*(x-μ)T*Σ^-1(x-μ)
    t = -0.5 * dot(dot((x-pmu).T,invpsigma),x-pmu)
    posprob = (2*pi)**(-D/2) * (detpsigma**(-0.5)) * exp(t)
    return posprob
```

2. 估计每个 Component 的参数：现在我们假设上一步中得到的 $\gamma(i,k)$ 就是正确的“数据 x_i由 Component k 生成的概率”，亦可以当做该 Component 在生成这个数据上所做的贡献，或者说，我们可以看作 x_i 这个值其中有 $\gamma(i,k)x_i$ 这部分是由 Component k 所生成的。集中考虑所有的数据点，现在实际上可以看作 Component 生成了 $\gamma(1,k)x_1,\ldots,\gamma(N,k)x_N$ 这些点。由于每个 Component 都是一个标准的 Gaussian 分布，可以很容易分布求出最大似然所对应的参数值：

$$\mu_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)x_i$$

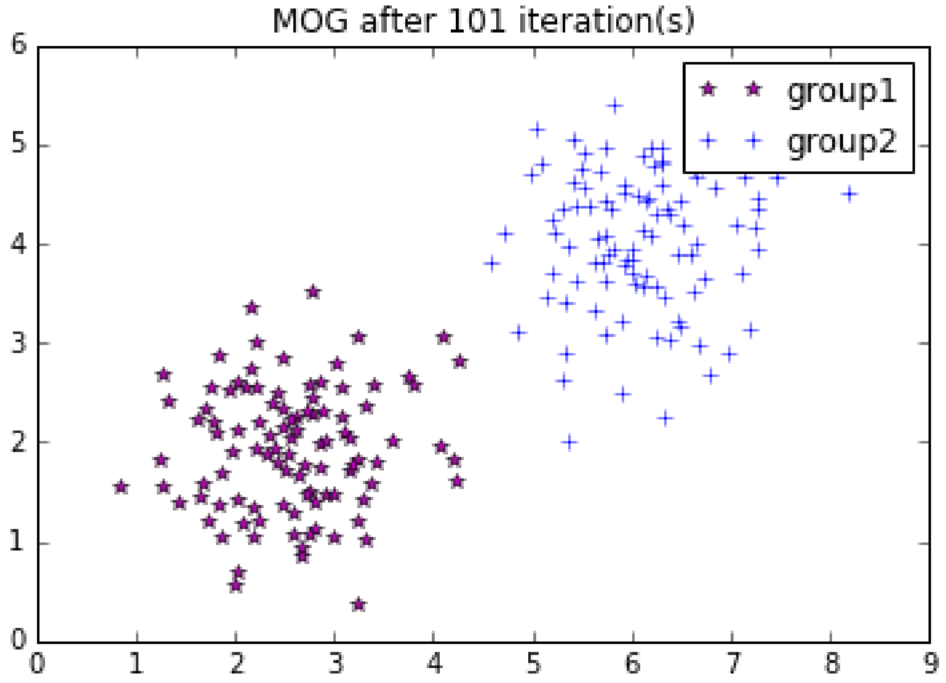
$$\Sigma_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)(x_i-\mu_k)(x_i-\mu_k)^T$$

其中 $N_k = \sum_{i=1}^{N}\gamma(i,k)$ ，并且 π_k 也顺理成章地可以估计为 N_k/N 。

```
pi = NK/datanum
for k in range(classnum):
    for i in range(datanum):
        pmu[k] = pmu[k] + gamma[i][k]*x[i]
    pmu[k] = pmu[k]/NK[k]
    for i in range(datanum):
        temp = mat(x[i]-pmu[k])
        psigma[k] = psigma[k] + gamma[i][k]*dot(temp.T,temp)
    psigma[k] = psigma[k]/NK[k]
```

3.重复迭代前面两步，直到似然函数的值收敛为止。在本实验中直接设定其进行100次迭代。

实验结果：



实验小结：EM 算法的基本思路是随机初始化一组参数 $\theta^{(0)}$，根据后验概率 $Pr(Y|X;\theta)$来更新 Y 的期望 E(Y)，然后用 E(Y)代替 Y 求出新的模型参数 $\theta^{(1)}$。如此迭代直到 θ 趋于稳定。E 步是假设模型参数已知的情况下求隐含变量 Z 分别取 $z_1,z_2,\ldots$的期望，亦即 Z 分别取 $z_1,z_2,\ldots$的概率。M 步是用最大似然的方法求出模型参数。