# GMM

高斯混合模型思想:

观察数据的分布时，有时会发现数据并不是呈单一高斯分布的，分布比较错乱抓不到规律时，这个时候可以采用高斯混合模型。

高斯混合模型有一个超参数需要确定，就是高斯的个数。

高斯混合模型使用 EM 算法进行求解。

EM 算法使用前需要损失函数是凸函数。

$$\sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \right\}$$

Logx 凸函数

与 hmm 一样，高斯混合模型使用 em 需要构造一个变量，这个变量可以用来表示均值方差

变量：E-step

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

参数（M-step）:

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

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

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

伪代码:

1）初始化均值方差和 pi

2）可以计算 E-step,

3) 通过 2）可以计算均值和 pi

4）通过 2）3）可以计算协方差

5）根据 3）4）更新第 2）步，一直循环计算 2）3）4）直到收敛