

# EM Algorithm

---

Author: 相忠良(Zhong-Liang Xiang)

Email: [ugood@163.com](mailto:ugood@163.com)

Date: Aug. 24st, 2017

## 0. 目的

---

求混合高斯分布(或其他分布, 但我们以 Gaussian Distribution 为例)模型的参数 .

## 1. 认识

---

谈及EM算法, 我们认为是几个高斯分布(也叫多个组份)协同地生成了一个样本 .  
若各组分方差相同, 就认为是 **k-means 算法模型** .

## 2. 故事

---

假设有这样一个问题 :

1. 有身高样本10000个, 既  $N = 10000$ ;
2. 身高由男人、女人组成 ;
3. 认为男性身高  $\sim Normal(\mu_1, \sigma_1^2)$ ;  
女性身高  $\sim Normal(\mu_2, \sigma_2^2)$ ;
4. 10000个人中男女比例未知, 既  $\pi_1, \pi_2$  未知  
求未知参数  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi_1$  and  $\pi_2$ .

上面的故事中, 样本  $x_i$  是一维的 . 若  $x_i$  是5维的, 则对应的 :

- $\mu$  是5维的;
- $\Sigma$  是5\*5的半正定方阵;
- $\pi$  是比例, 仍为标量.